Notes on Probability

University of Chicago, 2014

Vivak Patel

January 27, 2015

Contents

Part I
Fundamental Results

1 Basic Measure Theory

1.1 Special Structures on Sets

1. π-system, semi-algebra, algebra, monotone class, λ-system, and σ-algebra.

   Definition 1.1. *Let Ω be a non-empty set.*

   (a) *A π-system, $\mathcal{P}$, is a collection of subset of Ω that is closed under finite intersections and contains the empty set. That is, if $A, B \in \mathcal{P}$ then $A \cap B \in \mathcal{P}$, and $\emptyset \in \mathcal{P}$.*

   (b) *A semi-algebra, $\mathcal{S}$, is a collection of subsets of Ω which satisfy the following:*

      i. *$\mathcal{S}$ is a π-system (contains the empty set and is closed under finite intersections)*

      ii. *For $S \in \mathcal{S}$, S^c is the finite, disjoint union of elements in $\mathcal{S}$.*

   (c) *An algebra, $\mathcal{A}$, is a collection of subsets of Ω which satisfy:*

      i. *$\Omega \in \mathcal{A}$*

      ii. *$\mathcal{A}$ is closed under complementation (i.e. if $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$)*

      iii. *$\mathcal{A}$ is closed under finite unions (i.e. if $A, B \in \mathcal{A}$ then $A \cup B \in \mathcal{A}$)*

   (d) *A monotone class, $\mathcal{M}$, is a collection of subsets of Ω which satisfy:*

      i. *Closure under the union of increasing nested elements of $\mathcal{M}$. (i.e. if $A_i \in \mathcal{M}$ such that $A_i \subset A_{i+1}$ then $\bigcup_i A_i \in \mathcal{M}$).*

      ii. *Closure under the intersection of decreasing nested elements of $\mathcal{M}$. (i.e. if $A_i \in \mathcal{M}$ such that $A_i \supset A_{i+1}$ then $\bigcap_i A_i \in \mathcal{M}$).*

   (e) *A λ-system, $\mathcal{L}$, is a collection of subsets of Ω which satisfy:*

      i. *$\Omega \in \mathcal{L}$*

      ii. *$\mathcal{L}$ is closed under complementation*

      iii. *$\mathcal{L}$ is closed under countable disjoint unions*

   (f) *A σ-algebra, $\mathcal{F}$, is a collection of subset of Ω which satisfy:*

      i. *$\Omega \in \mathcal{F}$*

      ii. *$\mathcal{F}$ is closed under complementation*

      iii. *$\mathcal{F}$ is closed under countable unions*

2. Examples

   Example 1.1. *Let $\Omega = \{1, 2, 3, 4\}$.*

   (a) *$\mathcal{L} = \{\emptyset, \Omega, \{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}, \{3, 4\}\}$ is a λ-system which is not an algebra*

   (b) *$\mathcal{P} = \{\emptyset, \{1, 3\}\}$. Then $\mathcal{P}$ is a π-system. Moreover, the σ-algebra it generates is strictly contained in $\mathcal{L}$.*

(c) $\mathcal{M} = \{\{1\}, \{1,3\}\}$ *is a monotone class which is not an algebra.*

Example 1.2. *Let $\mathcal{S}_d$ be the empty set and subsets of $\mathbb{R}^d$ of the form:*

$$(a_1, b_1] \times \cdots \times (a_d, b_d] \qquad -\infty \leq a_i < b_i \leq \infty$$

Then $\mathcal{S}_d$ is a semi-algebra but not an algebra (none of the complements are in $\mathcal{S}_d$).

Example 1.3. *Let $\Omega = \mathbb{Z}$ and $\mathcal{A}$ be the collection of subsets of Ω which have finite cardinality or which have complements of finite cardinality. Then $\mathcal{A}$ is an algebra and it is not a σ-algebra since we can generate $2\mathbb{Z}$ by countable unions, but this set is not in $\mathcal{A}$.*

Example 1.4. *Let $\Omega = (0,1]$ and consider all left-half open intervals. Then though countable intersections we can generate $\{1\}$ but this is not in the collection of left-half open intervals. Hence, this collection is an algebra, but not a σ-algebra.*

3. Basic Properties

(a) Intersections preserve structure

Lemma 1.1. *Let $\mathcal{F}_i$, $i \in I$, be an arbitrary collection σ-algebras or algebras. Then their intersection is also a σ-algebra or algebra (respectively).*

Proof. Check each condition. □

(b) Converting semi-algebras into algebras

Lemma 1.2. *Let $\mathcal{S}$ be a semi-algebra. Let $\mathcal{A}$ be the collection of all finite disjoint unions of sets in S. Then $\mathcal{A}$ is an algebra.*

Proof. Since $\emptyset \in \mathcal{S}$ and its complement, Ω, can be expressed as the finite disjoint union of elements in $\mathcal{S}$, $\Omega \in \mathcal{A}$. Let $A, B \in \mathcal{A}$. Then there are $S_1, \ldots, S_m, T_1, \ldots, T_n \in \mathcal{S}$ such that $A = +_{i=1}^{m} S_i$ and $B = +_{j=1}^{n} T_j$. Then $A \cap B = +_{i=1,j=1}^{m,n} S_i \cap T_j$. $S_i \cap T_j \in \mathcal{S}$ and these are all disjoint. Hence, $A \cap B \in \mathcal{A}$. Finally, $A^c = \bigcap_{i=1}^{m} S_i^c$. Each S_i^c can be expressed as finitely many disjoint unions elements of $\mathcal{S}$. Hence, A^c can be expressed as the finite disjoint unions of elements of $\mathcal{S}$. Therefore, $A^c \in \mathcal{A}$. We can conclude that $\mathcal{A}$ is an algebra. □

(c) The intersection of monotone classes and algebras is σ-algebras.

Lemma 1.3. *Let $\mathcal{F}$ be a monotone class and an algebra. Then $\mathcal{F}$ is a σ-algebra.*

Proof. $\Omega \in \mathcal{F}$ and if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$, which follow directly from $\mathcal{F}$ being an algebra. Let $A_1, \ldots \in \mathcal{F}$. Let $B_1 = A_1$ and $B_n = \bigcup_{i=1}^{n} A_i$. Then $B_n \in \mathcal{F}$ since $\mathcal{F}$ is an algebra and $B_n \subset B_{n+1}$. Since $\mathcal{F}$ is a monotone class, $\bigcup_i A_i = \bigcup_i B_i \in \mathcal{F}$. Hence, $\mathcal{F}$ is a σ-algebra. □

(d) The intersection of π-systems and λ-systems is σ-algebras.

Lemma 1.4. *Let $\mathcal{F}$ be a π-system and a λ-system. Then $\mathcal{F}$ is a σ-algebra.*

Proof. $\Omega \in \mathcal{F}$ and if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$ follow from $\mathcal{F}$ being a λ-system. Let $A_1, \ldots \in \mathcal{F}$. Let $B_1 \in \mathcal{F}$. Let $B_1 = A_1$ and

$$B_n = A_n \cap A_{n-1}^c \cap \cdots \cap A_1^c$$

Hence, all B_n are disjoint and $B_n \in \mathcal{F}$ since it is a π-system. And so $\bigcup_i A_i = \bigcup_i B_i \in \mathcal{F}$. □

(e) A λ-system is closed under proper set differences

Lemma 1.5. *Let $\mathcal{L}$ be a λ-system and $A, B \in \mathcal{L}$ such that $A \subset B$ (strict). Then $B - A \in \mathcal{L}$.*

Proof. We want to show that $A^c \cap B \in \mathcal{L}$. Notice that $B^c = A^c \cap B^c$, $\emptyset = A \cap B^c$, $A = A \cap B$, and $A^c \cap B$ are disjoint. Since $B^c, A \in \mathcal{L}$, $B^c \cup A \in \mathcal{L}$. Hence, $B \cap A^c = (B^c \cup A)^c \in \mathcal{L}$. □

4. Halmos' Monotone Class Theorem

Theorem 1.1. *Let $\mathcal{A}$ be an algebra and $\mathcal{M}$ be a monotone class such that $\mathcal{A} \subset \mathcal{M}$. Then $\sigma(\mathcal{A}) \subset \mathcal{M}$.*

Proof. Define $m(\mathcal{A})$ to be the smallest monotone class containing $\mathcal{A}$. To show that $m(\mathcal{A})$ is closed under complementation, define $\bar{m}(\mathcal{A}) = \{A : A^c \in m(\mathcal{A})\}$. Show that $\bar{m}(\mathcal{A})$ is a monotone class containing $\mathcal{A}$. This implies that for any $A \in m(\mathcal{A})$ then $A^c \in m(\mathcal{A})$.

Now define

$$\mathcal{M}_1 = \{A : A \cup B \in m(\mathcal{A}), \forall B \in m(\mathcal{A})\}$$

This is a monotone class, and if we can show that $\mathcal{A} \subset \mathcal{M}_1$ then the result follows. So first we consider the more general case of showing $\mathcal{A} \subset \mathcal{M}_2$ which is:

$$\mathcal{M}_2 = \{A : A \cup B \in m(\mathcal{A}), \forall B \in \mathcal{A}\}$$

Clearly, $\mathcal{A} \subset \mathcal{M}_2$, which is a monotone class. Hence, $m(\mathcal{A}) \subset \mathcal{M}_2$. Now, if $A \in \mathcal{A}$ and $B \in m(\mathcal{A}) \subset \mathcal{M}_2$ then $A \cup B \in m(\mathcal{A})$. Hence, $\mathcal{A} \subset \mathcal{M}_1$. □

5. Dynkin's π-λ Theorem

Theorem 1.2. *Let $\mathcal{P}$ be a π-system contained in $\mathcal{L}$ a λ-system. Then $\sigma(\mathcal{P}) \subset \mathcal{L}$.*

Proof. Let L_0 be the smallest λ-system containing $\mathcal{P}$. We show that L_0 is a π-system. We follow the same strategy as in the Monotone Class Theorem. Define:

$$\mathcal{L}_1 = \{A : A \cap B \in L_0,\ \forall B \in L_0\}$$

If $A \in \mathcal{L}_1$ then $A \cap B \in L_0$ for any $B \in L_0$ and by **Lemma 1.5**, $A^c \cap B \in L_0$. This implies that $A^c \in L_0$. The property of countable disjoint unions follows from L_0 being a λ-system, hence $\mathcal{L}_1$ is a λ-system. We want to show that $\mathcal{P} \in \mathcal{L}_1$. To do this, we consider the more general case:

$$\mathcal{L}_2 = \{A : A \cap B \in L_0,\ \forall B \in \mathcal{P}\}$$

Again, $\mathcal{L}_2$ is a λ-system. Moreover, it contains $\mathcal{P}$. Hence, $L_0 \in \mathcal{L}_2$. Now let $P \in \mathcal{P}$ and $A \in L_0 \in \mathcal{L}_2$ then $P \cap A \in L_0$ for any $A \in L_0$. Hence, $\mathcal{P} \in \mathcal{L}_1$. Thus, $L_0 \subset \mathcal{L}_1$, so it is a π-system. It follows that L_0 is a σ-algebra. So $\sigma(\mathcal{P}) \subset L_0$. □

6. Generated Algebras and σ-Algebras

   (a) Generated Algebra. Generated σ-Algebra.

   Definition 1.2. *Let S be a collection of subsets of a non-empty set Ω.*

   i. The smallest algebra containing S, $a(S)$, is the algebra generated by S.

   ii. The smallest σ-algebra containing S, $\sigma(S)$, is the σ-algebra generated by S

   (b) Existence of a generated algebra.

   Lemma 1.6. *Let S be a collection of subsets of Ω. Then there exists a smallest algebra generated by S.*

   Proof. Let $\mathcal{M}$ be the collection of all algebras containing S. This is non-empty since the power set of Ω is an algebra containing S. Taking the intersection over all $\mathcal{M}$ results in $a(S)$. □

   (c) Existence of a generated σ-algebra

   Lemma 1.7. *Let S be a collection of subsets of Ω. Then there exists a smallest σ-algebra generated by S.*

   Proof. Let $\mathcal{M}$ be the collection of all σ-algebras containing S. This is non-empty since the power set of Ω is a σ-algebra containing S. Taking the intersection over all $\mathcal{M}$ results in $\sigma(S)$. □

7. Borel Sets

   Definition 1.3. *The Borel Sets are the σ-algebra generated by the open sets of a topology. Specifically, the Borel Sets usually refer to th open sets of $\mathbb{R}^d$ and, depending on the dimension, are denoted $\mathcal{B}^d$.*

8. Measurable Space

   Definition 1.4. *A Measurable Space is the double of a non-void set and a σ-algebra on that set.*

1.2 Measures

1. σ-finite. Premeasure. Measure. Probability Measure. Measure Space. Probability Space.

   Definition 1.5. *Let Ω be a non-void set.*

   (a) *A set function μ on subsets of Ω is σ-finite if on subsets A_n which satisfy $\bigcup_n A_n = \Omega$ also satisfy $\mu(A_n) < \infty$.*

   (b) *Let $\mathcal{A}$ be an algebra on Ω. A pre-measure, μ, is a set function on $\mathcal{A}$ that satisfies:*

      i. $\mu(\emptyset) = 0$

      ii. $\mu(A) \geq 0$ *for any* $A \in \mathcal{A}$

      iii. *If $A_i \in \mathcal{A}$ are disjoint AND $\bigcup_i A_i \in \mathcal{A}$ then*

$$\mu\left(\bigcup_i A_i\right) = \sum_i \mu(A_i)$$

   (c) *Let $\mathcal{F}$ be a σ-algebra on Ω. A (positive) measure is a set function, μ, on $\mathcal{F}$ which is non-negative (possibly infinite) and is countable additive. A probability measure is a special case of a measure in which $\mu(\Omega) = 1$.*

   (d) *A measure space or probability space is the triple of a non-void set, a σ-algebra, and the measure or probability measure (resp.).*

2. Caratheodory Extension Theorem

   Theorem 1.3. *Let Ω be a non-empty set.*

   (a) *Let $\mathcal{S}$ be a semi-algebra on Ω and μ be a non-negative set function on $\mathcal{S}$ such that $\mu(\emptyset) = 0$. Suppose*

      i. *Whenever $S \in \mathcal{S}$ such that $S = +_{i=1}^{m} S_i$ where $S_i \in \mathcal{S}$ then μ is finitely additive.*

      ii. *Whenever $S \in \mathcal{S}$ such that $S = +_{i=1}^{\infty} S_i$ where $S_i \in \mathcal{S}$ then μ is countably subadditive.*

      Then μ has a unique extension (agrees on $\mathcal{S}$) to $\bar{\mu}$, a premeasure on $a(\mathcal{S})$.

   (b) *If $\bar{\mu}$ is a pre-measure on an algebra, $\mathcal{A}$, on Ω, then it has an extension (agrees on $\mathcal{A}$) to $\sigma(\mathcal{A})$.*

   (c) *Moreover, if $\bar{\mu}$ is σ-finite then the extension is unique.*

3. Construction of Measures on the Real Line

   Example 1.5. *First we construct a Steltjes Measure Function, F, which satisfies:*

   (a) *F is non-decreasing*

   (b) *F is right-continuous*

Next, on the semi-algebra of intervals of the form $(a, b]$, *we define the non-negative set function* μ *such that* $\mu(\emptyset) = 0$ *and:*

$$\mu(a, b] = F(b) - F(a) \qquad -\infty \leq a < b \leq \infty$$

Finally, by the extension theorem, we have a unique measure on $(\mathbb{R}, \mathcal{B})$.

4. Basic Properties

Proposition 1.1. *Let* $(\Omega, \mathcal{F}, \mu)$ *be a measure space.*

(a) μ *is monotonic.*

(b) μ *is subadditive*

(c) On $A, A_i \in \mathcal{F}$ *such that* $A_i \uparrow A$, μ *is continuous from below.*

(d) On $A, A_i \in \mathcal{F}$ *such that* $\mu(A_1) < \infty$ *and* $A_i \downarrow A$, μ *is continuous from above.*

Proof. (a) Suppose $A, B \in \mathcal{F}$ and $A \subset B$. Then $C = A^C \cap B \in \mathcal{F}$. Hence:

$$\mu(A) \leq \mu(A) + \mu(C) = \mu(B)$$

(b) Let $A_i \in \mathcal{F}$. Let $B_1 = A_1$ and $B_i = A_i - B_{i-1}$. Then, by countable additivity:

$$\mu(\cup_i A_i) = \mu(\cup_i B_i) = \sum_i \mu(B_i) \leq \sum_i \mu(A_i)$$

(c) Let $B_1 = A_1$ and $B_n = A_n - A_{n-1}$. Then:

$$\mu(A) = \sum_{i=1}^{\infty} \mu(B_i) = \lim_{n \to \infty} \sum_{i=1}^{n} \mu(B_i) = \lim_{n \to \infty} \mu(A_n)$$

(d) We have $A_1 - A_i \uparrow A_1 - A$. Hence, $\mu(A_1 - A_i) \uparrow \mu(A_1 - A)$. Since $\mu(A_1) < \infty$, the following holds:

$$\mu(A_1) - \mu(A_i) \uparrow \mu(A_1) - \mu(A)$$

Which implies:

$$\mu(A_i) \downarrow \mu(A)$$

□

2 Random Variables

2.1 Measure-Theoretic Considerations

1. Measurable Function. Random Vector. Random Variable. σ-Algebra Generated by a Random Vector

Definition 2.1. *Let* $(\Omega, \mathcal{F})$ *and* $(S, \mathcal{S})$ *be measurable spaces.*

(a) *A function* $f : \Omega \to S$ *is measurable if the pre-images of measurable sets are measurable. That is for any* $B \in \mathcal{S}$, $f^{-1}(B) = \{\omega : f(\omega) \in B\} \in \mathcal{F}$.

(b) *Suppose* $(S, \mathcal{S}) = (\mathbb{R}^d, \mathcal{B}^d)$. *Then a measurable function is called a random vector. If* $d = 1$ *then it is called a random variable.*

(c) *The* σ*-algebra generated by a random vector is the smallest* σ*-algebra on the domain which makes the random vector measurable.*

2. Checking Measurability of a a Function

Proposition 2.1. *Suppose* $\mathcal{A}$ *generated* $\mathcal{S}$. *Moreover, suppose* $\forall A \in \mathcal{A}$, $f^{-1}(A) \in \mathcal{F}$. *Then* f *is measurable.*

Proof. We need to show that $\forall B \in \mathcal{S}$, $f^{-1}(B) \in \mathcal{F}$. To do this, we consider:

$$\mathcal{B} := \{B \subset S : f^{-1}(B) \in \mathcal{F}\}$$

$\mathcal{B}$ is a σ-algebra. Since $\mathcal{A} \subset \mathcal{B}$ then $\mathcal{S} \subset \mathcal{B}$. Therefore, f is measurable. □

3. Useful applications of checking measurability

Example 2.1. *Suppose* $X : \Omega \to \mathbb{R}^*$. *We can check that* $X^{-1}[-\infty, q)$ *for all* $q \in \mathbb{Q}$ *is measurable.*

Example 2.2. *If the co-domain is* $\mathbb{R}^d$, *we can simply check over all open rectangles with rational end-points.*

4. Existence of generated σ-algebra

Lemma 2.1. *Let* f *be a function from* Ω *to* $(S, \mathcal{S})$. *Then, there is a* $\sigma(f)$, *the smallest* σ*-algebra on* Ω, *which makes* f *measurable.*

Proof. Consider the collection $\mathcal{G} := \{\{\omega : f(\omega) \in B\} : B \in \mathcal{S}\}$. We show that $\mathcal{G}$ is a σ-algebra by checking all of the conditions.

(a) $\Omega \in \mathcal{G}$ since $f^{-1}(S) = \Omega$.

(b) If $\{f \in B\} \in \mathcal{G}$, then $\{f \in B\}^C = \{f \in B^c\} \in \mathcal{G}$.

(c) If $f^{-1}(B_i) \in \mathcal{G}$, then $\cup_i f^{-1}(B_i) = f^{-1}(\cup_i B_i) \in \mathcal{G}$.

Thus, the set of σ-algebras making f measurable is non-empty. Hence, there is a smallest σ-algebra making f measurable (through arbitrary intersections). □

5. Compositions of Measurable Functions

Proposition 2.2. *Let* $(\Omega, \mathcal{F})$ *be a measurable space.*

(a) *Let* $(S, \mathcal{S}), (T, \mathcal{T})$ *be measurable spaces. Suppose* f, g *are measurable functions such that* $f : \Omega \to S$ *and* $g : S \to T$. *Then* $g \circ f$ *is measurable.*

(b) Let $X_1, \ldots, X_n$ be random variables and $f : \mathbb{R}^n \to \mathbb{R}$ be a Borel function. Then $f(X_1, \ldots, X_n)$ is a random variable.

(c) If $X_1, \ldots, X_n$ are random variables then $X_1 + \ldots + X_n$ is a random variable.

Proof. (a) Let $B \in \mathcal{T}$. Then $g^{-1}(B) \in \mathcal{S}$ and so $f^{-1} \circ g^{-1}(B) \in \mathcal{F}$.

(b) We need to show that $(X_1, \ldots, X_n)$ is measurable since composing this with $\mathcal{F}$ does not cause any difficulties. Moreover, this requires that we know the product σ-algebra $\mathcal{B}^n$ is generated by cross-products of Borel sets. Hence, by **Prop 2.1**, we need to check measurable rectangles:

$$\{\omega : (X_1, \ldots, X_n) \in A_1 \times \cdots \times A_n\} \in \mathcal{F}$$

Since X_i are random variables, and this set is equal to $\cap_i X_i^{-1}(A_i)$, it is in $\mathcal{F}$.

(c) By the previous two points, we need only check that $f(x_1, \ldots, x_n) = x_1 + \cdots + x_n$ is a measurable function. f is continuous and hence the pre-images of open sets are open sets. Thus, by **Prop 2.1**, f is Borel measurable.

□

6. Limits, Infimums and Supremums of Measurable Functions

Proposition 2.3. *If $X_1, X_2, \ldots$ are random variables then so are:*

$$\inf_n X_n \quad \sup_n X_n \quad \limsup_n X_n \quad \liminf_n X_n$$

Proof.

(a) Let $q \in \mathbb{Q}$. $\{\inf_n X_n < q\} = \bigcup_n \{X_n < q\}$. The right hand side is a measurable set. By **Prop 2.1**, the infimum is measurable.

(b) Let $q \in \mathbb{Q}$. $\{\sup_n X_n < q\} = \bigcap_n \{X_n < q\}$. The right hand side is a measurable set. By **Prop 2.1**, the supremum is measurable.

(c) We have that for each $k \in \mathbb{N}$, $\sup_{n>k} X_n$ is a random variable. Taking the infimum over k of these random variables is measurable and is also the limit supremum.

(d) We have that for each $k \in \mathbb{N}$, $\inf_{n>k} X_n$ is a random variable. Taking the supremum over k of these random variables is measurable and is also the limit infimum.

□

2.2 Distributions

1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a Probability space. Let X be a random variable on the probability space.

2. Distribution of X. Distribution Function of X.

   Definition 2.2. *The set function $\mu : \mathcal{B} \to [0, 1]$ defined by $\mu(\mathcal{B}) = \mathbb{P}[X \in B]$ is the distribution of X. The Distribution Function of X is the function $F(x) = \mathbb{P}[X \leq x] = \mu([-\infty, x])$.*

3. Distributions are probability measures

   Lemma 2.2. *The distribution of X is a probability measure on $(\mathbb{R}, \mathcal{B})$.*

   Proof. Let μ be the distribution of X.

   (a) $\mu(\emptyset) = \mathbb{P}[X \in \emptyset] = \mathbb{P}[\emptyset] = 0$

   (b) Let $B_1, B_2, \ldots \in \mathcal{B}$ be disjoint. Then $\{X \in B_1\}, \{X \in B_2\}, \ldots$ are disjoint. Hence:

   $$\mu\left(\bigcup_i B_i\right) = \mathbb{P}\left[X \in \bigcup_i B_i\right] = \sum_i \mathbb{P}[X \in B_i] = \sum_i \mu(B_i)$$

   □

4. Properties of Distribution Functions

   Lemma 2.3. *Let F be a distribution function of X. Then F has the following properties.*

   (a) F is non-decreasing

   (b) $\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$.

   (c) F is right continuous

   (d) Denote $F(x-) = \lim_{y\uparrow x} F(y)$. Then $F(x-) = \mathbb{P}[X < x]$

   (e) $\mathbb{P}[X = x] = F(x) - F(x-)$

   Proof.

   (a) This follows from the monotonicty of measures.

   (b) Note that $\cap_n(-\infty, -n] = \emptyset$ and $\cup_n(-\infty, n] = \mathbb{R}$.

   (c) Let $y_n \downarrow x$. Then $\cap_n(-\infty, y_n] = (-\infty, x]$.

   (d) Let $y_n \uparrow x$. Then $\cup_n(-\infty, y_n] = (-\infty, x)$.

   (e) $\mathbb{P}[X = x] = \mathbb{P}[X \leq x] - \mathbb{P}[X < x] = F(x) - F(x-)$.

   □

5. Characterization of a Distribution Function. Inverse of a Distribution Function (in proof).

Lemma 2.4. *If F is non-decreasing, has limits 0 and 1, and is right continuous then there is a random variable for which F is its distribution function.*

Proof. Consider the interval $(0,1)$ with the usual σ-algebra and the Lebesgue measure. Define random variable X for $\omega \in (0,1)$ by:

$$X(\omega) = \sup y : F(y) < \omega$$

We need to show $\{X \leq x\} = \{\omega \leq F(x)\}$.

(a) If $X(\omega) \leq x$ then $F(x) \geq \omega$. Hence $\{X \leq x\} \subset \{\omega \leq F(x)\}$

(b) If $X(\omega) > x$ then $F(x) < \omega$. Hence, $\{X \leq x\}^c \subset \{\omega \leq F(x)\}^c$.

Hence, $\mu X \leq x = \mu\omega \leq F(x) = F(x)$. □

2.3 Defining the Integral

1. First, we define the integral for simple functions, bounded functions, non-negative functions and finally all measurable functions with respect to a measure space. Then, we prove the following results for each set of functions:s

   Lemma 2.5. *Let f, g be measurable functions w.r.t $(\Omega, \mathcal{F}, \mu)$.*

   (a) (Positivity) If $f \geq 0$ then $\int f d\mu \geq 0$

   (b) (Scalar Multiplication) If $a \in \mathbb{R}$ then $\int af d\mu = a \int f d\mu$

   (c) (Linearity) $\int (f+g) d\mu = \int f d\mu + \int g d\mu$

   (d) (Monotonicity) If $f \leq g$ a.e. then $\int f d\mu \leq \int g d\mu$

   (e) (a.e-Equal) If $f = g$ a.e. then $\int f d\mu = \int g d\mu$

   (f) $|\int f d\mu| \leq \int |f| d\mu$

2. The last three items can be proven from the first three given some extra assumptions:

   Lemma 2.6. *Let f, g be measurable functions w.r.t. $(\Omega, \mathcal{F}, \mu)$.*

   *(a) If (a) and (c) from **Lemma 2.5** hold then (d) holds.*

   *(b) If (a) and (c) from **Lemma 2.5** hold then (e) holds.*

   *(c) If (a) and (c) from **Lemma 2.5** hold and (b) holds for $a = -1$ then (f) holds.*

   Proof.

   (a) If (a) and (c) hold then since $f - g \geq 0$ a.e. then $\int (f - g) d\mu \geq 0$. Applying (c) we get the result.

   (b) By the first part of this lemma, the inequality in both directions implies the equality of the integrals.

(c) $f \leq |f|$ a.e. and $-f \leq |f|$. By (b) we have that $\int f d\mu \leq \int |f| d\mu$ and $-\int f d\mu \leq \int |f| d\mu$.

□

3. Simple Functions

   (a) Simple Functions. Integral of Simple Function.

   Definition 2.3. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space.*

   i. *ϕ is a simple function if ϕ can be written as $\phi(\omega) = \sum_{i=1}^{n} \alpha_i \mathbf{1}\{A_i\}$ where $n < \infty$, $A_i \in \mathcal{F}$ are disjoint and $\mu(A_i) < \infty$.*

   ii. *The integral of a simple function is defined as:*

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

   (b) Proof of **Lemma 2.5**

   Proof.

   i. If $\phi \geq 0$ then $\alpha_i \geq 0$. Since $\mu(A_i) \geq 0$, then $\sum_{i=1}^{n} \alpha_i \mu(A_i) \geq 0$.

   ii. Let $a \in \mathbb{R}$. Then:

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

   iii. If $\phi = \sum_{i=1}^{n} \alpha_i \mathbf{1}\{A_i\}$ and $\psi = \sum_{j=1}^{m} \beta_j \mathbf{1}\{B_j\}$ then $\{A_i \cap B_j\}$ are disjoint. Therefore $\phi + \psi = \sum_{i,j}^{n,m} (\alpha_i + \beta_j) \mathbf{1}\{A_i \cap B_j\}$. Hence by countable additivity of μ:

$$\sum_{i=1}^{n} \sum_{j=1}^{m} (\alpha_i + \beta_j) \mu(A_i \cap B_j) = \sum_{i=1}^{n} \alpha_i \mu(A_i) + \sum_{j=1} \beta_j \mu(B_j)$$

□

4. Bounded Functions

   (a) Bounded functions with Finite Measure Support (BFMS). Integral of Bounded Function.

   Definition 2.4. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space.*

   i. *Let f be measurable. f is bounded with finite-measure support if f is non-zero only on a set $E \in \mathcal{F}$ such that $\mu(E) < \infty$.*

   ii. *Let $\mathcal{S}$ be the collection of all simple functions. The integral of a BFMS function is:*

$$\int f d\mu = \sup \left\{ \int \psi d\mu : \psi \in \mathcal{S},\ \psi \leq f \ a.e. \right\}$$

   (b) Equivalent characterization of integral

Lemma 2.7. *Let f be a BFMS function. Then:*

$$\sup\left\{\int \psi d\mu : \psi \in \mathcal{S},\ \psi \leq f \ a.e.\right\} = \inf\left\{\int \phi d\mu : \phi \in \mathcal{S},\ \phi \geq f \ a.e.\right\}$$

Proof. By assumption, $\exists M < \infty$ such that $|f| \leq M$. And let E be the support of f. Then define:

$$E_{k,n} = \left\{\omega \in E : \frac{k}{n}M \leq f(\omega) < \frac{k+1}{n}M\right\}$$

and let $\psi_n = \sum_{k=0}^{n-1} \frac{k}{n}M\mathbf{1}\{E_{k,n}\}$ and $\phi_n = \sum_{k=0}^{n-1} \frac{k+1}{n}M\mathbf{1}\{E_{k,n}\}$. Then: $\psi_n \leq f \leq \phi_n$ for all n. Moreover:

$$\int \phi_n - \int \psi_n = \int \phi_n - \psi_n = \frac{M}{n}\mu(E) \to 0 \quad n \to \infty$$

Hence, the result holds. □

(c) Proof of **Lemma 2.5**

Proof.

i. If $f \geq 0$, 0 is a simple function hence by the same property for simple functions $\int f d\mu \geq \int 0 d\mu = 0$.

ii. Let $a \in \mathbb{R}$. Then for $a \geq 0$:

$$\int af d\mu = \sup_{\phi \leq af} \int \phi = \sup_{\phi' \leq f} \int a\phi' = a \int f d\mu$$

If $a < 0$ then:

$$\int af d\mu = \inf_{\phi' \geq f} \int a\phi' = a \int f d\mu$$

iii. Let $\phi_1 \geq f$ and $\phi_2 \geq g$ then $\phi_1 + \phi_2 \geq f + g$, then:

$$\int (f+g) d\mu \leq \int \phi_1 + \int \phi_2$$

Since this holds for any $\phi_1 \geq f$ and $\phi_2 \geq g$ we have that:

$$\int (f+g) d\mu \leq \int f d\mu + \int g d\mu$$

For the other direction, let $\psi_1 \leq f$ and $\psi_2 \leq g$. Then:

$$\int_{\psi_1} + \int_{\psi_2} \leq \int (f+g) d\mu$$

Hence, $\int f d\mu + \int g d\mu \leq \int (f+g) d\mu$.

□

5. Non-negative functions.

(a) Integral of non-negative function.

Definition 2.5. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Let $f \geq 0$ be a measurable function. Let $\mathcal{M}$ be the set of all BFMS functions. Then the integral of a non-negative function f is defined as:*

$$\int f d\mu = \sup\left\{\int g d\mu : g \leq f,\ g \in \mathcal{M}\right\}$$

(b) Particular construction converging to the integral of $f \geq 0$.

Lemma 2.8. *Suppose μ is σ-finite. Let $E_n \uparrow \Omega$ be such that $\mu(E_n) < \infty$. Then:*

$$\int_{E_n} f \wedge n d\mu \uparrow \int f d\mu$$

Proof. First note that $f \wedge n \mathbf{1}\{E_n\} \leq f$ and is in $\mathcal{M}$. Hence, $\int_{E_n} f \wedge n d\mu \leq \int f d\mu$. To get the inequality in the other direction, let $0 \leq g \leq f$, $g \in \mathcal{M}$, and $g \leq M$ for some $M > 0$. Then for $n \geq M$:

$$\int_{E_n} f \wedge n d\mu \geq \int_{E_n} g d\mu = \int g d\mu - \int_{E_n^c} g d\mu$$

Moreover as $n \to \infty$:

$$\left|\int_{E_n^c} g d\mu\right| \leq M \mu(E_n^c \cap \operatorname{supp}(g)) \to 0$$

Hence:

$$\liminf \int_{E_n} f \wedge n d\mu \geq \int g d\mu$$

We can do this for every function $g \in \mathcal{M}$ such that $g \leq f$, and so the result holds. □

(c) Proof of **Lemma 2.5**

Proof.

i. 0 is a bounded function which is less than or equal to f and it has integral of 0.

ii. If $a > 0$ then from the lemma:

$$\int af = \lim_{n\to\infty} \int_{E_n} af \wedge n = a \lim_{n\to\infty} \int_{E_n} f \wedge n = a \int f$$

iii. We have that:

$$\int_{E_n} (f+g) \wedge n \leq \int_{E_n} f \wedge n + \int_{E_n} g \wedge n$$

Hence, $\int (f+g) \leq \int f + \int g$. For the other direction, let $h_1, h_2 \in \mathcal{M}$ such that $h_1 \leq f$ and $h_2 \leq g$. Then $h_1 + h_2 \in \mathcal{M}$ and $h_1 + h_2 \leq f + g$. Hence:

$$\int h_1 + \int h_2 \leq \int (f+g)$$

Since this holds for all $h_1 \leq f$ and $h_2 \leq g$, then the other direction follows.

□

6. Integral of a Measurable Functions

   (a) Positive part. Negative part.

   Definition 2.6. *Let f be a function. Its positive part is $f^+ = f \vee 0$ and its negative part is $f^- = (-f) \vee 0$.*

   (b) Integrable. Integral of a measurable function.

   Definition 2.7. *A function f is integrable if $\int |f| d\mu < \infty$. If f is integrable, the integral of a measurable function is defined as:*

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

   (c) Minimality of Positive and Negative Parts and Integrability.

   Lemma 2.9. *Suppose $\exists f_1, f_2 \geq 0$ such that $f = f_1 - f_2$. Then $f_1 \geq f^+$ and $f_2 \geq f^-$ and*

$$\int f d\mu = \int f_1 d\mu - \int f_2 d\mu$$

Proof. To prove the first part. For any point $\omega \in \Omega$, we have a few cases:

   i. $f(\omega) = 0$ then $f^+(\omega) = f^-(\omega) = 0$. Then, $f_1(\omega) \geq f^+(\omega)$ and $f_2(\omega) \geq f^-(\omega)$.
   ii. $f(\omega) > 0$ then $f^+(\omega) = f(\omega)$ and $f^-(\omega) = 0$. For a contradiction, if $f_1(\omega) < f(\omega)$ then $f_1(\omega) - f_2(\omega) > f_1(\omega) \implies 0 > f_2(\omega)$. Hence, $f_1(\omega) \geq f^+(\omega)$ and $f_2(\omega) \geq f^-(\omega)$.
   iii. This holds for the negative case as well.

For the second part, since we have only proved **Lemma 2.5(b)** for $a > 0$:

$$\begin{aligned} f^+ + f_2 = f^- + f_1 &\implies \int f^+ + \int f_2 = \int f^- + \int f_1 \\ &\implies \int f^+ - \int f^- = \int f_1 - \int f_2 \\ &\implies \int f = \int f_1 - \int f_2 \end{aligned}$$

□

(d) Proof of **Lemma 2.5**.

Proof.

i. This is the non-negative case.

ii. Suppose $a \geq 0$, then this follows from the non-negative case. If $a < 0$ then $(af)^+ = (-a)f^-$ and $(af)^- = (-a)f^+$. Hence:

$$\int af = \int (-a)f^- - \int (-a)f^+ = -a \int f^- + a \int f^+ = a \int f$$

iii. Note that $(f+g)^+ - (f+g)^- = f^+ + g^+ - (f^- + g^-)$. By the previous lemma:

$$\int (f+g) = \int f^+ + \int g^+ - \int f^- - \int g^- = \int f + \int g$$

□

2.4 Properties of Integrals

1. Jensen's Inequality

Theorem 2.1. *Suppose ϕ is convex. If μ is a probability measure and $f, \phi(f)$ are both integrable, then:*

$$\phi\left(\int f d\mu\right) \leq \int \phi \circ f d\mu$$

Proof. Note that there is always at least one line passing through any point of ϕ which remains below ϕ. We use show this and prove the result based by considering the point $c = \int f d\mu$. By convexity:

$$\frac{\phi(c) - \phi(x)}{c - x} \leq \frac{\phi(y) - \phi(c)}{y - c}$$

for $x < c$ and $y > c$. So the increasing limit on the left exists and the decreasing limit on the right exists. Let m be any value between these two limits. Then:

$$m(y-c) + \phi(c) \leq \phi(y)$$

and

$$m(x-c) + \phi(c) \leq \phi(x)$$

Hence, for any x, we found a line below ϕ passing through $\phi(c)$. Letting $f = x$, and integrating:

$$\begin{aligned}\int \phi \circ f d\mu &\geq \int \phi(c) d\mu + m \int (f - c) d\mu \\ &\geq \phi(c) + m\left(\int f d\mu - c\right) \\ &\geq \phi\left(\int f d\mu\right)\end{aligned}$$

□

2. Holder's Inequality.

Theorem 2.2. *If (p, q) are conjugate exponents then* $\|fg\|_{L^1} \leq \|f\|_{L^p} \|g\|_{L^q}$

Proof. If the either term on the right hand side is infinite, we are done. If either term on the right hand side if 0, this implies that the corresponding function is 0 almost everywhere and so the left hand side is 0. Suppose both terms on the right are non-zero and finite. Let F and G be f and g normalized by the p and q norms respectively. Then by the convexity of the exponential function:

$$FG = \exp\left(\frac{1}{q}\log(F^q) + \frac{1}{p}\log(G^p)\right) \leq \frac{1}{q}F^q + \frac{1}{p}G^p$$

Integrating both sides results in $\|FG\|_1 \leq 1$. □

3. Bounded Convergence Theorem

Theorem 2.3. *Suppose f_n are BFMS with respect to a set E and a bound $M > 0$ such that $f_n \to f$ a.e.. Then:*

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

Proof. Let $\epsilon > 0$. Let $B_n = \{|f_n - f| > \epsilon\}$. Then:

$$\begin{aligned}\int_\Omega |f_n - f| d\mu &= \int_{E-B_n} |f_n - f| d\mu + \int_{B_n} |f_n - f| d\mu \\ &\leq \epsilon\mu(E) + 2M\mu(B_n)\end{aligned}$$

As $n \to \infty$, $\mu(B_n) \to 0$ and since ϵ is arbitrary, we have convergence in L^1 and we thus have the result. □

4. Fatou's Lemma

Theorem 2.4. *Suppose $f_n \geq 0$ then* $\liminf \int f_n d\mu \geq \int \liminf f_n d\mu$.

Proof. Let $g_n = \inf_{k\geq n} f_k$ and let $g = \lim_{n\to\infty} g_n$. Now we use σ-finiteness, truncation and **Lemma 2.8** to conclude:

(a) Let $E_m \uparrow \Omega$ such that $\mu(E_M) < \infty$

(b) Then, by bounded convergence theorem,

$$\int_{E_m} g \wedge m d\mu = \lim_{n\to\infty} \int_{E_m} g_n \wedge m d\mu \leq \int g_n d\mu$$

(c) By the reference lemma, we have that:

$$\lim_{m\to\infty} \int_{E_m} g \wedge m d\mu = \int g d\mu$$

□

5. Monotone Convergence Theorem

   Theorem 2.5. *If $f_n \geq 0$ and $f_n \uparrow f$ a.e. then $\int f_n d\mu \uparrow \int f d\mu$.*

   Proof. By monotonicity, $\int f_n d\mu \leq \int f_{n+1} \leq \int f d\mu$ for all $n \geq 1$. By Fatou's lemma, $\int f d\mu \leq \liminf_{n\to\infty} \int f_n d\mu$. The inequalities together imply the result. □

6. Dominated Convergence Theorem

   Theorem 2.6. *If $f_n \to f$ a.e., $|f_n| \leq g$ a.e. for all n and g is integrable then:*

   $$\int f_n d\mu \to \int f d\mu$$

   and the $f_n \to f$ in L^1.

   Proof. By assumption, $0 \leq 2g - |f - f_n| \leq 2g$. Using Fatou's lemma:

   $$\begin{aligned}\int 2g &\leq \liminf \int 2g - |f - f_n| \\ &\leq \int 2g - \limsup \int |f - f_n| d\mu\end{aligned}$$

   Therefore, $\limsup_n \int |f - f_n| d\mu = 0$. So we have L^1 convergence which implies the result. □

2.5 Expectation

1. Expected Value. Existence of Expectation.

   Definition 2.8. *Let X be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$.*

   (a) The Expected Value of X is $\mathbb{E}[X] = \int X d\mathbb{P}$

   (b) The expected value of X exists if either $\mathbb{E}[X^+] < \infty$ and/or $\mathbb{E}[X^-] < \infty$.

2. Chebyshev's General Inequality.

   Lemma 2.10. *Suppose $\phi : \mathbb{R} \to \mathbb{R}_{\geq 0}$. Let $A \in \mathcal{B}$ and $i_A = \inf\{\phi(y) : y \in A\}$. Let X be a random variable:*

   $$i_A \mathbb{P}[X \in A] \leq \mathbb{E}[\phi(X)\mathbf{1}\{X \in A\}] \leq \mathbb{E}[\phi(X)]$$

   Proof. By definition, $i_A \mathbf{1}\{X \in A\} \leq \phi(X)\mathbf{1}\{X \in A\} \leq \phi(X)$. Take expectations. □

3. An important limit theorem

Theorem 2.7. *Suppose $X_n \to X$ a.s.. Let g, h be continuous functions such that:*

(a) $g \geq 0$ and $g(x) \to \infty$ as $|x| \to \infty$
(b) $|h(x)|/g(x) \to 0$ as $|x| \to \infty$
(c) $\mathbb{E}[g(X_n)] \leq K < \infty$ for all n.

Then $\mathbb{E}[h(X_n)] \to \mathbb{E}[h(X)]$

Proof. Let $M > 0$ such that $\mathbb{P}[|X| = M] = 0$ and for $|x| > M$, $|h(x)| < \epsilon g(x)$.

(a) Truncate. Define $\bar{Y} = Y\mathbf{1}\{|Y| \leq M\}$. Then by bounded convergence theorem, $\mathbb{E}\left[h(\bar{X}_n)\right] \to \mathbb{E}\left[h(\bar{X})\right]$. Let $G = \{|X| > M\}$ and $G_n = \{|X_n| > M\}$.
(b) Estimation. Then for sufficiently large n:

$$\begin{aligned}
|\mathbb{E}[h(X)] - \mathbb{E}[h(X_n)]| &\leq \mathbb{E}\left[|h(X) - h(\bar{X})|\right] + \mathbb{E}\left[|h(X_n) - h(\bar{X}_n)|\right] \\
&\quad + |\mathbb{E}\left[h(\bar{X}_n)\right] - \mathbb{E}\left[h(\bar{X})\right]| \\
&\leq \int_G |h(X)| + \int_G |h(X_n)| + \epsilon \\
&\leq \int_G \epsilon g(X) + \int_{G_n} \epsilon g(X_n) + \epsilon \\
&\leq 2\epsilon K + \epsilon
\end{aligned}$$

□

4. An important example.

Example 2.3. *The previous theorem can be used when $g(x) = \|x\|_p$ and $h(x) = x$. This implies that if the L^p norms are uniformly bounded, and $X_n \to X$ a.s. then $\mathbb{E}[X_n] \to \mathbb{E}[X]$.*

5. Computing Expectation - Change of Variables.

Theorem 2.8. *Let X be a random variable from $(\Omega, \mathcal{F}, \mathbb{P})$ into $(S, \mathcal{S})$. Let μ be its distribution. Let $f : (S, \mathcal{S}) \to (\mathbb{R}, \mathcal{B})$ such that $f \geq 0$ or $\mathbb{E}[|f(X)|] < \infty$. Then:*

$$\mathbb{E}[f(X)] = \int_S f(y)\mu(dy)$$

Proof. Let $A \in \mathcal{S}$. Then:

$$\mathbb{E}[\mathbf{1}\{X \in A\}] = \mathbb{P}[X \in A] = \mu(A) = \int_S \mathbf{1}\{s \in A\}\, d\mu$$

By linearity, this extends to simple functions. By monotone convergence, this extends to all non-negative functions f. And given the integrability assumptions, we can use the positive and negative parts to finish the conclusion. □

3 Product Measures and Fubini's Theorem

1. Rectangles. Product σ-algebra.

   Definition 3.1. *Let $(X, \mathcal{X}, \mu)$ and $(Y, \mathcal{Y}, \nu)$ be σ-finite measure spaces. Let $\Omega = X \times Y$ and $\mathcal{S} = \{A \times B \in \mathcal{X} \times \mathcal{Y}\}$. The elements of the (semi-algebra) $\mathcal{S}$ are called rectangles and $\mathcal{F} = \sigma(\mathcal{S})$ is the Product σ-algebra.*

2. Existence and Uniqueness of Product Measure

   Theorem 3.1. *There is a unique measure π on $(\Omega, \mathcal{F})$ such that $\pi(A \times B) = \mu(A)\nu(B)$.*

   Proof. This is an application of the Caratheodory Extension theorem. To satisfy the hypotheses of this result, we must show that π is a pre-measure on $(\Omega, \mathcal{S})$. Let $\pi(A \times B) = \mu(A)\nu(B)$ and suppose:

$$A \times B = \oplus_{i \in \mathbb{N}} (A_i \times B_i)$$

   Let $I(x) = \{i : x \in A_i\}$ and note that since these are rectangles that $B = \cup_{i \in I(x)} B_i$. Then, by monotone convergence theorem:

$$\begin{aligned}
\mathbb{E}\left[\mathbf{1}\left\{A\right\}(x)\nu(B)\right] &= \mathbb{E}\left[\sum_{i \in I(x)} \mathbf{1}\left\{A_i\right\}(x)\nu(B_i)\right] \\
&= \mathbb{E}\left[\sum_{i \in \mathbb{N}} \mathbf{1}\left\{A_i\right\}(x)\nu(B_i)\right] \\
&= \sum_{i \in \mathbb{N}} \mathbb{E}\left[\mathbf{1}\left\{A_i\right\}(x)\right]\nu(B_i) \\
&= \sum_i \mu(A_i)\nu(B_i) \\
&= \sum_i \pi(A_i \times B_i)
\end{aligned}$$

   Hence, by the extension theorem a unique measure exists on the algebra generated by $\mathcal{S}$. Since both μ and ν are σ-finite, then this measure extends uniquely to $\mathcal{F}$. □

3. Fubini's Theorem

   Theorem 3.2. *Let $(\Omega, \mathcal{F}, \pi)$ be the product measure space by σ-finite measure spaces with measure μ and ν. If $f \geq 0$ or $\int_\Omega |f| d\pi < \infty$ then:*

$$\int_X \int_Y f(x,y) d\nu(y) d\mu(x) = \int_\Omega f(z) d\pi(z) = \int_Y \int_X f(x,y) d\mu(x) d\nu(y)$$

   Proof.

(a) Let $E \subset \Omega$ and define $E_x = (Y \times \{x\}) \cap E$. If $E \in \mathcal{F}$ then $E_x \in \mathcal{Y}$
Clearly, this is satisfied for rectangles. Let $\mathcal{E} = \{E : E_x \in \mathcal{Y}\}$. Then, this is satisfied for compliments and countable unions. Hence, $\mathcal{F} \subset \mathcal{E}$.

(b) If $E \in \mathcal{F}$ then $g(x) = \nu(E_x)$ is $\mathcal{X}$ measurable and $\int_X g(x)d\mu = \pi(E)$.
Let $\mathcal{L}$ be the collection for which this holds, and we can assume that $\mu(X)$ or $\nu(Y)$ are finite and use σ-finiteness to get the result. We show that $\mathcal{L}$ is a λ-system.

i. If $E = \Omega$ then $E_x = Y$. Hence:

$$\int_X \nu(Y)d\mu(x) = \nu(Y)\mu(X) = \pi(\Omega)$$

ii. If $E \in \mathcal{L}$ then:

$$\pi(E^c) = \int_X (\nu(Y) - \nu(E_x))\, d\mu(x) = \int_X \nu(E_x^c)d\mu(x)$$

iii. If $E^i \in \mathcal{L}$ are disjoint, by monotone convergence:

$$\pi(\cup_i E^i) = \sum_i \pi(E_i) = \int_X \nu\left((\cup_i E_i)_x\right) d\mu(x)$$

Since $\mathcal{S} \subset \mathcal{L}$, $\mathcal{F} \subset \mathcal{L}$ by $\pi - \lambda$ theorem.

(c) Use standard machinery to build up to simple functions, positive functions and then integrable functions.

□

4 Independence

4.1 Measure-Theoretic Definition and Properties

1. Independent σ-algebras, independent random variables, and independent events.

Definition 4.1.

(a) Let $\mathcal{F}_1, \ldots, \mathcal{F}_n$ be σ-algebras. These are independent σ-algebras if for any $A_i \in \mathcal{F}_i$ and $I \subset \{1, \ldots, n\}$:

$$\mathbb{P}\left[\cap_{i \in I} A_i\right] = \prod_{i \in I} \mathbb{P}\left[A_i\right]$$

(b) $X_1, \ldots, X_n$ are independent random variables whenever $\sigma(X_1), \ldots, \sigma(X_n)$ are independent σ-algebras.

(c) $A_1, \ldots, A_n$ are independent events whenever $\sigma(A_1), \ldots, \sigma(A_n)$ are independent.

2. Equivalent Definition

Lemma 4.1. *In the definition of independent σ-algebras, the condition that $I \subset \{1, \ldots, n\}$ is equivalent to $I = \{1, \ldots, n\}$.*

Proof. Clearly, $I = \{1, \ldots, n\}$ requires $I \subset \{1, \ldots, n\}$. Now suppose the definition requires $I = \{1, \ldots, n\}$. Choose $I \subset \{1, \ldots, n\}$. For $j \in \{1, \ldots, n\} \setminus I$, let $A_j = \Omega$. The result follows. □

4.2 Sufficient Conditions for Independence

1. Independent π-systems

   Lemma 4.2. *Let $\mathcal{P}_1, \ldots, \mathcal{P}_n$ be independent π-systems. Then, the σ-algebras they generate are independent.*

   Proof. Let $F \in \mathcal{P}_2 \times \ldots \times \mathcal{P}_n$. Let $\mathcal{L} = \{A : \mathbb{P}[A \cap F] = \mathbb{P}[A]\mathbb{P}[F], \forall F\{$. Note that $\mathcal{P}_1 \subset \mathcal{L}$. Now, we show that $\mathcal{L}$ is a λ-system.

   (a) $\Omega \in \mathcal{L}$
   (b) If $A \in \mathcal{L}$. Then $\mathbb{P}[A^c \cap F] = \mathbb{P}[F] - \mathbb{P}[A \cap F] = \mathbb{P}[A^c]\mathbb{P}[F]$
   (c) If $A_i \in \mathcal{L}$ are disjoint, then

   $$\mathbb{P}\left[\left(\bigcup_i A_i\right) \cap F\right] = \sum_i \mathbb{P}[A_i]\mathbb{P}[F] = \mathbb{P}[F]\mathbb{P}\left[\bigcup_i A_i\right]$$

   By $\pi - \lambda$ theorem, we have that $\sigma(\mathcal{P}_1), \mathcal{P}_2, \ldots, \mathcal{P}_n$ are independent. Now we simply repeat the proof for each i with

   $$F \in \sigma(\mathcal{P}_1) \times \cdots \times \mathcal{P}_{i-1} \times \mathcal{P}_{i+1} \times \cdots \times \mathcal{P}_n$$

   □

2. Functions of Independent Random Variables

   Theorem 4.1. *Let $\mathcal{F}_{i,j}$ for $1 \leq i \leq n$ and $1 \leq j \leq m(i)$ be independent. Let $\mathcal{G}_i = \sigma(\times_j \mathcal{F}_{i,j})$. Then $\mathcal{G}_1, \ldots, \mathcal{G}_n$ are independent.*

   Proof. Let $\mathcal{S}_i$ be the rectangles generated by $\mathcal{F}_{i,j}$ for $1 \leq j \leq m(i)$. These are π-systems and are independent. By the previous lemma, the σ-algebras they generate are independent. □

   Corollary 4.1. *Let $X_{i,j}$ be independent random variables for $1 \leq i \leq n$ and $1 \leq j \leq m(i)$. Let f_i be measurable functions on $\mathbb{R}^{m(i)}$. Then $f_i(X_{i,1}, \ldots, X_{m(i),i})$ are independent.*

4.3 Independent, Distributions and Expectation

1. Independence and Distributions

   Theorem 4.2. *Let $X_1, \ldots, X_n$ be independent random variables such that $X_i \sim \mu_i$. Then:*

   $$(X_1, \ldots, X_n) \sim \mu_1 \times \cdots \mu_n$$

Proof. Let $A = A_1 \times \cdots \times A_n \in \mathcal{B}^n$. Then:

$$\begin{aligned}
\mathbb{P}\left[(X_1, \ldots, X_n) \in A\right] &= \mathbb{P}\left[X_1 \in A_1, \ldots, X_n \in A_n\right] \\
&= \prod_{i=1}^{n} \mathbb{P}\left[X_i \in A_i\right] \\
&= \prod_{i=1}^{n} \mu_i(A_i) \\
&= \mu_1 \times \cdots \times \mu_n(A)
\end{aligned}$$

Hence, this holds for the π-system generated by the rectangles. By the Extension Theorem, since the product measure is unique, $\mathbb{P} = \mu_1 \times \cdots \times \mu_n$. □

2. Independence and Integration

Theorem 4.3. *Suppose X, Y are independent with distributions μ and ν.*

(a) If $h : \mathbb{R}^2 \to \mathbb{R}$ is measurable such that $h \geq 0$ and $\mathbb{E}\left[|h(X,Y)|\right] < \infty$ then:

$$\mathbb{E}\left[h(X,Y)\right] = \int \int h(x,y)\mu(dx)\nu(dy)$$

(b) In particular, if $f, g : \mathbb{R} \to \mathbb{R}$ are measurable such that either $f, g \geq 0$ or f, g are integrable then:

$$\mathbb{E}\left[f(X)g(Y)\right] = \mathbb{E}\left[f(X)\right]\mathbb{E}\left[g(Y)\right]$$

Proof. By change of variables formula, independence, and Fubini:

$$\mathbb{E}\left[h(X,Y)\right] = \int_{\Omega} h(x,y)d(\mu \times \nu)((x,y)) = \int \int h(x,y)d\mu(x)d\nu(y)$$

If $h(x,y) = f(x)g(y)$ then:

$$\mathbb{E}\left[h(X,Y)\right] = \int \int h(x,y)d\mu(x)d\nu(y) = \int g(y) \int f(x)d\mu(x)d\nu(y)$$

The result follows. □

4.4 Sums of Independent Random Variables

1. Distribution Function of Sums

Theorem 4.4. *If X, Y are independent with distribution functions F and G respectively, then:*

$$\mathbb{P}\left[X + Y \leq z\right] = \int_{y \in \mathbb{R}} F(z-y)dG(y)$$

Proof. Let μ and ν be probability measures such that $X \sim \mu$ and $Y \sim \nu$. Then by change of variables, independence and Fubini's theorem:

$$\begin{aligned}
\mathbb{P}\left[X+Y \leq z\right] &= \int_{X+Y\leq z} d\mathbb{P} \\
&= \int_{x+y\leq z} d(\mu \times \nu)((x,y)) \\
&= \int_{y\in\mathbb{R}} \int_{-\infty}^{z-y} d\mu(x) d\nu(y) \\
&= \int_{y\in\mathbb{R}} F(z-y) dG(y)
\end{aligned}$$

□

2. Simple change of measure

Lemma 4.3. *If μ is a probability measure such that for some Lebesgue measurable function f, $\mu(A) = \int_A f(x)dx$ then for any measurable function g non-negative or integrable:*

$$\int g d\mu = \int g(x)f(x)dx$$

Proof. Let A be a measurable set. Then:

$$\int \mathbf{1}\left\{A\right\} d\mu = \mu(A) = \int \mathbf{1}\left\{A\right\} f(x)dx$$

Using linear combinations, we have simple functions. Then using monotone convergence, we will have non-negative functions. Using positive and negative parts will give us integrable functions. □

3. Density of Sums

Theorem 4.5. *Suppose that X has density f and Y has distribution function G and both are independent.*

(a) $X+Y$ has density $h(x) = \int f(x-y)dG(y)$

(b) If Y admits a density then $X+Y$ has density $h(x) = \int f(x-y)g(y)dy$

Proof. We have from Fubini:

$$H(x) = \int F(x-y)dG(y) = \int \int_{-\infty}^{x} f(z-y)dzdG(y) = \int_{-\infty}^{x} \int f(z-y)dG(y)dz$$

Hence, $X+Y$ admits a density, which is given by $\int f(z-y)dG(y)$. Secondly, if G admits a derivative, then we can just apply the previous lemma. □

Part II
Laws of Large Numbers

5 Weak Laws of Large Numbers

5.1 L^2 Weak Laws

1. Uncorrelated Random Variables

   Lemma 5.1. *Let $X_1, \dots, X_n$ have $\mathbb{E}\left[|X_i|^2\right] < \infty$ and be uncorrelated. Then $\mathbb{V}\left[\sum_i X_i\right] = \sum_i \mathbb{V}\left[X_i\right]$.*

   Proof. Assume W.L.O.G. that $\mathbb{E}\left[X_i\right] = 0$. Then:

   $$\mathbb{V}\left[\sum_i X_i\right] = \sum_i \mathbb{V}\left[X_i\right] + 2\sum_{i<j} \mathbb{E}\left[X_i X_j\right] = \sum_i \mathbb{V}\left[X_i\right]$$

   □

2. Convergence in L^p implies convergence in measure

   Lemma 5.2. *Fix $p > 0$. If $\mathbb{E}\left[|X_n|^p\right] \to 0$ then $X_n \to 0$ in P.*

   Proof. Use Chebyshev's inequality with $\phi(x) = |x|^p$. □

3. Weak Law in L^2

   Theorem 5.1. *Let $X_1, \dots$ be uncorrelated with $\mathbb{E}\left[X_i\right] = 0$ and $\mathbb{V}\left[X_i\right] \leq C < \infty$. Then $\frac{S_n}{n} \to 0$ in L^2 and $\frac{S_n}{n} \to 0$ in P.*

   Proof.

   $$\frac{\mathbb{E}\left[|S_n|^2\right]^2}{n} \leq \frac{C}{n} \to 0$$

   Convergence in probability follows from Chebyshev's inequality. □

4. Polynomial Approximation

   Example 5.1. *Let f be continuous on $[0,1]$. Let*

   $$f_n(x) = \sum_{m=0}^{n} \binom{n}{m} x^m (1-x)^{n-m} f(m/n)$$

   Then $\sup_{x\in[0,1]} |f_n(x) - f(x)| \to 0$.

Proof. Let $X_1, \dots, X_n$ be bernoulli with probability p. Then:

$$\begin{aligned} f_n(p) &= \sum_{m=0}^{n} \binom{n}{m} p^m (1-p)^{n-m} f(m/n) \\ &= \sum_{m=0}^{n} \mathbb{P}\left[S_n = m\right] f(m/n) \\ &= \mathbb{E}\left[f(S_n/n)\right] \end{aligned}$$

Now we must show that $\mathbb{E}\left[f(S_n/n)\right] \to f(p)$.

(a) Let $\epsilon > 0$. Then $\exists \delta > 0$ such that (by uniform continuity):

$$\mathbb{P}\left[|f(S_n/n) - f(p)| > \epsilon\right] \leq \mathbb{P}\left[|S_n/n - p| > \delta\right]$$

The right hand side tends to 0 by the L^2 weak convergence theorem. Hence $f(S_n/n) \to f(p)$ in probability.

(b) To get convergence in expectation (Borel-Cantelli would be easier to use), let $G_n = \{|f(S_n/n) - f(p)| > \epsilon\}$. Then:

$$\begin{aligned} \mathbb{E}\left[|f(S_n/n) - f(p)|\right] &\leq \mathbb{E}\left[|f(S_n/n) - f(p)| \mathbf{1}\left\{G_n\right\}\right] + \epsilon \\ &\leq 2 \|f\|_\infty \mathbb{P}\left[G_n\right] + \epsilon \end{aligned}$$

Hence, the result holds for all p.

□

5.2 Triangular Arrays

1. Controlling Variance of Sum

Lemma 5.3. *Let $\mu_n = \mathbb{E}\left[X_n\right]$ and $\sigma_n^2 = \mathbb{V}\left[X_n\right]$. If $\frac{\sigma_n^2}{b_n^2} \to 0$ then*

$$\frac{X_n - \mu_n}{b_n} \to 0 \sim \mathbb{P}$$

Proof. Chebyshev.

$$\mathbb{P}\left[\left|\frac{X_n - \mu_n}{b_n}\right| > \epsilon\right] \leq \frac{\mathbb{V}\left[X_n\right]}{b_n^2} \leq \frac{\sigma_n^2}{b_n^2} \to 0$$

□

2. Coupon Collector

Example 5.2. *Suppose there are n distinct items, and we choose one item, independently uniformly with replacement. How long does it take to collect a complete set? What is the asymptotic behavior as $n \to \infty$?*

Solution. We can only choose at distinct moments in time, which we index by t. Let X_t be the number of distinct coupons collected by time t. Let $N(k) = \inf\{t : X_t = k\}$. Then, $N(n) - N(n-1), N(n-1) - N(n-2), \cdots, N(1)$ are independent random variables with a geometric distribution. The probability of success for $N(k) - N(k-1)$ is

$$\frac{n-(k-1)}{n}$$

Hence:

$$\mathbb{E}[N(n)] = \sum_{i=1}^{n} \mathbb{E}[N(i) - N(i-1)] = \sum_{i=1}^{n} \frac{n}{n-(i-1)} = n \sum_{i=1}^{n} \frac{1}{i}$$

Also:

$$\mathbb{V}[N(n)] = \sum_{i=1}^{n} \mathbb{V}[N(i) - N(i-1)] = n^2 \sum_{k=1}^{n} \frac{1}{k^2}$$

Therefore, by the previous lemma:

$$\frac{N(n) - n\sum_{k=1}^{n} \frac{1}{k}}{n \log n} \to 0 \sim \mathbb{P}$$

That is:

$$\frac{N(n)}{n \log n} \to 1 \sim \mathbb{P}$$

□

5.3 Truncation

1. General Theorem Weak Law for Triangular Arrays

   Theorem 5.2. *For each n, let $X_{n,k}$, $1 \leq k \leq n$ be independent. Let $b_n > 0$ such that $b_n \to \infty$, and let $\bar{X}_{n,k} = \mathbf{1}\{X_{n,k}\mathbf{1}\{|Y_{n,k}| \leq b_n\}\}$. Suppose as $n \to \infty$:*

   (a) Deviation from Truncation Control:

   $$\sum_{k=1}^{n} \mathbb{P}[|X_{n,k}| \leq b_n] \to 0$$

   (b) Variance of Truncation Control:

   $$\frac{\sum_{k=1}^{n} \mathbb{E}\left[\bar{X}_{n,k}^2\right]}{b_n^2} \to 0$$

   Letting $a_n = \sum_{k=1}^{n} \mathbb{E}\left[\bar{X}_{n,k}\right]$. Then:

   $$\frac{S_n - a_n}{b_n} \to 0 \sim \mathbb{P}$$

Proof. There are two parts. First, we show that for T_n, the partial sums of the truncations, the result holds. Then, we approximate the difference between S_n and T_n

(a) By the second assumption and Chebyshev's theorem:

$$\mathbb{P}\left[\left|\frac{T_n - a_n}{b_n}\right| > \epsilon\right] \leq \frac{\sum_{k=1}^n \mathbb{E}\left[\bar{X}_{n,k}^2\right]}{b_n^2} \to 0$$

(b) By the first assumption:

$$\mathbb{P}\left[S_n \neq T_n\right] = \sum_{k=1}^n \mathbb{P}\left[|X_{n,k}| > b_n\right] \to 0$$

□

2. I.I.D. Weak Law of Large Numbers

Theorem 5.3. *Let $X_1, \ldots$ be i.i.d. such that*

$$x\mathbb{P}\left[|X_i| > x\right] \to 0 \qquad as\ x \to \infty$$

Let $\mu_n = \mathbb{E}\left[X_1 \mathbf{1}\left\{|X_1| \leq n\right\}\right]$. Then:

$$\frac{S_n}{n} - \mu_n \to 0 \sim \mathbb{P}$$

Proof. First we must control the deviance from the truncation. That is:

$$\sum_{i=1}^n \mathbb{P}\left[|X_i| \geq n\right] = n\mathbb{P}\left[|X_1| \geq n\right] \to 0$$

Now, we must control the variance. First note that:

$$\mathbb{E}\left[|\bar{X}_i|^2\right] = \int_0^n 2y\mathbb{P}\left[|X_1| \geq y\right] dy$$

Let $M = \sup_{y \geq 0} y\mathbb{P}\left[|X_1| \geq y\right]$. Let $\epsilon > 0$ and find K_ϵ such that if $y \geq K$ then:

$$y\mathbb{P}\left[|X_1| \geq y\right] < \epsilon$$

Then:

$$\mathbb{E}\left[|\bar{X}_i|^2\right] \leq 2\int_0^{K_\epsilon} M + 2\epsilon(n - K_\epsilon)$$

So to get the control over variance, we note that:

$$\frac{\sum_{k=1}^n \mathbb{E}\left[\bar{X}_k^2\right]}{n^2} = \frac{\mathbb{E}\left[X_1^2\right]}{n} \leq \frac{2K_\epsilon M + 2\epsilon(n - K_\epsilon)}{n} \to 2\epsilon$$

Since ϵ is arbitrary, we have to only apply the Weak Law for Triangular Arrays. □

3. I.I.D. Integrable Weak Law of Large Numbers

Theorem 5.4. *Let $X_1, \ldots$ be i.i.d. integrable random variables with $\mu = \mathbb{E}[X_1]$. Then:*

$$\frac{S_n}{n} \to \mu \sim \mathbb{P}$$

Proof. We simply must show that $x\mathbb{P}[|X_1| \geq x] \to 0$ as $x \to \infty$. Since we are on a probability measure, $\mathbb{P}[|X_1| \geq x] \to 0$ as $x \to \infty$. Since X_1 is integrable, letting $\epsilon > 0$ we can find a y such that if $x > y$:

$$x\mathbb{P}[|X_1| \geq x] \leq \mathbb{E}[|X_1|\mathbf{1}\{|X_1| \geq x\}] < \epsilon$$

Hence, we have satisfied the first condition. To show that $\mu_n \to \mu$, we have that:

$$\lim_{n\to\infty} |\mu - \mathbb{E}[X_1\mathbf{1}\{|X_1| \leq n\}]| = |\mu - \mathbb{E}\left[\lim_{n\to\infty} X_1\mathbf{1}\{|X_1| \leq n\}\right] \to 0$$

□

6 Strong Laws of Large Numbers

6.1 Borel-Cantelli Lemmas: from Weak to Strong

1. Infinitely Often. Almost Always.

Definition 6.1. *Let A_n be a sequence of measurable subsets of a probability space.*

(a) ω occurs infinitely often in A_n if

$$\omega \in \{A_n \ i.o\} = \limsup_n A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m$$

(b) ω occurs almost always in A_n if

$$\omega \in \{A_n \ a.a.\} = \liminf_n A_n = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m$$

Note 6.1. *We can interpret these both as follows. ω occurs infinitely often means that no matter what N is chosen, there is an $m \geq N$ such that $\omega \in A_m$. If ω does not occur infinitely often, it will occur only in finitely many A_m. If ω occurs almost always that means that after some N, $\omega \in A_m$ for every $m \geq N$.*

6.1.1 First Borel-Cantelli Lemma

2. First Borel-Cantelli Lemma

   Lemma 6.1. *If* $\sum_{k=1}^{\infty} \mathbb{P}[A_k] < \infty$ *then* $\mathbb{P}[A_n \ i.o.] = 0$.

   Proof. This is a consequence of monotonic limits of measures:

   $$\begin{aligned} \mathbb{P}[A_n \ i.o.] &= \lim_{n\to\infty} \mathbb{P}\left[\bigcup_{m\geq n} A_m\right] \\ &\leq \lim_{n\to\infty} \sum_{m=n}^{\infty} \mathbb{P}[A_m] \\ &= 0 \end{aligned}$$

   □

3. Counter-Example for Converse.

   Example 6.1. *Consider* $A_n = [0, 1/n]$ *with Lebesgue measure on* $[0,1]$. *Here we have that* $\mathbb{P}[A_n \ i.o.] = \mathbb{P}[0] = 0$. *However,* $\sum_{n=1}^{\infty} \mathbb{P}[A_n] = \sum_{n=1}^{\infty} 1/n = \infty$.

4. Applications

   (a) Convergence in probability and sub(sub)sequence convergence almost everywhere

   Theorem 6.1. $X_n \to X \sim \mathbb{P}$ *if and only if for every subsequence indexed by* $n(m)$ *there is a subsequence for which* $X_{n(m,k)} \to X \sim a.s.$

   Proof. Suppose $X_n \to X$ in probability. Let $X_n(m)$ be an subsequence. We can then find a subsequence indexed by $n(m,k)$ such that for $\epsilon_k \downarrow 0$:

   $$\mathbb{P}\left[|X_{n(m,k)} - X| > \epsilon_k\right] < \frac{1}{2^k}$$

   By Borel-Canteilli, the probability that these events occur infinitely often is 0. Therefore, for almost every $\omega \in \Omega$, $X_{n(m,k)}(\omega) \to X(\omega)$.

   For the other direction, suppose $X_n \not\to X$ in probability. Then there is a ball about X of radius ϵ and a subsequence of X_n indexed by $n(m)$ which remains outside of this ball with probability 1. However, we can extract a subsequence $X_n(m,k)$ which converges to X almost surely, which is a contradiction. □

   (b) Continuous functions of sequences converging in Probability

   Theorem 6.2. *If* f *is continuous and* $X_n \to X$ *in probability then* $f(X_n) \to f(X)$ *in probability. Also, if* f *is bounded then* $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$.

Proof. Select any subsequence of X_n and extract a further subsequence which converges to X almost surely. Then by continuity, $f(X_{n(m,k)})$ converges to $f(X)$ almost surely. Hence, by the previous theorem, we have the first result. For the second result, by the dominated convergence theorem, $\mathbb{E}\left[f(X_{n(m,k)})\right] \to \mathbb{E}\left[f(X)\right]$. For a contradiction, we can suppose a subsequence of $f(X_n)$ does not converge in expectation to $f(X)$, but we can find a subsequence which does of this subsequence which does. □

(c) Fatou's Lemma for Probability

Theorem 6.3. *Suppose $X_n \geq 0$ and $X_n \to X$ in probability. Then:*

$$\liminf_{n\to\infty} \mathbb{E}\left[X_n\right] \geq \mathbb{E}\left[X\right]$$

Proof. Take sub-subsequences of X_n on which almost sure convergence holds. Then apply Fatou's lemma. Then, argue by contradiction on subsequences that the result holds. □

(d) Dominated Convergence for Probability

Theorem 6.4. *Suppose $|X_n|$ is dominated by a random variable Y which is integrable, and $X_n \to X$ in probability. Then, $\mathbb{E}\left[X_n\right] \to \mathbb{E}\left[X\right]$.*

Proof. Go to sub-subsequences and use the dominated convergence theorem for almost surely converging sequences. Then, by contradiction on sub-subsequences, we have that $\mathbb{E}\left[X_n\right] \to \mathbb{E}\left[X\right]$ □

(e) Fourth-Moment Bound Strong Law of Large Numbers

Theorem 6.5. *Let $X_1, \dots$ be independent mean 0 random variables such that $\mathbb{E}\left[X_i^4\right] \leq C < \infty$. Then:*

$$\frac{S_n}{n} \to 0 \sim a.s.$$

Proof. This will be a consequence of Chebyshev, and because there is the fourth moment, we will be able to apply borel cantelli. So:

$$\mathbb{P}\left[\left|\frac{S_n}{n}\right| > \epsilon_n\right] \leq \frac{\mathbb{E}\left[S_n^4\right]}{n^4\epsilon_n^4}$$

Note that $(x_1 + \cdots + x_n)^4$ will have:

i. n terms of the form x_i^4
ii. $\binom{n}{2}$ ways of choosing (i,j) for the term $x_i^2x_j^2$ and $\frac{4!}{2!2!}$ ways of assigning the exponent. Hence, there are $3n(n-1)$ such terms.
iii. $\binom{n}{3}$ ways of choosing (i,j,k) and $\frac{4!}{2!1!1!}$ ways of assigning the exponent.

iv. $\binom{n}{4}$ ways of choosing (i,j,k,l) and 4! ways of assigning the exponent.

Since the latter two terms will be 0 because of independence, we have:

$$\frac{\mathbb{E}\left[S_n^4\right]}{n^4\epsilon_n^4} \leq \frac{nC+3n(n-1)C^2}{\epsilon_n^4 n^4} = \mathcal{O}(n^{-2})$$

Now we can choose $\epsilon_n \downarrow 0$ such that $\epsilon_n = n^{-1-\delta}$ where $0 < \delta < 1$. Then:

$$\sum_{n=1}^{\infty} \mathbb{P}\left[\left|\frac{S_n}{n}\right| > \epsilon_n\right] < \infty$$

By Borel-Cantelli, we have that these events occur infinitely often with probability 0. Hence, we have almost sure convergence. □

6.1.2 Second Borel-Cantelli Lemma

5. Second Borel-Cantelli Lemma

Lemma 6.2. *Suppose A_n are independent events and $\sum_{n=1}^{\infty} \mathbb{P}[A_n] = \infty$. Then, $\mathbb{P}[A_n \text{ i.o.}] = 1$.*

Proof. Let $M > N$. Then:

$$\mathbb{P}\left[\cap_{n=N}^{M} A_n^c\right] = \prod_{n=N}^{M}(1-\mathbb{P}[A_n]) \leq \exp\left(-\sum_{n=N}^{M}\mathbb{P}[A_n]\right)$$

Taking the limit over M, we see that $\mathbb{P}\left[\bigcup_{n\geq N} A_n\right] = 1$. Thus, we have the result. □

6. Application: Existence of Limit

Theorem 6.6. *If $X_1, \ldots$ are i.i.d. with $\mathbb{E}[|X_i|] = \infty$ then*

$$\mathbb{P}[|X_n| \geq n \text{ i.o}] = 1$$

Also:

$$\mathbb{P}[\lim S_n/n \text{ exists in } (-\infty, \infty)] = 0$$

Proof. For the first point. Note that:

$$\mathbb{E}[|X_i|] = \int_0^{\infty} \mathbb{P}[|X_i| \geq x]\,dx \leq \sum_{n=0}^{\infty} \mathbb{P}[|X_i| \geq n]$$

Applying the Second Borel-Cantelli Lemma gives the result. For the second result, we have that:

$$\frac{S_n}{n} - \frac{S_{n+1}}{n+1} = \frac{S_n}{n(n+1)} - \frac{X_{n+1}}{n+1}$$

Let C be the set on which the limit converges. Then on $C \cap \{|X_n| \geq n \ i.o\}$:

$$\left| \frac{S_n}{n} - \frac{S_{n+1}}{n+1} \right| > \frac{2}{3} \ i.o.$$

So we have a contradiction. □

7. Extension.

Theorem 6.7. *If $A_1, \ldots$ are pairwise independent events and $\sum_{n=1}^{\infty} \mathbb{P}[A_n] = \infty$ then*

$$\frac{\sum_{m=1}^{n} \mathbf{1}\{A_m\}}{\sum_{m=1}^{n} \mathbb{P}[A_m]} \to 1 \sim a.s.$$

6.2 Etemadi's Proof of SLLN

1. Summation Bounding Lemma

Lemma 6.3. *Let $y > 0$. Then $2y \sum_{k>y} \frac{1}{k^2} \leq 4$.*

Proof. If $y > 1$ then

$$\begin{aligned} 2y \sum_{k>y} \frac{1}{k^2} &\leq 2y \sum_{k>\lceil y \rceil} \frac{1}{k^2} \\ &\leq 2y \int_{\lfloor y \rfloor}^{\infty} \frac{1}{z^2} \\ &\leq \frac{2y}{\lfloor y \rfloor} \\ &\leq 4 \end{aligned}$$

If $0 < y \leq 1$ then

$$\begin{aligned} 2y \sum_{k>y} \frac{1}{k^2} &\leq 2 + 2 \sum_{k \geq 2} \frac{1}{k^2} \\ &\leq 2 + 2 \int_{1}^{\infty} \frac{1}{z^2} \\ &\leq 4 \end{aligned}$$

When $y = 0$, the entire term is 0. □

2. Bounding Variance

Lemma 6.4. *Let $X_1, \ldots$ be identically distributed, integrable random variables. Then:*

$$\sum_{k=1}^{\infty} \frac{\mathbb{V}\left[X_k^2\right]}{k^2} \leq 4\mathbb{E}\left[|X_i|\right]$$

Proof. Bounding variance above by the second moment, and applying the previous lemma:

$$\sum_{k=1}^{\infty} \frac{1}{k^2} \int_0^{\infty} 2x\mathbb{P}\left[|X_1| \geq x\right] dx \leq \int_0^{\infty} \mathbb{P}\left[|X_1| \geq x\right] 2x \sum_{k=1} \frac{1}{k^2} dx$$
$$\leq 4 \int \mathbb{P}\left[|X_1| \geq x\right]$$
$$\leq 4\mathbb{E}\left[|X_1|\right]$$

□

3. Pairwise Independent SLLN

Theorem 6.8. *Let $X_1, \ldots$ be pairwise independent identically distributed random variables such that $\mathbb{E}\left[|X_i|\right] < \infty$. Let $\mathbb{E}\left[X_i\right] = 0$. Then $\frac{S_n}{n} \to 0$ a.s..*

Proof. First, the hypotheses are satisfies by the positive and negative parts of X_i, so we need only consider the positive case. As per usual, we truncate the random variable. Use convergence convergence in probability with Borel-Cantelli to get the convergence of subsequences a.s.. We then sandwich the remaining terms between the subsequences.

(a) Truncation. Let $Y_k = X_k \mathbf{1}\left\{|X_k| \leq k\right\}$. Then Y_k are pairwise independent. Moreover:

$$\sum_{k=1}^{\infty} \mathbb{P}\left[X_k \neq Y_k\right] = \sum_{k=1}^{\infty} \int_{|X_k| \geq k} d\mathbb{P} \leq \mathbb{E}\left[|X_1|\right]$$

Hence, by Borel Cantelli, $X_k \neq Y_k$ for almost every ω at most finitely many times. Letting T_n be the partial sums of Y_k, for a.e. $\omega \in \Omega$:

$$|S_n(\omega) - T_n(\omega)| \leq R(\omega) < \infty$$

Hence, in the limit $\frac{S_n - T_n}{n} \to 0$ a.s.

Convergence in Probability of Subsequence. Let $\epsilon > 0$ and $\alpha > 1$. Let $k(n) = \lceil \alpha^n \rceil$. We consider the behavior of the subsequence $T_{k(n)}$.

$$\sum_{n=1}^{\infty} \mathbb{P}\left[\left|T_{k(n)} - \mathbb{E}\left[T_{k(n)}\right]\right| > \epsilon n(k)\right] \leq \sum_{n=1}^{\infty} \frac{\mathbb{V}\left[T_{k(n)}\right]}{\epsilon^2 k(n)^2}$$
$$\leq \sum_{m=1}^{\infty} \frac{\mathbb{V}\left[Y_m\right]}{\epsilon^2} \sum_{k(n) \geq m} \frac{1}{\alpha^{2n}}$$
$$\leq \sum_{m=1}^{\infty} \frac{\mathbb{V}\left[Y_m\right]}{\epsilon^2} \frac{1}{m^2} \frac{1}{1 - \alpha^{-2}}$$
$$\leq \frac{4\mathbb{E}\left[|X_1|\right]}{\epsilon^2 (1 - \alpha^{-2})}$$

By Borel Cantelli, these events happen only finitely often a.s. Since ϵ is arbitrary, and by dominated convergence theorem:

$$\frac{T_{k(n)}}{k(n)} \to 0 \ a.s.$$

Sandwiching. Let $k(n) \leq m \leq k(n+1)$. $T_{k(n)}m \leq T_m k(n+1)$ and $T_m k(n) \leq T_{k(n+1)} m$ a.s. Therefore:

$$\frac{T_{k(n)}}{k(n+1)} \leq \frac{T_m}{m} \leq \frac{T_{k(n+1)}}{k(n)}$$

Since the limit of $T_{k(n)}/T_{k(n+1)} = \frac{1}{\alpha}$, and $\alpha > 1$ is arbitrary, the result follows. □

4. Unbounded Positive Part SLLN

Theorem 6.9. *Let $X_1, \dots$ be i.i.d. such that $\mathbb{E}\left[X_i^+\right] = \infty$ and $\mathbb{E}\left[X_i^-\right] < \infty$. Then $\frac{S_n}{n} \to \infty$ a.s.*

Proof. Let $X_i^m = X_i^+ \mathbf{1}\left\{X_i^+ \leq m\right\} - X_i^-$. Then, X_i^m are i.i.d. and integrable. By the SLLN:

$$\frac{S_n^m}{n} \to \mathbb{E}\left[X_1^m\right]$$

Since $\frac{S_n}{n} \geq \frac{S_n^m}{n}$ and X_i^m are monotonically increasing, then by the monotone convergence theorem, the result follows. □

5. Application to Renewal Theory

Theorem 6.10. *Let $X_1, \dots$ be i.i.d. with $0 < X_i < \infty$. Let $T_n = \sum_{i=1}^n X_i$ and let $N_t = \sup\{n : T_n \leq t\}$. If $\mathbb{E}\left[X_1\right] = \mu \leq \infty$, as $t \to \infty$:*

$$\frac{N_t}{t} \to \frac{1}{\mu} \ a.s.$$

Proof. Let $\Omega_0 = \{T_n/n \to \mu\}$. Since $T_n < \infty$ for all n, then as $t \to \infty$, $N_t \to \infty$. Hence:

$$\frac{T_{N_t(\omega)}}{N_t(\omega)} \to \mu \qquad \frac{N_t(\omega)+1}{N_t(\omega)} \to 1$$

Finally:

$$\frac{T_{N_t}}{N_t} \leq \frac{t}{N_t} \leq \frac{T_{N_t+1}}{N_t+1}\frac{N_t+1}{N_t}$$

□

6.3 Random Series Approach

1. The results above are approached using a similar set of techniques:
   (a) Truncating the random variable and controlling truncation error.
   (b) Applying Chebyshev's Inequality to get convergence in probability of the sequence or subsequence of S_n. Achieved by some type of variance control.
   (c) Applying Bore-Cantelli's First Lemma to a subsequence to demonstrate almost sure convergence.
   (d) Sandwiching the original sequence between these terms to demonstrate almost sure convergence.
2. In the following treatment, there may be no need for Truncation and the last three steps can be replaced by controlling the series using a maximal inequality (which is derived using Chebyshev).
3. Tail σ-algebras
   (a) Tail σ-algebra

   Definition 6.2. *Let $\mathcal{F}'_n = \sigma(X_n, X_{n+1}, \ldots)$. Then the tail σ-algebra is $\mathcal{T} = \bigcap_n \mathcal{F}'_n$*

   (b) Examples

   Example 6.2. *If $B_n \in \mathcal{B}$ then $\{X_n \in B_n \ i.o\} \in \mathcal{T}$. This holds since:*

$$\bigcup_{m \geq n} \{X_m \in B_m\} \in \mathcal{F}'_n$$

   Example 6.3. *Let $S_n = X_1 + \cdots + X_n$. Then:*

   i. $\{\lim S_n$ exists $\} \in \mathcal{T}$. The existence of the limit does not depend on a finite number of initial random variables, hence, the event is in $\mathcal{F}'_n$ for all n.

   ii. $\{\limsup S_n > 0\} \notin \mathcal{T}$. The sum always depends on the initial $X_1, \ldots, X_{n-1}$, hence the event cannot be a tail event.

   iii. $\{\limsup S_n > ab_n\} \in \mathcal{T}$ for $b_n \to \infty$. This holds since for sufficiently large n, $X_1, \ldots, X_{n-1}$ will have little effect on the event since their contribution is diminished by b_n.

   (c) Kolmogorov's 0-1 Law

   Theorem 6.11. *Let $X_1, \ldots$ be independent random variables. If A is a tail event in the tail σ-algebra generated by these random variables then $\mathbb{P}[A] \in \{0, 1\}$.*

   Proof. There are two parts

   i. Let $k > 0$. Then, $\sigma(X_1, \ldots, X_{k-1})$ is independent of $\sigma(X_k, \ldots, X_{k+M})$. Then, $\sigma(X_1, \ldots, X_{k-1})$ and $\bigcup_M \sigma(X_k, \ldots, X_{k+M})$ are independent π-systems. Therefore, $\sigma(X_1, \ldots, X_{k-1})$ and $\mathcal{F}'_k$ are independent.

ii. Let $k > 0$. Then, by the first point, $\sigma(X_1, \ldots, X_{k-1})$ is independent of $\mathcal{T} \subset \mathcal{F}'_k$. Hence, $\bigcup_k \sigma(X_1, \ldots, X_{k-1})$ and $\mathcal{T}$ are independent π systems. Therefore, $\sigma(X_1, \ldots)$ is independent of $\mathcal{T}$.

iii. So if $\mathcal{A} \in \mathcal{T}$ then $A \in \sigma(X_1, \ldots)$. Therefore:

$$\mathbb{P}[A] = \mathbb{P}[A \cap A] = \mathbb{P}[A]\,\mathbb{P}[A]$$

□

4. Kolmogorov's Maximal Inequality

Theorem 6.12. *Suppose $X_1, \ldots, X_n$ are independent with mean zero and finite variance. Then:*

$$\mathbb{P}\left[\max_{1 \leq k \leq n} |S_k| \geq x\right] \leq \frac{\mathbb{V}[S_n]}{x^2}$$

Proof. Let $A_k = \{S_k \geq x : \; j < k \implies S_j < x\}$. Then, $\{\max_{1 \leq k \leq n} |S_k| \geq x\} = \bigcup_{k=1}^n A_k$, and A_k are disjoint. We achieve the result by working with both ends:

(a) By Chebyshev's Inequality: $x^2 \sum_{k=1}^n \mathbb{P}[A_k] \leq \sum_{k=1}^n \int S_k^2 \mathbf{1}\{A_k\}\, d\mathbb{P}$

(b) From the other side:

$$\begin{aligned}
\int_{A_k} S_n^2 &= \int_{A_k} (S_k + S_n - S_k)^2 \\
&= \int_{A_k} S_k^2 + 2S_k(S_n - S_k) + (S_n - S_k)^2 \\
&= \int_{A_k} S_k^2 + (S_n - S_k)^2 \\
&\geq \int_{A_k} S_k^2 \\
&\geq x^2 \mathbb{P}[A_k]
\end{aligned}$$

Taking the sum over k, the result follows. □

5. Applications of Kolmogorov's Maximal Inequality

(a) Convergence of Random Series

Theorem 6.13. *Let $X_1, \ldots$ be mean zero, independent random variables. If $\sum_{n=1}^\infty \mathbb{V}[X_n] < \infty$ then $\sum_{n=1}^\infty X_n$ converges a.s.*

Proof. Let S_n be the partial sums. We show that the Cauchy sequence converges almost surely, which implies the existence of a limiting random variable. We want to show that for any $\epsilon > 0$ as $M \to \infty$:

$$\mathbb{P}\left[\sup_{n,m \geq M} |S_n - S_m| > 2\epsilon\right] \leq 2\mathbb{P}\left[\sup_{n \geq M} |S_n - S_M| > \epsilon\right] \to 0$$

Kolmogrov's Maximal Inequality only applies in the finite case. Hence, we apply it to the finite case and then take a limit:

$$\begin{aligned}\mathbb{P}\left[\max_{M\leq n\leq N}|S_n - S_M| > \epsilon\right] &\leq \frac{1}{\epsilon^2}\mathbb{V}\left[S_N - S_M\right]\\ &\leq \frac{1}{\epsilon^2}\sum_{i=M+1}^{N}\mathbb{V}\left[X_i\right]\\ &\leq \sum_{i=M+1}^{\infty}\mathbb{V}\left[X_i\right]\end{aligned}$$

The right hand side no longer depends on N, and so taking the monotonic limit, we have that the supremum is bounded by the tail sum of variances from $M+1$ to ∞, which tends to 0 as $M \to \infty$ by the assumption. Hence, the result holds. □

(b) Kolmogrov's Three Series Theorem

Theorem 6.14. *Let $X_1, \ldots$ be independent random variables. Let $A > 0$ and $Y_i = X_i \mathbf{1}\{|X_i| \leq A\}$. Then $\sum_{i=1}^{\infty} X_i$ converges if and only if the following three hold:*

i. *Control Truncation Error: $\sum_n \mathbb{P}\left[|X_n| > A\right] < \infty$*
ii. *Control Variance of Truncation: $\sum_n \mathbb{V}\left[Y_n\right] < \infty$*
iii. *Convergence of Truncation Mean: $\sum_n \mathbb{E}\left[Y_n\right]$ converges*

Proof. We prove the ($\Leftarrow$) direction first. By the second condition, $\sum_n Y_n - \mathbb{E}\left[Y_n\right]$ converge a.s.. By the second condition, $\sum_n Y_n$ converges almost surely. By Borel Cantelli and the first condition, $\mathbb{P}\left[X_n \neq Y_n\right] = 0$, which implies the result.

For ($\Rightarrow$). Suppose the series converges but the truncation error control does not hold. Then by the second Borel-Cantelli lemma, $\mathbb{P}\left[|X_n| > A \ i.o.\right] = 1$ which implies $\sum X_n$ cannot converges. Hence, the first condition holds.

Now suppose the variance goes to infinity and define $c_n = \sum_{m=1}^{n} \mathbb{V}\left[Y_m\right]$

i. Since $\sum X_m$ converges a.s. then $\sum Y_m$ converges a.s. and so

$$T_n = \frac{\sum Y_m}{\sqrt{c_n}} \to 0 \sim d$$

ii. Letting

$$X_{n,m} = \frac{Y_m - \mathbb{E}\left[Y_m\right]}{\sqrt{c_n}}$$

and $S_n = \sum_m X_{n,m}$, by Lindberg Feller CLT, $S_n \to \chi \sim d$.

Therefore, $S_n - T_n \to \chi \sim d$ but $S_n - T_n$ is not random. Now for the final condition, we have that $\sum Y_n$ converges a.s. and $\sum Y_n - \mathbb{E}\left[Y_n\right]$ converges a.s. □

6. Alternative Proof of SLLN

   (a) Kronecker's Lemma

   Lemma 6.5. *If* $a_n \uparrow \infty$ *and* $\sum_{n=1}^{\infty} \frac{x_n}{a_n}$ *converges then* $\frac{\sum_{m=1}^{n} x_m}{a_n} \to 0$.

   (b) Strong Law of Large Numbers

   Theorem 6.15. *Let* $X_1, \ldots$ *be i.i.d. random variables with* $\mathbb{E}[|X_i|] < \infty$*. Let* $\mathbb{E}[X_i] = 0$*. Then:*

$$\frac{S_n}{n} \to 0 \sim a.s.$$

   Proof. We will apply Kronecker's Lemma by showing that $\sum_{n=1}^{\infty} \frac{X_n}{n}$ converges a.s.. To do this, we need only bound the sum of the variances, which follows from **Lemma 6.4**:

$$\sum_{n=1}^{\infty} \frac{\mathbb{V}[X_n]}{n^2} \leq 4\mathbb{E}[|X_1|] < \infty$$

□

Part III
Central Limit Theorems

7 DeMoivre-Laplace Theorem

1. Properties of Exponential Function

   Lemma 7.1. *The following results are in the real field, but extend to complex values.*

   (a) If $c_j \to 0$*,* $a_j \to \infty$*, and* $a_j c_j \to \lambda$ *then* $(1 + c_j)^{a_j} \to \exp(\lambda)$

   (b) If $\max_{1 \leq j \leq n} |c_{j,n}| \to 0$*,* $\sum_{j=1}^{n} c_{j,n} \to \lambda$ *and* $\sup_n \sum_{j=1}^{n} |c_{j,n}| < \infty$ *then* $\prod_{j=1}^{n}(1 + c_{j,n}) \to \exp(\lambda)$

   Proof. For the first result, we have that: $\log(1 + c_j)/c_j \to 1$ as $c_j \to 0$. Hence:

$$a_j \log(1 + c_j) = a_j c_j \frac{\log(1 + c_j)}{c_j} \to \lambda$$

   For the second result, we have that;

$$\sum_{j=1}^{n} c_{j,n} \frac{1 + c_{j,n}}{c_{j,n}} \to \lambda$$

□

2. Let X_i be independent Rademacher random variables and S_n be their partial sums

3. DeMoivre-Laplace Theorem

Theorem 7.1. *If $a < b$ and $m \to \infty$ then*

$$\mathbb{P}\left[a \leq \frac{S_m}{\sqrt{m}} \leq b\right] \to \int_a^b \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right)$$

Proof. Note that when n is even, S_n is even. When n is odd, S_n is odd. Since $S_{2n+1} \in \{S_{2n}+1, S_{2n}-1\}$ it is simpler to consider the even case and then extend to the odd case.

(a) For $n, k \in \mathbb{Z}$, we have that:

$$\begin{aligned}\mathbb{P}[S_{2n} = 2k] &= \mathbb{P}[n+k : +1, n-k : -1] \\ &= \binom{2n}{n+k} 2^{-2n}\end{aligned}$$

For sufficiently large n, we can use Stirling's Approximation:

$$\begin{aligned}\mathbb{P}[S_{2n} = 2k] &= \frac{(2n)!}{(n+k)!(n-k)!} 2^{-2n} \\ &\approx \frac{n^{2n}\sqrt{4\pi n}}{(n+k)^{n+k}\sqrt{2\pi(n+k)}(n-k)^{n-k}\sqrt{2\pi(n-k)}} \\ &\approx \frac{1}{\sqrt{\pi n}}\left(1-\frac{k^2}{n^2}\right)^{-n-1/2}\left(1+\frac{k}{n}\right)^{-k}\left(1-\frac{k}{n}\right)^{k}\end{aligned}$$

(b) Suppose $2k^2/n \to x^2 \in \mathbb{R}$. Then, applying **Lemma 7.1**:

$$\begin{aligned}&\mathbb{P}[S_{2n} = 2k] \\ &\approx \frac{1}{\sqrt{\pi n}}\left(1-\frac{x^2/2}{n}\right)^{-n-1/2}\left(1+\frac{x}{\sqrt{2n}}\right)^{-x\sqrt{n/2}}\left(1-\frac{x}{\sqrt{2n}}\right)^{x\sqrt{n/2}} \\ &\approx \frac{1}{\sqrt{\pi n}} \exp(x^2/2)\exp(-x^2/2)\exp(-x^2/2) \\ &\approx \frac{1}{\sqrt{\pi n}} \exp(-x^2/2)\end{aligned}$$

(c) Therefore:

$$\begin{aligned}\mathbb{P}\left[a \leq S_{2n}/\sqrt{2n} \leq b\right] &= \sum_{m \in [a\sqrt{2n}, b\sqrt{2n}] \cap 2\mathbb{Z}} \mathbb{P}[S_{2n} = m] \\ &\sim \sum_{m \in [a,b] \cap \sqrt{2/n}\mathbb{Z}} \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)\left(\frac{2}{n}\right)^{1/2}\end{aligned}$$

This is a Riemann sum, and so in the limit, we have the result.

□

8 Weak Convergence

8.1 Definition and Basic Results

1. Converge Weakly. Converge in Distribution

   Definition 8.1.

   (a) *A sequence of distribution functions, F_n, converges weakly to a limit F if $F_n(y) \to F(y)$ at all continuity points of F.*

   (b) *A sequence of random variables X_n converges weakly or converge in distribution if their distribution functions converge weakly.*

2. Examples

   Example 8.1. *The DeMoivre Laplace Theorem which states that Rademacher random variables converge weakly to χ, a normally distributed random variable, is an example of weak convergence.*

   Example 8.2. *Convergence at Discontinuities. Let $X \sim F$ and $X_n = X + \frac{1}{n}$. Then $F_n(x) = F(x - 1/n)$ and*

   $$\lim_n F_n(x) = F(x-)$$

   Therefore, if x is a discontinuity, we have an issue if weak convergence required convergence at points of discontinuity as well.

3. Weak Convergence and Identically Distributed Random Variables Converging a.s.

   Theorem 8.1. *If $F_n \to F_\infty \sim d$, then $\exists Y_n$ for $1 \leq n \leq \infty$ such that $Y_n \sim F_n$ and $Y_n \to Y_\infty \sim$ a.s..*

   Proof. Let $F = F_\infty$. Let $Y_n(\omega) = \sup\{y : F_n(y) < \omega\}$. Let $\Omega_0 \subset [0,1]$ for which F is continuous, then on Ω_0, F^{-1} exists and $F \circ F^{-1}(\omega) = \omega$. For $\omega \in \Omega_0$:

   (a) Suppose $y < F^{-1}(\omega)$. Then $F(y) < \omega$. For sufficiently large n, $F_n(y) < \omega$. Hence, $y \leq Y_n(\omega)$. Letting $y \uparrow F^{-1}(\omega)$, $Y_\infty(\omega) \leq \liminf_n Y_n(\omega)$.

   (b) Suppose $y > F^{-1}(\omega)$. Then $F(y) > \omega$. For sufficiently large n, $F_n(y) > \omega$. Hence, $y \geq Y_n(\omega)$. Letting $y \downarrow F^{-1}(\omega)$, $Y_\infty(\omega) \geq \limsup_n Y_n(\omega)$.

   □

4. Convergence Results

   (a) Fatou's Lemma

   Lemma 8.1. *If $g \geq 0$ and continuous, and $X_n \to X_\infty \sim d$, then* $\liminf_n \mathbb{E}[g(X_n)] \geq \mathbb{E}[g(X_\infty)]$

Proof. By **Theorem 8.1**, $\exists Y_n$ such that $Y_n \to Y_\infty \sim a.s.$. Since g is continuous, $g(Y_n) \to g(Y)$ a.s.. Thirdly, since $g \geq 0$, by Fatou's lemma:

$$\liminf \mathbb{E}\left[g(X_n)\right] = \liminf \mathbb{E}\left[g(Y_n)\right] \geq \mathbb{E}\left[g(Y_\infty)\right] = \mathbb{E}\left[g(X_\infty)\right]$$

□

(b) Integration to Limit

Lemma 8.2. *Let g, h be continuous, $g > 0$, and $|h(x)|/g(x) \to 0$ as $|x| \to \infty$. If $F_n \to F_\infty \sim d$ and $\int g(x)dF_n(x) \leq C < \infty$ then*

$$\int h(x)dF_n(x) \to \int h(x)dF_\infty(x)$$

Proof. This is the standard proof using truncation and approximating the error with g. Let $\epsilon > 0$. Then, there is an $M > 0$ such that M is a continuity point of F and $|h(x)|/g(x) < \epsilon$ for $|x| > M$. From **Theorem 8.1**, there are $Y_n \to Y_\infty$ a.s.. Letting $\bar{Y}_n = Y_n \mathbf{1}\left\{|Y_n| \leq M\right\}$, we have that $\bar{Y}_n \to \bar{Y}_\infty$ a.s. By dominated convergence:

$$\int_{[-M,M]} h(x)dF_n(x) \to \int_{[-M,M]} h(x)dF_\infty(x)$$

For the error term:

$$\int_{[-M,M]^c} |h(x)|dF_n(x) \leq \epsilon \int_{[-M,M]^c} g(x)dF_n(x) \leq \epsilon C$$

Since $\epsilon > 0$ is arbitrary, we have the result. □

(c) Equivalent Definitions of Weak Convergence

i. Convergence in Topology

Theorem 8.2. *$X_n \to X \sim d$ if and only if for every bounded continuous g, $\mathbb{E}\left[g(X_n)\right] \to \mathbb{E}\left[g(X)\right]$*

Proof. For ($\Rightarrow$), we can find $Y_n \to Y$ a.s. and if g is bounded and continuous then by dominated convergence theorem:

$$\mathbb{E}\left[g(X_n)\right] = \mathbb{E}\left[g(Y_n)\right] \to \mathbb{E}\left[Y\right] = \mathbb{E}\left[X\right]$$

For ($\Leftarrow$), we choose a specific form of g so that we have

$$\mathbb{P}\left[X_n \leq x\right] \to \mathbb{P}\left[X \leq x\right]$$

at all continuity points of X. Let x be a point of continuity and $\epsilon > 0$. Consider

$$g_{x,\epsilon}(y) = \begin{cases} 1 & y \leq x \\ 1 - \frac{1}{\epsilon}(y - x) & y \in (x, x+\epsilon) \\ 0 & y > x + \epsilon \end{cases}$$

Then $g_{x,\epsilon}$ is bounded and continuous. Therefore:

$$\mathbb{P}[X_n \leq x] \leq \mathbb{E}[g_{x,\epsilon}(X_n)] \to \mathbb{E}[g_{x,\epsilon}(X)] \leq \mathbb{P}[X \leq x+\epsilon]$$

Hence, $\limsup_n \mathbb{P}[X_n \leq x] \leq \mathbb{P}[X \leq x]$. We now reverse the roles slightly.

$$\mathbb{P}[X_n \leq x] \geq \mathbb{E}[g_{x-\epsilon,\epsilon}(X_n)] \to \mathbb{E}[g_{x-\epsilon,\epsilon}(X)] \geq \mathbb{P}[X \leq x-\epsilon]$$

Hence, $\liminf_n \mathbb{P}[X_n \leq x] \geq \mathbb{P}[X \leq x]$. □

ii. Other Equivalences

Lemma 8.3. *The following are equivalent:*

A. $X_n \to X$ *in distribution*

B. $\forall G$*, open,* $\liminf \mathbb{P}[X_n \in G] \geq \mathbb{P}[X \in G]$

C. $\forall K$*, closed,* $\limsup \mathbb{P}[X_n \in K] \leq \mathbb{P}[X \in K]$

D. For all A *such that* $\mathbb{P}[X \in \partial A] = 0$*,* $\lim \mathbb{P}[X_n \in A]\,\mathbb{P}[X \in A]$

Proof. For ($A \implies B$), since $\liminf \mathbf{1}\{X_n \in G\} \geq \mathbf{1}\{X \in G\}$ (consider boundary points), by Fatou's Lemma:

$$\liminf_n \mathbb{P}[X_n \in G] \geq \mathbb{P}[X \in G]$$

For ($B \implies C$), we can use complements of open sets. We now show ($B, C \implies D$). Noting that the interior of A is open and the closure of A is the interior with the boundary, the result follows. To show ($D \implies A$), we consider $A = (-\infty, x]$ where x is a continuity point. □

(d) Continuous Mapping Theorem

Theorem 8.3. *Let g be measurable and* $D_g = \{x : \lim_{y \to x} g(y) \neq g(x)\}$*. Suppose* $X_n \to X$ *in distribution and* $\mathbb{P}[X \in D_g] = 0$.

i. Then $g(X_n) \to g(X)$ *in distribution*

ii. If g is bounded then $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)]$

Proof. We will use the toplogical equivalence to show that for any bounded, continuous function f, $\mathbb{E}[(f \circ g)(X_n)] \to \mathbb{E}[(f \circ g)(X)]$ implying that $g(X_n) \to g(X)$ in distribution. So let f be an continuous, bounded function. Then $\mathbb{P}[X \in D_{f \circ g}] = 0$ by assumption. Letting $Y_n \to Y$ a.s. with the same distributions as X_n and X, then

$$\mathbb{E}[(f \circ g)(X_n)] = \mathbb{E}[(f \circ g)(Y_n)] \to \mathbb{E}[(f \circ g)(Y)] = \mathbb{E}[(f \circ g)(X)]$$

Moreover, when g is bounded then this follows from dominated convergence. □

8.2 Sequential Compactness of Distribution Functions

1. Helly's Selection Theorem

   Theorem 8.4. *For every sequence F_n of distributions, there is a subsequence which converges to a right continuous, non-decreasing function F at the continuity points of F.*

   Proof. Consider $x \in \mathbb{Q}$. Then $F_n(x)$ is a compact sequence in $[0, 1]$. Hence, it has a subsequence. We use this fact, and subsequences of subsequences to extract a converging subsequence. Let $x_1, \dots$ be some enumeration of the rationals. Let $F_{n(j,1)}(x_1) \to G(x_1)$. From $n(j, 1)$ we extract $n(j, 2)$ such that $F_{n(j,2)} \to G(x_2)$. Continuing, we have that $F_{n(m,j)}(x_j) \to G(x_j)$. Then, we have that $F_{n(j,j)}$ is a sequence of functions such that $F_{n(j,j)}(x) \to G(x)$ for ever rational.

   Let $x \in \mathbb{R}$. Let $s_n \in \mathbb{Q}$ such that $s_n \downarrow x$. And define $G(x) = \lim_n G(s_n)$. Hence, G is right continuous and it is increasing. We now must show that $F_{n(j,j)}(x) \to G(x)$ for any x which is a continuity point of G. Let x be a continuity point of G. Let $\epsilon > 0$ and let $s < x < t$ where $s, t \in \mathbb{Q}$ and

   $$G(x) - \epsilon < G(s) < G(x) < G(t) < G(x) + \epsilon$$

   For sufficiently large j:

   $$G(x) - \epsilon < F_{n(j,j)}(s) < G(x) < F_{n(j,j)}(t) < G(x) + \epsilon$$

   Hence, $G(x) - \epsilon < F_{n(j,j)}(x) < G(x) + \epsilon$. Since $\epsilon > 0$ is arbitrary, letting it go to 0 forces $j \to \infty$, and so $G(x) = \lim_j F_{n(j,j)}(x)$. □

2. The function F may not necessarily be a distribution.

   Example 8.3. *Let $a+b = 1$ and $a, b > 0$. Consider $F_n(x) = a\mathbf{1}\{x \geq -n\} + b\mathbf{1}\{x \geq n\}$. Then as $n \to \infty$, it converges to $F(x) = a$ which does not satisfy the requirements for a distribution.*

   Note 8.1. *We need a notion of "tightness" to determine when subsequences converge to a distribution.*

3. Tightness

   Definition 8.2. *A sequence μ_n of measures is tight if $\forall \epsilon > 0$, $\exists M_\epsilon > 0$ such that $\forall n$:*

   $$\limsup_n \mu_n[-M_\epsilon, M_\epsilon]^c \leq \epsilon$$

   A sequence of distributions F_n is tight if

   $$\limsup_n 1 - F_n(M_\epsilon) + F_n(-M_\epsilon) \leq \epsilon$$

4. Tightness and Helly

Corollary 8.1. *Consider a sequence of distribution functions. Every subsequential limit is a distribution function if and only if the sequence is tight.*

Proof. First ($\Rightarrow$). Suppose F_n are not tight, but every subsequence converges to a distribution function. Let F be this limit. Let $\epsilon > 0$. Then, there is an M such that $1 - F(M) + F(-M) < \epsilon/2$. For sufficiently large n,

$$0 \leq F(M) - \inf_{m \geq n} F_m(M) \leq \epsilon/4$$

and

$$0 \leq \sup_{m \geq n} F_m(-M) - F(-M) \leq \epsilon/4$$

Hence:

$$\sup_{m \geq n} 1 - F_m(M) - F_m(-M) \leq \epsilon$$

Therefore, F_n are tight.

Now ($\Leftarrow$). Since F is right continuous and non-decreasing, we need only check that the measure associated with F has measure 1 on the whole space. Let $\epsilon > 0$, then there $\exists M$, for which $-M, M$ are continuity points of F, such that:

$$\begin{aligned} \epsilon &\geq \limsup 1 - F_m(M) + F_m(-M) \\ &\geq 1 - \liminf F_m(M) + \limsup F_m(-M) \\ &\geq 1 - F(M) + F(M) \end{aligned}$$

Letting $\epsilon \to 0$, we see that $M \to \infty$ and so the whole space has measure 1. □

5. Sufficient Condition for Tightness

Lemma 8.4. *If $\exists \phi$ such that $\phi(x) \geq 0$, $\phi(x) \to \infty$ as $|x| \to \infty$, and*

$$\infty > C = \sup_n \int \phi(x) dF_n(x)$$

Then $F_n(x)$ are tight.

Proof. Let $\epsilon > 0$. Then there is an N such that $C \leq \epsilon N$. And there is an M such that $\phi(x) \geq N$ if $|x| > M$. Then:

$$\sup_n N(1 - F_n(M) + F_n(-M)) \leq \sup_n \int_{[-M,M]^c} \phi(x) dF_n(x) \leq C$$

□

9 Characteristic Functions

9.1 Definition and Properties

1. Characteristic Function

   Definition 9.1. *Let X be a random variables. Its characteristic function is*
   $$\phi_X(t) = \mathbb{E}\left[\exp(itX)\right]$$

2. Basic Properties

   Proposition 9.1. *Let $\phi(t)$ be a characteristic function and X be a random variable.*

   (a) $\phi(0) = 1$

   (b) $|\phi(t) \leq 1$

   (c) $\phi(-t) = \overline{\phi(t)}$

   (d) $\phi(t)$ *is uniformly continuous on* $(-\infty, \infty)$

   (e) $\mathbb{E}\left[\exp(it(aX+b))\right] = \exp(itb)\phi_X(ta)$

   Proof. For the first item, we notice that $\phi(0) = \mathbb{E}[1] = 1$. For the second, we note that:
   $$|\phi(t)| \leq \mathbb{E}\left[|\exp(itX)|\right] \leq 1$$
   For the third:
   $$\phi(-t) = \mathbb{E}\left[\exp(-itX)\right] = \overline{\phi(t)}$$
   For the fourth:
   $$|\phi(t+h) - \phi(t)| = |\mathbb{E}\left[\exp(itX)\left(\exp(ith) - 1\right)\right]| \leq \mathbb{E}\left[|\exp(ihX) - 1|\right]$$
   The last result follows from the linearity of integrals and properties of exponentials. □

3. Sums of Independent Random Variables

   Lemma 9.1. *If X_1, X_2 are independent random variables then*
   $$\phi_{X_1}(t)\phi_{X_2}(t) = \phi_{X_1+X_2}(t)$$

   Proof.
   $$\mathbb{E}\left[\exp\left(itX_1 + itX_2\right)\right] = \mathbb{E}\left[\exp(itX_1)\right]\mathbb{E}\left[\exp(itX_2)\right]$$
   □

4. Moments and Derivatives

   (a) ϕ is differentiable (if X has finite moments)

Theorem 9.1. *If $\mathbb{E}[|X|^n] < \infty$ then its characteristic function has continuous derivative up to order n given by*

$$\phi^{(n)}(t) = \mathbb{E}[(iX)^n \exp(itX)]$$

Proof. Fix n, then for any $1 \le m \le n$, $\mathbb{E}[|X|^m] < \infty$. Moreover, for small h:

$$\left|\frac{\phi(t+h)-\phi(t)}{h}\right| \le \mathbb{E}\left[\left|\frac{\exp(ihX)-1}{h}\right|\right]$$

So by dominated convergence: $\phi'(t) = \mathbb{E}[iX\exp(itX)]$. Higher order derivatives follows from dominated convergence as well. □

(b) Error Estimation

Theorem 9.2. *For any x:*

$$\left|\exp(ix) - \sum_{m=0}^{n}\frac{(ix)^m}{m!}\right| \le \min\left(\frac{|x|^{n+1}}{(n+1)!}, \frac{2|x|^n}{n!}\right)$$

(c) Quadratic Error Estimation in ϕ

Corollary 9.1. *Let X have finite second moment. Then:*

$$\left|\phi_X(t) - \sum_{j=0}^{2}\frac{\phi_X^{(j)}(0)}{j!}t^j\right| \le t^2\mathbb{E}\left[|t||X|^3 \wedge 2|X|^2\right]$$

or

$$\phi_X(t) = 1 + it\mathbb{E}[X] - t^2\frac{\mathbb{E}[X^2]}{2} + o(t)$$

9.2 Inversion Formula and Weak Convergence

1. Inversion Formula

Theorem 9.3. *Let $\phi(t) = \int \exp(itx)\mu(dx)$. If $a < b$ then*

$$\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{\exp(-ita)-\exp(-itb)}{it}\phi(t)dt = \mu(a,b) + \frac{1}{2}\mu\{a,b\}$$

Proof. Let

$$I_T = \int_{-T}^{T}\frac{\exp(-ita)-\exp(-itb)}{it}\phi(t)dt$$

Then, using Fubini's Theorem:

$$\begin{aligned} I_T &= \int_{-T}^{T} \int \frac{\exp(-it(a-x) - \exp(-it(b-x))}{it} \mu(dx)dt \\ &= \int \int_{-T}^{T} \frac{\exp(-it(a-x) - \exp(-it(b-x))}{it} dt \ \mu(dx) \\ &= \int \int_{-T}^{T} \frac{\sin(t(x-a)) - \sin(t(b-x))}{t} dt \ \mu(dx) \end{aligned}$$

Let

$$R(\theta, T) = \int_{-T}^{T} \frac{\sin(t\theta)}{t} dt \qquad S(T) = \int_{0}^{T} \frac{\sin(t)}{t} dt$$

Hence, $R(\theta, T) = 2sgn(\theta)S(T|\theta|)$. So as $T \to \infty$, $R(\theta, T) \to \pi sgn(\theta)$. So:

$$R(x-a, T) - R(x-b, T) \to \begin{cases} 0 & x > a, x > b \\ 0 & x < a, x < b \\ \pi & x = a \\ \pi & x = b \\ 2\pi & a < x < b \end{cases}$$

Since $|R(\theta, T)| \leq 2 \sup_y S(y) < \infty$ by dominated convergence:

$$\frac{1}{2\pi} I_T \to \mu(a, b) + \frac{1}{2}\mu\{a, b\}$$

□

2. Continuity Theorem

Theorem 9.4. *Let μ_n for $1 \leq n \leq \infty$ be probability measures with characteristic functions ϕ_n.*

(a) If $\mu_n \to \mu_\infty$ in distribution then $\phi_n(t) \to \phi_\infty(t)$ for all t

(b) If $\phi_n(t) \to \phi_\infty(t)$ and $\phi_\infty(t)$ is continuous at 0 then μ_n are tight, $\mu_n \to \mu_\infty \sim d$ and μ_∞ has characteristic function $\phi_\infty(t)$.

Proof. We only prove the first claim. Let $Y_n \sim \mu_n$ such that $Y_n \to Y_\infty$ a.s.. Then by dominated convergence:

$$\phi_n(t) = \mathbb{E}[\exp(itY_n)] \to \mathbb{E}[\exp(itY_\infty)] = \phi_\infty(t)$$

□

10 Central Limit Theorem

1. Independent, Identically Distributed Sequence

   Theorem 10.1. *Let* $X_1, \ldots$ *be i.i.d with* $\mathbb{E}[X_i] = 0$ *and* $\mathbb{V}[X_i] = \sigma^2 \in (0, \infty)$. *Then:*

   $$\frac{S_n}{\sqrt{n}\sigma} \to \chi \sim d$$

   Proof. We have that

   $$\phi_{S_n/\sqrt{n\sigma^2}}(t) = \prod_{i=1}^{n} \phi_{X_i/\sqrt{n\sigma^2}}(t) = \left(1 - \frac{t^2/2}{n} + o(n^{-1})\right)^n \to \exp(-t^2/2)$$

   From the continuity theorem, the result follows. □

2. Lindberg-Feller Central Limit Theorem

   Theorem 10.2. *For each* n*, let* $X_{n,m}$*,* $1 \le m \le n$*, be independent random variables with* $\mathbb{E}[X_{n,m}] = 0$*. Suppose*

   (a) *Converge in Variance:* $\sum_{m=1}^{n} \mathbb{E}\left[X_{n,m}^2\right] \to \sigma^2 > 0$

   (b) *Error Control: For all* $\epsilon > 0$*,*

   $$\lim_n \sum_{m=1}^{n} \mathbb{E}\left[|X_{n,m}|^2 \mathbf{1}\{|X_{n,m} > \epsilon\}\right] = 0$$

   Then $S_n \to \sigma\chi \sim d$*.*

   Proof. Again we use characteristic functions to demonstrate this result. We need to show that

   $$\prod_m \phi_{n,m}(t) \to \exp(-\sigma^2 t^2/2)$$

   By continuity and **Lemma 7.1**:

   $$\prod_m (1 - \mathbb{E}\left[X_{n,m}^2\right] t^2/2) \to \exp(-\sigma^2 t^2/2)$$

   It is also not too difficult to show the first inequality, and the second follows from the error bound

   $$\begin{aligned}
   &\left|\prod_m \phi_{n,m}(t) - \prod_m (1 - \mathbb{E}\left[X_{n,m}^2\right] t^2/2)\right| \\
   &\le \sum_m \left|\phi_{n,m}(t) - (1 - \mathbb{E}\left[X_{n,m}^2\right] t^2/2))\right| \\
   &\le \sum_m t^3 \mathbb{E}\left[|X_{n,m}|^3 \mathbf{1}\{|X_{n,m} \le \epsilon\}\right] + 2t^2 \mathbb{E}\left[|X_{n,m}^2 \mathbf{1}\{|X_{n,m} \ge \epsilon\}\right] \\
   &\to \epsilon t^3 \sigma^2
   \end{aligned}$$

   □

11 Lindberg's Method

1. Lindberg's Condition

   Definition 11.1. *Let $X_{n,m}$ be a triangular array of random variables. The following is Lindberg's Condition:*

   $$\lim_n \sum_{m=1}^{n} \mathbb{E}\left[|X_{n,m}|^2 \mathbf{1}\{|X_{n,m} > \epsilon\}\right] = 0$$

2. Lindberg's Method

   (a) First, we use the fact that $\sum_{m=1}^n X_{n,i} \to \chi$ is equivalent to convergence in topology.

   (b) Second, we note that C^2 functions are dense in all continuous bounded functions.

   (c) Third, find $\zeta_{n,i}$ which are normally distributed such that $\sum_{i=1}^n \zeta_{n,i} \sim N(0,1)$

   (d) Fourth, using the triangle inequality (consider $n = 2$):

   $$\left|\mathbb{E}\left[f(\sum_i X_{n,i})\right] - \mathbb{E}\left[f(\sum_i \zeta_{n,i})\right]\right|$$
   $$\leq \sum_k \left|\mathbb{E}\left[f(\sum_{i=1}^{k-1} X_{n,i} + \sum_{i=k}^{n} \zeta_{n,i})\right] - \mathbb{E}\left[f(\sum_{i=1}^{k} X_{n,i} + \sum_{i=k+1}^{n} \zeta_{n,i})\right]\right|$$

   (e) Fifth, using Taylor's theorem and Lindberg's condition, show that errors go to 0.

Part IV
Conditional Expectation & Martingales

12 Conditional Expectation

12.1 Existence and Uniqueness

1. Conditional Expectation. Versions.

   Definition 12.1. *Let $(\Omega, \mathcal{F}_0, \mathbb{P})$ be a probability space and $\mathcal{F} \subset \mathcal{F}_0$ be a sub-σ-algebra. Let $X \in L^1(\Omega, \mathcal{F}_0, \mathbb{P})$. The conditional expectation of X given $\mathcal{F}$, $\mathbb{E}[X|\mathcal{F}]$, is a random variable such that:*

   (a) $\mathbb{E}[X|\mathcal{F}]$ *is* $\mathcal{F}$ *measurable*

   (b) $\forall A \in \mathcal{F}$, $\int_A X d\mathbb{P} = \int_A \mathbb{E}[X|\mathcal{F}]\, d\mathbb{P}$

   Any random variables satisfying these two properties are called versions.

2. Conditional Random Variables are Integrable

Lemma 12.1. *Suppose Y is a version of the conditional expectation of X with respect to $\mathcal{F}$. Then $\mathbb{E}[|Y|] < \infty$.*

Proof. Let $A = \{Y > 0$. Then:

$$\int Y^+ = \int_A Y = \int_A X \leq \mathbb{E}[|X|] < \infty$$

and:

$$\int Y^- = \int_{A^c} Y = \int_{A^c} X \leq \mathbb{E}[|X|] < \infty$$

□

3. Uniqueness

   (a) Uniqueness

   Lemma 12.2. *Let $X \in L^1$. Then $\mathbb{E}[X|\mathcal{F}]$ is a.s. unique.*

   Proof. Let Y and Z be two versions of $\mathbb{E}[X|\mathcal{F}]$. Let $A_\epsilon = \{Z - Y > \epsilon > 0\} \in \mathcal{F}$. Then:

$$\int_{A_\epsilon} Z = \int_{A_\epsilon} X = \int_{A_\epsilon} Y$$

   Therefore:

$$0 = \int_{A_\epsilon} Z - Y \geq \epsilon \mathbb{P}[A_\epsilon]$$

   Hence, $\mathbb{P}[A_\epsilon] = 0$. This holds for all ϵ and interchanging the roles of Z and Y gives that:

$$Z = Y \sim a.s.$$

□

   (b) Generalization of Uniqueness

   Lemma 12.3. *Suppose $X_1 = X_2$ on $B \in \mathcal{F}$. Then $\mathbb{E}[X_1|\mathcal{F}] = \mathbb{E}[X_2|\mathcal{F}]$ a.s. on B.*

   Proof. Let $Y_1 = \mathbb{E}[X_1|\mathcal{F}]$ and $Y_2 = \mathbb{E}[X_2|\mathcal{F}]$ Let $A_\epsilon = \{Y_1 - Y_2 > \epsilon > 0\}$. Then:

$$0 = \int_{A_\epsilon \cap B} X_1 - X_2 = \int_{A_\epsilon \cap B} Y_1 - Y_2 \geq \epsilon \mathbb{P}[A_\epsilon \cap B]$$

   Therefore, $\mathbb{P}[A_\epsilon \cap B] = 0$. This holds for all ϵ and interchanging the roles of Y_1 and Y_2 the result follows. □

   Note 12.1. *We can get uniqueness from this proof quite easily. Just by letting $X = X_1 = X_2$ which holds on Ω.*

4. Existence

(a) Absolutely Continuous.

Definition 12.2. *Let ν and μ be measures. ν is absolutely continuous with respect to μ, $\nu \ll \mu$, if whenever $\mu(A) = 0$ then $\nu(A) = 0$.*

(b) Radon-Nikodym Theorem

Theorem 12.1. *Let μ and ν be σ-finite measures on $(\Omega, \mathcal{F})$. If $\nu \ll \mu$ then $\exists f \in \mathcal{F}$ such that $\forall A \in \mathcal{F}$:*

$$\nu(A) = \int_A f d\mu$$

f is called the Radon-Nikodym derivative and is denoted $d\nu/d\mu$

(c) Existence

Corollary 12.1. *Let $\mathbb{E}[|X|] < \infty$ and $\mathcal{F} \subset \mathcal{F}_0$ be a σ-algebra. Then, $\exists Y$ which is a version of $\mathbb{E}[X|\mathcal{F}]$.*

Proof. Define $\nu(A) = \int_A X d\mathbb{P}$ where $A \in \mathcal{F}$. Then ν is a probability measure and $\nu \ll \mathbb{P}$ on $\mathcal{F}$. Hence there is a Y which is measurable in $\mathcal{F}$ such that:

$$\int_A Y d\mathbb{P} = \nu(A) = \int_A X d\mathbb{P}$$

Hence, $Y = \mathbb{E}[X|\mathcal{F}]$ a.s. □

12.2 Basic Properties

1. Linearity

Lemma 12.4. *Supposing $X, Y \in L^1$ and a is a number:*

$$\mathbb{E}[aX + y|\mathcal{F}] = a\mathbb{E}[X|\mathcal{F}] + \mathbb{E}[Y|\mathcal{F}]$$

Proof. Let $A \in \mathcal{F}$:

$$\int_A \mathbb{E}[aX + Y|\mathcal{F}] = \int_A aX + Y = a\int_A X + \int_A Y = \int a\mathbb{E}[X|\mathcal{F}] + \mathbb{E}[Y|\mathcal{F}]$$

□

2. Monotonicity

Lemma 12.5. *If $X \le Y$ a.s. then $\mathbb{E}[X|\mathcal{F}] \le \mathbb{E}[Y|\mathcal{F}]$ a.s.. In particular, if $X = 0$ then $0 \le \mathbb{E}[Y|\mathcal{F}]$*

Proof. Let $A \in \mathcal{F}$. Then:

$$\int_A \mathbb{E}[Y|\mathcal{F}] - \mathbb{E}[X|\mathcal{F}] = \int_A Y - X \ge 0$$

This holds for all $A \in \mathcal{F}$. We can specifically look at $A_\epsilon = \{\mathbb{E}[X|\mathcal{F}] - \mathbb{E}[Y|\mathcal{F}] > \epsilon > 0\}$ and this will give us that $\mathbb{P}[A_\epsilon] = 0$. □

3. Monotonic Convergence

Lemma 12.6. *If $X_n \geq 0$ and $X_n \uparrow X$ a.s. with $\mathbb{E}[X] < \infty$ then $Y_n := \mathbb{E}[X_n | \mathcal{F}] \uparrow \mathbb{E}[X | \mathcal{F}] =: Y$*

Proof. By monotonicity, $Y_n \leq Y_{n+1}$ and so by monotonic convergence theorem:

$$\int_A \lim Y_n = \lim \int_A Y_n = \lim \int_A X_n = \int_A X = \int_A Y$$

This holds for all $A \in \mathcal{F}$ and so $\lim Y_n = Y$ a.s. by uniqueness. □

4. Extension to all positive random variables.

Lemma 12.7. *Suppose $X \geq 0$ and $\mathbb{E}[X] = \infty$. There is a unique r.v. $Y \in \mathcal{F}$ such that:*

$$\int_A X d\mathbb{P} = \int_A Y d\mathbb{P}$$

for all $A \in \mathcal{F}$.

Proof. Let $X_M = X \wedge M$. Then, $X_M \in L^1$ and let its conditional expectation be Y_M. So:

$$\int_A Y_M = \int_A X_M$$

Using monotonic convergence theorem and denoting $\lim Y_M = Y$

$$\int_A Y = \int_A X$$

Uniqueness follows as it did before, we just consider it on Ω_0 where Y is finite. □

5. Chebyshev's Lemma

Lemma 12.8.

$$\mathbb{P}[|X| > a | \mathcal{F}] \leq \frac{1}{a^2} \mathbb{E}\left[X^2 \middle| \mathcal{F}\right]$$

Proof. By monotonicity:

$$\mathbb{E}\left[a^2 \mathbf{1}\{|X| > a\} \middle| \mathcal{F}\right] \leq \mathbb{E}\left[X^2 \mathbf{1}\{|X| > a\} \middle| \mathcal{F}\right] \leq \mathbb{E}\left[X^2 \middle| \mathcal{F}\right]$$

□

6. Jensen's Inequality

Lemma 12.9. *If ϕ is convex and $\mathbb{E}[|X|], \mathbb{E}[|\phi(X)|] < \infty$ then*

$$\phi(\mathbb{E}[X | \mathcal{F}]) \leq \mathbb{E}[\phi(X) | \mathcal{F}]$$

Proof. Let $S(x) = \{(a, b) \in \mathbb{Q}^2 : ax + b \leq \phi(x)\}$ and note then that:

$$\phi(x) = \sup_{S(x)} ax + b$$

Therefore:

$$a\mathbb{E}[X|\mathcal{F}] + b = \mathbb{E}[aX + b|\mathcal{F}] \leq \mathbb{E}[\phi(X)|\mathcal{F}]$$

Taking the supremum on the left hand side over S gives the result. □

7. Cauchy-Schwartz

Lemma 12.10.

$$\mathbb{E}[XY|\mathcal{F}]^2 \leq \mathbb{E}[X^2|\mathcal{F}]\,\mathbb{E}[Y^2|\mathcal{F}]$$

Proof.

$$\begin{aligned} 0 &\leq \mathbb{E}[(X + \theta Y)^2|\mathcal{F}] \\ &\leq \mathbb{E}[X^2|\mathcal{F}] + 2\theta\mathbb{E}[XY|\mathcal{F}] + \theta^2\mathbb{E}[Y^2|\mathcal{F}s] \end{aligned}$$

This means that the discriminant $b^2 - 4ac \leq 0$. Or:

$$4\mathbb{E}[XY|\mathcal{F}]^2 \leq 4\mathbb{E}[X^2|\mathcal{F}]\,\mathbb{E}[Y^2|\mathcal{F}]$$

□

8. Conditional Expectation & Contraction

Lemma 12.11. *Conditional Expectation is a contraction in L^p for $p \geq 1$*

Proof. By Jensen's Inequality:

$$|\mathbb{E}[X|\mathcal{F}]|^p \leq \mathbb{E}[|X|^p|\mathcal{F}]$$

Taking expectation of both sides gives the desired result. □

9. Russian Doll Properties

Lemma 12.12. *Let $\mathcal{F} \subset \mathcal{G} \subset \mathcal{F}_0$ be σ-algebras and $X \in \mathcal{F}_0$.*

(a) If $\mathbb{E}[X|\mathcal{G}] \in \mathcal{F}$ then $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X|\mathcal{F}]$ a.s..
(b) $\mathbb{E}[\mathbb{E}[X|\mathcal{F}]|\mathcal{G}] = \mathbb{E}[X|\mathcal{F}]$
(c) $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{F}] = \mathbb{E}[X|\mathcal{F}]$

Proof.

(a) By assumption the first condition for conditional expectation is satisfied. Noting that $A \in \mathcal{F} \subset \mathcal{G}$, the second condition is satisfied. By uniqueness, the result follows.

(b) Note that $Y := \mathbb{E}[X|\mathcal{F}] \in \mathcal{G}$ since $\mathcal{F} \subset \mathcal{G}$. Therefore, from the definition:

$$\mathbb{E}[Y|\mathcal{G}] = Y \sim a.s.$$

(c) Let $A \in \mathcal{F}$.s Then:

$$\int_A \mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{F}] = \int_A \mathbb{E}[X|\mathcal{G}] = \int_A X = \int_A \mathbb{E}[X|\mathcal{F}]$$

□

10. Measurability & Conditionals

Lemma 12.13. *If $X \in \mathcal{F}$, and $\mathbb{E}[|Y|], \mathbb{E}[|XY|] < \infty$ then $\mathbb{E}[XY|\mathcal{F}] = X\mathbb{E}[Y|\mathcal{F}]$ a.s.*

Proof. Let $A \in \mathcal{F}$. Then:

$$\int_B \mathbb{E}[Y\mathbf{1}\{A\}|\mathcal{F}] = \int_{B\cap A} Y = \int_B \mathbf{1}\{A\}\mathbb{E}[Y|\mathcal{F}]$$

This extends to simple functions by linearity. Then to positive measurable functions through monotonicity. And finally to general measurable functions through linearity. □

11. Independence & Conditionals

Lemma 12.14. *Suppose $\mathbb{E}[|X|], \mathbb{E}[|Y|], \mathbb{E}[|XY|] < \infty$. (Letting $\perp$ indicate independence)*

$$X \perp Y \implies \mathbb{E}[Y|\sigma(X)] = \mathbb{E}[Y]\, a.s. \implies \mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$

Proof. Suppose $X \perp Y$. Let $A \in \sigma(X)$ and B be independent of $\sigma(X)$. Then:

$$\int_A \int_\Omega Y = \int_\Omega \int_A Y = \int_\Omega \int_A \mathbb{E}[Y|X]$$

That is, for all $A \in \sigma(X)$:

$$\mathbb{E}[(\mathbb{E}[Y|X] - \mathbb{E}[Y])\mathbf{1}\{A\}] = 0$$

Since $\mathbb{E}[Y]$ is a constant, choosing $A_\epsilon = \{\mathbb{E}[Y|X] - \mathbb{E}[Y] > \epsilon > 0$, shows that $\mathbb{P}[A_\epsilon] = 0$. For the second part:

$$\mathbb{E}[XY] = \int_\Omega \mathbb{E}[XY|\sigma(X)] = \int_\Omega X\mathbb{E}[Y|\sigma(X)] = \mathbb{E}[X]\mathbb{E}[Y]$$

□

12. Orthogonality and Conditionals in L^2

Lemma 12.15. *Suppose $X \in L^2(\mathcal{F}_0)$. Then for $Y_0 = \mathbb{E}[X|\mathcal{F}]$:*

$$\mathbb{E}\left[(X - Y_0)^2\right] = \min_{Y \in L^2(\mathcal{F})} \mathbb{E}\left[(X - Y)^2\right]$$

Proof. Let $Y \in L^2(\mathcal{F})$ and $W = Y - Y_0$. Then:

$$\begin{aligned}
\mathbb{E}\left[(X - Y)^2\right] &= \mathbb{E}\left[(X - Y_0)^2\right] - 2\mathbb{E}\left[W(X - Y_0)\right] + \mathbb{E}\left[W^2\right] \\
&= \mathbb{E}\left[(X - Y_0)^2\right] + \mathbb{E}\left[W^2\right] - 2\mathbb{E}\left[\mathbb{E}\left[W(X - Y_0)|\mathcal{F}\right]\right] \\
&= \mathbb{E}\left[(X - Y_0)^2\right] + \mathbb{E}\left[W^2\right] - 2\mathbb{E}\left[W(\mathbb{E}[X|\mathcal{F}] - Y_0)\right] \\
&= \mathbb{E}\left[(X - Y_0)^2\right] + \mathbb{E}\left[W^2\right]
\end{aligned}$$

□

12.3 Regular Conditional Distributions

1. Markov Kernel. Conditional Distribution of an Event. Regular Conditional Distribution.

   Definition 12.3.

   (a) *Let $(\mathcal{X}, \mathcal{F}_{\mathcal{X}})$ and $(\mathcal{Y}, \mathcal{F}_{\mathcal{Y}})$ be measurable spaces. A Markov Kernel is a family $\{\mu_x(dy) : x \in \mathcal{X}\}$ of probability measures on $(\mathcal{Y}, \mathcal{F}_{\mathcal{Y}})$ such that $\forall F \in \mathcal{F}_{\mathcal{Y}}$, $x \mapsto \mu_x(F)$ is $\mathcal{F}_{\mathcal{X}}$ measurable.*

   (b) *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{G} \subset \mathcal{F}$ be a σ-algebra. then for any real valued measurable r.v. X on $(\Omega, \mathcal{F}, \mathbb{P})$, $\mathbb{P}[X \in B|\mathcal{G}](\omega)$ is a conditional distribution of the event $\{X \in B\}$ on $\mathcal{G}$.*

   (c) *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{G} \subset \mathcal{F}$ be a σ-algebra. Let X be a r.v. as above. Then the conditional distributions*

   $$\{\mathbb{P}[X \in dy|\mathcal{G}])(\omega) : \omega \in \Omega\}$$

   is a regular conditional distribution if this is a markov kernel. (that is $\mu_\omega(F) = \mathbb{P}[X \in F|\mathcal{G}](\omega)$ a.s.)

2. Existence

   Theorem 12.2. *If X is a real-valued r.v. on $(\Omega, \mathcal{F}, \mathbb{P})$ then for any σ-algebra $\mathcal{G} \subset \mathcal{F}$ there is a regular conditional distribution of X given $\mathcal{G}$.*

   Proof. We consider events $(-\infty, q)$ where $q \in \mathbb{Q}$. That is, we want to construct distribution functions:

   $$F_\omega(q) := \mathbb{P}[X \leq q|\mathcal{G}]$$

   We consider ω for which F_ω is not a distribution function.

(a) As $q \to 0$, $F_\omega(q) \to 0$ a.s. since $\mathbb{P}[X \leq q|\mathcal{G}] \to 0$ by monotonic convergence (consider $1 - \mathbf{1}\{X \leq q\}$).
(b) As $q \to \infty$, $F_\omega(q) \to 1$ a.s. since $\mathbb{P}[X \leq q|\mathcal{G}] \to 1$ also by monotonic convergence.
(c) Consider $q < p \in \mathbb{Q}$. Then:

$$\mathbb{P}[X \leq q|\mathcal{G}] \leq \mathbb{P}[X \leq p|\mathcal{G}] \sim a.s.$$

by monotonicity.
(d) Finally, let $q_i\mathbb{Q}$ such that $q_i \downarrow q$. Then $1-\mathbf{1}\{X \leq q_i\} \uparrow 1-\mathbf{1}\{X \leq q\}$. By monotonic convergence, $\mathbb{P}[X \leq q_i|\mathcal{G}] \downarrow \mathbb{P}[X \leq q|\mathcal{G}]$ a.s.

Hence, the $\mathbb{P}[\omega : F_\omega$ is not a distribution$] = 0$. Now let

$$F_\omega(x) = \lim_{q_i \downarrow x} F_\omega(q_i)$$

where $q_i \in \mathbb{Q}$ for any $x \in \mathbb{R}$. Then F_ω are distribution functions. And so there exists a probability measure μ_ω such that

$$\mu_\omega(-\infty, x] = F_\omega(x) = \mathbb{P}[X \leq x|\mathcal{G}] \sim a.s.$$

Noting that $\{(-\infty, x] : x \in \mathbb{R}\}$ forms a π-system, we have that:

$$\mu_\omega(B) = \mathbb{P}[X \in B|\mathcal{G}]$$

for any $B \in \mathcal{B}$ by π-λ theorem. We must now show that $\omega \mapsto \mu_\omega(F)$ is measurable. This follows since $\mu_\omega(F) = \mathbb{P}[X \in F|\mathcal{G}] \in \mathcal{G}$. □

3. Change of Variables

Theorem 12.3. *Let $\{\mu_\omega(dx)\}$ be a regular conditional distribution of X given $\mathcal{G}$. And let Y be $\mathcal{G}$ measurable. Suppose $f(x, y)$ is a jointly measurable real valued function such that $\mathbb{E}[|f(X,Y)|] < \infty$. Then:*

$$\mathbb{E}[f(X,Y)|\mathcal{G}] = \int f(x, Y(\omega))\mu_\omega(dx) \sim a.s.$$

Proof. First consider $\mathbf{1}\{(X,Y) \in A \times B\}$ where $B \in \mathcal{G}$. Then:

$$\begin{aligned}
\mathbb{E}[\mathbf{1}\{(X,Y) \in A \times B\}|\mathcal{G}] &= \mathbf{1}\{Y \in B\}\mathbb{E}[\mathbf{1}\{X \in A\}|\mathcal{G}] \\
&= \mathbf{1}\{Y \in B\}\mu_\omega(A) \\
&= \mathbf{1}\{Y \in B\}\int \mathbf{1}\{x \in A\}\mu_\omega(dx) \\
&= \int \mathbf{1}\{(x, Y(\omega)) \in A \times B\}\mu_\omega(dx)
\end{aligned}$$

This holds for all measurable rectangles, and so by π-λ theorem it holds for all measurable $F \in \mathcal{B}^2$. That is when $\mathbf{1}\{(X,Y) \in F\}$. By linearity, it holds for all simple functions. By monotonicity it holds for all positive functions. Finally, by linearity it holds for all measurable functions f. □

13 Martingales

13.1 Definition and Basic Properties

1. Filtration. Adapted. Super Martingale. Sub Martingale. Martingale. Martingale Difference Sequence.

   Definition 13.1.

   (a) *A sequence of increasing σ-algebras is a filtration*

   (b) *A sequence X_n of random variables for which $X_n \in \mathcal{F}_n$ where $(\mathcal{F}_n)$ are a filtration is adapted.*

   (c) *A sequence X_n adapted to $\mathcal{F}_n$ is a super martingale (sup mg) if*

   $$\mathbb{E}\left[X_n \middle| \mathcal{F}_{n-1}\right] \leq X_{n-1}$$

   (d) *A sequence X_n adapted to $\mathcal{F}_n$ is a sub martingale (sub mg) if*

   $$\mathbb{E}\left[X_n \middle| \mathcal{F}_{n-1}\right] \geq X_{n-1}$$

   (e) *A sequence X_n adapted to $\mathcal{F}_n$ is a martingale (mg) if it is both a sup mg and sub mg.*

   (f) *Suppose X_n is a mg. Then $\xi_n = X_n - X_{n-1}$ are called martingale differences.*

2. Superhamonic Functions and sup mg

   Example 13.1. *A superharmonic function is a function f for which:*

   $$f(x) \geq \frac{1}{|B(0,r)|} \int_{B(x,r)} f(y)dy$$

   Suppose f is a superharmonic function on $\mathbb{R}^d$ and let $\xi_1, \ldots$ be i.i.d. uniform over the unit sphere and S_n be the partial sums. Then $f(S_n)$ is a sup mg.

   Proof. The second inequality follows from a change of variable and noting that S_{n-1} is known.

   $$\begin{aligned}\mathbb{E}\left[f(S_{n+1}) \middle| \mathcal{F}_n\right] &= \frac{1}{|B(0,1)|} \int_{B(0,1)} f(y + S_{n-1})dy \\ &\leq f(S_{n-1})\end{aligned}$$

   □

3. Generalization of sup mg/sub mg/mg properties

   Lemma 13.1. *Let X_n be adapted to $\mathcal{F}_n$.*

   (a) *If X_n is a sup mg then for $n > m$, $\mathbb{E}\left[X_n \middle| \mathcal{F}_m\right] \leq X_m$*

   (b) *If X_n is a sub mg then for $n > m$, $\mathbb{E}\left[X_n \middle| \mathcal{F}_m\right] \geq X_m$*

(c) *If X_n is a mg then for $n > m$, $\mathbb{E}[X_n | \mathcal{F}_m] = X_m$*

Proof. Use the Russian doll property finitely many times. □

4. Martingale Differences are Uncorrelated

   Lemma 13.2. *Let X_n be a mg w.r.t $\mathcal{F}_n$ and with difference ξ_n. If $X_0 = 0$ and $\mathbb{E}\left[X_n^2\right] < \infty$ then ξ_i are uncorrelated and*

   $$\mathbb{E}\left[X_n^2\right] = \sum_{i=1}^{n} \mathbb{E}\left[\xi_i^2\right]$$

   Proof. Note that $\xi_n = X_n - X_{n-1} \in \mathcal{F}_n$ and $\mathbb{E}[\xi_n | \mathcal{F}_m] = 0$ for $m < n$. W.L.O.G. suppose $i < j$ then

   $$\mathbb{E}[\xi_i \xi_j] = \mathbb{E}[\xi_i \mathbb{E}[\xi_j | \mathcal{F}]] = 0$$

   Therefore

   $$\mathbb{E}\left[X_n^2\right] = \sum_{i=1}^{n} \mathbb{E}\left[\xi_i^2\right] + \sum_{i<j} 2\mathbb{E}[\xi_i \xi_j] = \sum_{i=1}^{n} \mathbb{E}\left[\xi_i^2\right]$$

   □

5. Convexity and mg

   Lemma 13.3. *If X_n is a mg, ϕ is convex, and $\mathbb{E}[|\phi(X_n)|] < \infty$ for all n then $\phi(X_n)$ is a sub mg. In particular, this holds when $\phi(x) = |x|^p$ for $p \geq 1$.*

   Proof. By Jensen's Inequality:

   $$\mathbb{E}[\phi(X_n) | \mathcal{F}_{n-1}] \geq \phi(\mathbb{E}[X_n | \mathcal{F}_{n-1}]) = \phi(X_{n-1})$$

   □

6. Convexity and sub mg

   Lemma 13.4. *If X_n is a sub mg., ϕ is convex and increasing, and $\mathbb{E}[|\phi(X_n)|] < \infty$ for all n then $\phi(X_n)$ is a sub mg. In particular, this holds when $\phi(x) = (x-a)^+$. If X_n is a sup mg then $X_n \wedge a$ is a sup mg.*

   Proof. By Jensen and then sub mg property with the increasing property of ϕ:

   $$\mathbb{E}[\phi(X_n) | \mathcal{F}_{n-1}] \geq \phi(\mathbb{E}[X_n | \mathcal{F}_{n-1}]) \geq \phi(X_{n-1})$$

   Note that $(x-a)^+$ is increasing and convex. Note that $x \vee a = (x-a)^+ + a$ is increasing and convex. So when X_n is a super mg, $-X_n \vee -a$ is a sub mg, and so $X_n \wedge a$ is a sup mg. □

13.2 Martingale Transforms & Convergence Theorems

1. Predictable Sequence. Martingale Transform.

   Definition 13.2. *Let $\mathcal{F}_n$ be a filtration.*

   (a) *A sequence of random variables H_n adapted to $\mathcal{F}_{n-1}$ is called a Predictable Sequence*

   (b) *Given a predictable sequence H_n and a sequence X_n adapted to $\mathcal{F}_n$,*

   $$(H \cdot X)_n = \sum_{i=1}^{n} H_i(X_i - X_{i-1})$$

   is called a Martingale transform.

2. Generating sub mg/sup mg/mg martingale transforms

   Lemma 13.5. *Let $H_n \geq 0$ be a predictable, bounded sequence.*

   (a) *If X_n is a sub mg then $(H \cdot X)_n$ is a sub mg*

   (b) *If X_n is a sup mg then $(H \cdot X)_n$ is a sup mg*

   (c) *If X_n is a mg then $(H \cdot X)_n$ is a mg*

   Proof. Since each $H_n \leq c_n$ for $n \geq 0$,

   $$\mathbb{E}\left[|(H \cdot X)_n|\right] \leq \sum_{i=1}^{n} c_i(\mathbb{E}\left[|X_i|\right] + \mathbb{E}\left[|X_{i-1}|\right]) < \infty$$

   If X_n is a sub mg then:

   $$\begin{aligned}\mathbb{E}\left[(H \cdot X)_{n+1} | \mathcal{F}_n\right] &= (H \cdot X)_n + H_n(\mathbb{E}\left[X_{n+1} | \mathcal{F}_n\right] - X_n) \\ &\geq (H \cdot X)_n\end{aligned}$$

   If X_n is a sup mg, consider $-X_n$, which is a sub mg. The result then holds. For the mg case, this holds because it is both a sup and sub mg. □

3. Optionally Stopped sub mg/sup mg/mg using martingale transforms

   Corollary 13.1. *Let N be a stopping time. If X_n is a sub mg/sup mg/mg then $X_{N \wedge n}$ is a sub mg/sup mg/mg respectively.*

   Proof. Note $X_{N\wedge n} - X_0 = \sum_{i=1}^{n} \mathbf{1}\{N > i\}(X_i - X_{i-1})$. Since $\mathbf{1}\{N > i\} \in \mathcal{F}_{i-1}$, this is a martingale transform. We can apply the previous lemma to conclude. □

4. Upcrossings

   (a) Suppose X_n is a sub mg, $a < b$ and $N_0 = -1$. Define the stopping times

   $$N_{2k-1} = \inf\{n > N_{2k-2} : X_n \leq a\} \quad N_{2k} = \inf\{n > N_{2k-1} : X_n \geq b\}$$

(b) We are interested in the number of times the martingale crosses from below a to above b. That is, the number of upcrossings, $U_n = \sup k : N_{2k} \leq n$.

(c) We can upper-bound U_n since

$$U_n(b-a) \leq \sum_{i=1}^{n} \mathbf{1}\left\{N_{2k-1} < i \leq N_{2k},\ \text{for some } k\right\}(X_i - X_{i-1})$$

(d) The right hand side is a martingale transform since $\{N_{2k-1} < i\} \in \mathcal{F}_{i-1}$ and $\{N_{2k} \geq i\} \in \mathcal{F}_{i-1}$.

(e) Doob's Upcrossing Lemma

Lemma 13.6.

$$\mathbb{E}\left[U_n\right](b-a) \leq \mathbb{E}\left[(X_n - a)^+\right] - \mathbb{E}\left[(X_0 - a)^+\right]$$

Proof. Define $Y_m = (X_m - a)^+ + a$, which is a sub mg. Letting $K_n = 1 - H_n$ where $H_n = \mathbf{1}\left\{N_{2k-1} < i \leq N_{2k},\ \text{for some } k\right\}$:

$$Y_m - Y_0 = (K \cdot Y)_m + (H \cdot Y)_m$$

Since $(K \cdot Y)_m$ is a sub-martingale, $\mathbb{E}\left[(K \cdot Y)_m\right] \geq \mathbb{E}\left[(K \cdot Y)_1\right] \geq 0$. And note that $(H \cdot Y)_m \geq U_m(b-a)$. The result follows. □

5. Submartingale Convergence Theorem

Theorem 13.1. *Let X_n be a sug mg such that $\sup_n \mathbb{E}\left[X_n^+\right] < \infty$. Then $\lim_n X_n =: X$ exists and is integrable.*

Proof. Let $a < b$ be rational numbers. From the upcrossing lemma and the upper bound on the supremum:

$$\mathbb{E}\left[U_m\right] \leq \frac{1}{(b-a)} \mathbb{E}\left[(X_n - a)^+\right] \leq C < \infty$$

Using monotone convergence theorem, $\mathbb{E}\left[U_\infty\right] < \infty$. Hence $U < \infty$ a.s. And since this holds for all $a < b$ rational:

$$\mathbb{P}\left[\bigcup_{a,b \in \mathbb{Q}} \{\liminf_n X_n < a < b < \limsup_n X_n\right] = \sum_{a,b \in \mathbb{Q}} 0 = 0$$

Therefore, $\liminf_n X_n = \limsup_n X_n$ a.s. Let X denote this limit. By Fatou's Lemma:

$$\mathbb{E}\left[X^+\right] \leq \liminf \mathbb{E}\left[X_n^+\right] < \infty$$

And:

$$\mathbb{E}\left[X^-\right] \leq \liminf \mathbb{E}\left[X_n^-\right] = \liminf \mathbb{E}\left[X_n^+\right] - \mathbb{E}\left[X_n\right] \leq \sup_n \mathbb{E}\left[X_n^+\right] - \mathbb{E}\left[X_0\right]$$

□

6. Supermartingale Convergence Theorem

Theorem 13.2. *Suppose $X_n \geq 0$ is a sup mg. Then $\lim_n X_n =: X$ exists and $\mathbb{E}[X] \leq \mathbb{E}[X_0]$.*

Proof. $-X_n \leq 0$ is a sub mg bounded above. Submartingale Convergence Theorem, gives the existence of X and Fatou's lemma gives the bound on the expectation. □

13.3 Applications

1. Bounded Increments

Theorem 13.3. *Let $X_1, \ldots$ be a mg with $|X_{n+1} - X_n| \leq M < \infty$. Then for:*

(a) $C = \{\lim X_n$ exists and is finite$\}$

(b) $D = \{\limsup_n X_n = \infty, \liminf_n X_n = -\infty\}$

$\mathbb{P}[C \cup D] = 1$

Proof. We need to show that on $A = \{\liminf_n X_n > -\infty\}$ and $B = \{\limsup X_n < \infty\}$ that the $\lim X_n$ exists and is finite. Let $0 < K < \infty$ and consider the stopping time $N(K) = \inf\{n : X_n \leq -K\}$. Then $X_{N(K)\wedge n} + K + M \geq 0$ and is a mg. By sup mg convergence, $\lim_n X_{N(K)\wedge n}$ exists, specifically when $N(K) = \infty$. Since K is arbitrary, we have that the limit exists for

$$\bigcup_{K>0} \{N(K) = \infty\} = \{\liminf_n X_n > 0\}$$

For the other direction, we can consider $-X_n$ and the result follows. □

2. Generalization of Second Borel Cantelli Lemma

Theorem 13.4. *Let $\mathcal{F}_n$ be a filtration with $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $A_n \in \mathcal{F}_n$. Then $\{A_n$ i.o.$\} = \{\sum \mathbb{P}[A_n|\mathcal{F}_{n-1}] = \infty\}$.*

Proof. Consider $Z_n = \sum_{i=1}^n \mathbf{1}\{A_i\} - \mathbb{P}[A_i|\mathcal{F}_{i-1}]$. Then Z_n is a martingale and has bounded increments. Therefore:

(a) On C, $\lim Z_n$ exists and is finite. Therefore, $\sum \mathbf{1}\{A_n\} = \infty$ if and only if $\sum \mathbb{P}[A_n|\mathcal{F}_{n-1}] = \infty$.

(b) On D, $\limsup Z_n = \infty$ and $\liminf Z_n = -\infty$, hence $\sum \mathbf{1}\{A_n\} = \infty$ and $\sum \mathbb{P}[A_n|\mathcal{F}_n] = \infty$.

□

3. Branching Process

(a) Galton Watson Process

Definition 13.3. *Let ξ_i^n for $i, n \geq 1$ be i.i.d., non-negative, integer valued random variables. Define $Z_0 = 1$ and*

$$Z_{n+1} = \begin{cases} \sum_{i=1}^{Z_n} \xi_i^{n+1} & Z_n > 0 \\ 0 & Z_n = 0 \end{cases}$$

Then (Z_n) is called a Galton Watson Process.

(b) Martingale

Lemma 13.7. *If $\mathbb{E}\left[\xi_1^1\right] = \mu$ then Z_n/μ^n is a mg.*

Proof. $\mathbb{E}\left[Z_{n+1} | Z_n\right] = \sum_{i=1}^{\infty} \mathbf{1}\left\{Z_n > i\right\} \mathbb{E}\left[\xi_i^{n+1}\right] = \mu Z_n$ □

(c) Convergence when $\mu < 1$

Lemma 13.8. *If $\mu < 1$ then $Z_n \to 0$ a.s.*

Proof. Since $\mathbb{E}\left[Z_0\right] = 1$, $\mathbb{E}\left[Z_n\right] = \mu^n \to 0$. Hence, $Z_n \to 0$ a.s. □

(d) Convergence when $\mu = 1$

Lemma 13.9. *If $\mu = 1$ and $\mathbb{P}\left[\xi_1^1 = 1\right] < 1$ then $Z_n \to 0$ a.s.*

Proof. We have that Z_n is a martingale which is non-negative. By super mg convergence theorem $\lim Z_n =: Z_\infty$ exists. Consider when $Z_\infty = k$. Then for sufficiently large m, $Z_m = k$, but the probability of this is 0. Hence, $Z_\infty = 0$. □

13.4 Uniform Integrability

1. Uniformly Integrable

Definition 13.4. *A collection of random variables $\{X_i, i \in I\}$ is uniformly integrable (u.i.) if*

$$\lim_{M \to \infty} \sup_{i \in I} \mathbb{E}\left[|X_i| \mathbf{1}\left\{|X_i| > M\right\}\right] = 0$$

2. Canonical Examples

(a) Uniformly Dominated Random Variables

Lemma 13.10. *Suppose $|X_i| \leq Y$ for all $i \in I$ and $\mathbb{E}[Y] < \infty$. Then X_i are u.i.*

Proof. Let $M > 0$

$$\sup_{i \in I} \mathbb{E}\left[|X_i| \mathbf{1}\left\{|X_i| > M\right\}\right] \leq \mathbb{E}\left[Y \mathbf{1}\left\{Y > M\right\}\right]$$

Since Y is integrable, as $M \to \infty$, the right hand side goes to 0. □

(b) Generated Family of Conditionals

Lemma 13.11. *Let $X \in L^1(\mathcal{F})$. Then $\{\mathbb{E}\left[X | \mathcal{F}\right] : \mathcal{F} \subset \mathcal{F}_0\}$ is u.i.*

Proof. By Jensen's inequality and the definition of conditional expectation:

$$\mathbb{E}\left[|\mathbb{E}\left[X|\mathcal{F}\right]|\mathbf{1}\left\{|\mathbb{E}\left[X|\mathcal{F}\right]|>M\right\}\right]=\mathbb{E}\left[|X|\mathbf{1}\left\{|\mathbb{E}\left[X|\mathcal{F}\right]|>M\right\}\right]$$

To uniformly bound the domain of integration, we use Chebyshev's Inequality and Jensen:

$$\mathbb{P}\left[|\mathbb{E}\left[X|\mathcal{F}\right]|>M\right]\leq\frac{\mathbb{E}\left[|X|\right]}{M}$$

So for any $\epsilon>0$ there is a δ and M such that $\mathbb{E}\left[|X|\right]/M<\delta$ and so $\forall\mathcal{F}$:

$$\mathbb{E}\left[|X|\mathbf{1}\left\{|\mathbb{E}\left[X|\mathcal{F}\right]|>M\right\}\right]\leq\epsilon$$

□

(c) Checking u.i.

Lemma 13.12. *Let $\phi\geq 0$ be an function such that $\phi(x)/x\to\infty$ as $x\to\infty$. If $\mathbb{E}\left[\phi(|X_i|)\right]\leq C$ for all $i\in I$ then $\{X_i:i\in I\}$ is u.i.*

Proof. For all $z>1$, $\exists M>1$ such that if $|X_i|>M$ then

$$\phi(|X_i|)/|X_i|>z$$

Therefore:

$$\mathbb{E}\left[|X_i|\mathbf{1}\left\{|X_i|>M\right\}\right]\leq\frac{1}{z}\mathbb{E}\left[\phi(|X_i|)\mathbf{1}\left\{|X_i|>M\right\}\right]\leq\frac{C}{z}$$

Letting $z\to\infty$ we have that $M\to\infty$ and the result follows. □

3. Convergence in L^1 and u.i.

Theorem 13.5. *If $X_n\to X$ in probability then the following are equivalent:*

(a) X_n *are u.i.*
(b) $X_n\to X$ *in* L^1
(c) $\mathbb{E}\left[|X_n|\right]\to\mathbb{E}\left[|X|\right]<\infty$

Proof. We first start with (a) $\implies$ (b). Let $\epsilon>0$ and find $M>0$ such that:

$$\sup_n\mathbb{E}\left[|X_n|\mathbf{1}\left\{|X_n|>M\right\}\right]\leq\epsilon$$

and let $Y_n=X_n\mathbf{1}\left\{|X_n|\leq M\right\}$. Hence, by dominated convergence theorem:

$$\mathbb{E}\left[|Y_n-Y|\right]\to 0$$

Now:

$$|X_n-X|\leq|Y_n-Y|+|X_n|\mathbf{1}\left\{|X_n|>M\right\}+|X|\mathbf{1}\left\{|X|>M\right\}$$

Taking expectation, we have already bounded the first term and second term. The last term follows by Fatou's lemma. To show (b) $\implies$ (c) is a consequence of the reverse triangle inequality. To show (c) $\implies$ (a) we consider the following function:

$$\phi_M(x) = \begin{cases} x & x \in [0, M-1] \\ linear & x \in [M-1, M] \\ 0 & x \in [M, \infty] \end{cases}$$

Taking $\epsilon > 0$, for $i > n_0$:

$$\mathbb{E}\left[|X_i| \mathbf{1}\left\{|X_i| > M\right\}\right] \leq \mathbb{E}\left[|X_i|\right] - \mathbb{E}\left[\phi_M(|X_i|)\right] \leq \mathbb{E}\left[|X|\right] - \mathbb{E}\left[\phi_M(|X|)\right] + \epsilon$$

since ϕ_M is continuous and bounded. Since $\mathbb{E}\left[|X|\right]$ is integrable for sufficiently large M the right hand side is arbitrarily small. For $0 \leq i \leq n_0$, we can choose M large enough so that the bound will hold uniformly over all i. □

13.5 Uniform Integrability and Martingale Convergence in L^1

1. Sub mg in L^1

   Corollary 13.2. *For a sub mg, X_n, the following are equivalent.*

   (a) X_n are u.i.

   (b) X_n converge in L^1 and a.s.

   (c) X_n converge in L^1.

   Proof. To show (a) $\implies$ (b), we first note that $\exists M > 0$ such that:

   $$\sup_n \mathbb{E}\left[|X_n| \mathbf{1}\left\{|X_n| > M\right\}\right] < 1$$

   Hence, $\sup_n \mathbb{E}\left[|X_n|\right] \leq M + 1$. Hence, applying the sub mg. convergence theorem we have a.s. convergence. By **Theorem 13.5** convergence in L^1 holds. (b) $\implies$ (c) trivially. To show (c) $\implies$ (a), we need to demonstrate that X_n converges in probability. But this is implied by Chebyshev and L^1. □

2. mg in L^1 and Generated Family of Conditionals

   Corollary 13.3. *For a mg, X_n, the following are equivalent.*

   (a) X_n are u.i.

   (b) X_n converge in L^1 and a.s.

   (c) X_n converge in L^1

   (d) $\exists X \in L^1$ such that $X_n = \mathbb{E}\left[X | \mathcal{F}_n\right]$.

Proof. (a) $\implies$ (b) $\implies$ (c) follows from the sub mg case. To prove that (d) $\implies$ (a), let $A \in \mathcal{F}_n$. Then $|\mathbb{E}[X_n \mathbf{1}\{A\}] - \mathbb{E}[X \mathbf{1}\{A\}]| \leq \mathbb{E}[|X_n - X|] \to 0$. Hence, $\mathbb{E}[X | \mathcal{F}_n] = X_n$ a.s. To show that (d) $\implies$ (a), this follows because X_n are a generated family of conditionals. □

3. Families Generated by Increasing Filtrations

   (a) Convergence over Increasing Filtration

   Lemma 13.13. *Suppose $\mathcal{F}_n \uparrow \mathcal{F}_\infty$, where $\mathcal{F}_\infty = \sigma(\cup_n \mathcal{F}_n)$. Then $\mathbb{E}[X | \mathcal{F}_n] \to \mathbb{E}[X | \mathcal{F}_\infty]$ in L^1 and a.s. for any $X \in L^1$.*

   Proof. By construction $\mathbb{E}[X | \mathcal{F}_n]$ are uniformly integrable and are a m.g. Hence, they converge to some Z in L^1 and a.s. by **Corollary 13.3**. Note that $Z \in \mathcal{F}_\infty$. Also, for any $A \in \mathcal{F}_n$:

   $$\mathbb{E}[Z \mathbf{1}\{A\}] = \mathbb{E}[X_n \mathbf{1}\{A\}] = \mathbb{E}[X \mathbf{1}\{A\}]$$

   This holds for all A in the π-system $\bigcup_n \mathcal{F}_n$. By π-λ theorem, it extends to $\mathcal{F}_\infty$. □

   (b) Levy's 0-1 Law

   Corollary 13.4. *If $\mathcal{F}_n \uparrow \mathcal{F}_\infty$ and $A \in \mathcal{F}_\infty$ then $\mathbb{P}[A | \mathcal{F}_n] \to \mathbf{1}\{A\}$.*

   (c) Kolmogorov's 0-1 Law

   Corollary 13.5. *If $X_1, \ldots,$ are i.i.d. random variables and $A \in \mathcal{T}$ then $\mathbb{P}[A] \in \{0, 1\}$.*

4. Dominated Convergence Theorem

   Theorem 13.6. *Suppose $|Y_n| \leq Z$ a.s. and $\mathbb{E}[Z] < \infty$. Suppose $\mathcal{F}_n \uparrow \mathcal{F}_\infty$.*

   (a) If $Y_n \to Y$ a.s. then $\mathbb{E}[Y_n | \mathcal{F}_n] \to \mathbb{E}[Y | \infty]$ a.s.

   (b) If $Y_n \to Y$ in L^1 then $\mathbb{E}[Y_n | \mathcal{F}_n] \to \mathbb{E}[Y | \infty]$ in L^1

   Proof. For the first claim, let $W_M = \sup\{|Y_n - Y_m| : n, m \geq M\}$. Then $W_M \leq 2Z$ and so $\mathbb{E}[W_M] < \infty$. Then:

   $$|\mathbb{E}[Y_n | \mathcal{F}_n] - \mathbb{E}[Y | \mathcal{F}_n]| \leq \mathbb{E}[|Y_n - Y| | \mathcal{F}_n] \leq 2\mathbb{E}[W_M | \mathcal{F}_n]$$

   The last term goes to $2\mathbb{E}[W_M | \mathcal{F}_\infty]$. Hence, by monotone convergence theorem, as $M \to \infty$, $W_M \downarrow 0$. Therefore:

   $$\lim_n \mathbb{E}[Y_n | \mathcal{F}_n] = \lim_n \mathbb{E}[Y | \mathcal{F}_n] = \mathbb{E}[Y | \mathcal{F}_\infty]$$

   For the second claim:

   $$\mathbb{E}[|\mathbb{E}[Y_n | \mathcal{F}_n] - \mathbb{E}[Y | \mathcal{F}_n]|] \leq \mathbb{E}[|Y_n - Y|] \to 0$$

   Hence: $\mathbb{E}[|\mathbb{E}[Y_n | \mathcal{F}_n] - \mathbb{E}[Y | \mathcal{F}_\infty]|] \to 0$. □

13.6 Optional Stopping Theorems

1. Bounded Stopping Times

   Theorem 13.7. *Let N be a stopping time.*

   (a) If X_n is a sub mg, then $\mathbb{E}[X_0] \leq \mathbb{E}[X_{n\wedge N}] \leq \mathbb{E}[X_n]$

   (b) If X_n is a sup mg, then $\mathbb{E}[X_0] \geq \mathbb{E}[X_{n\wedge N}] \geq \mathbb{E}[X_n]$

   (c) If X_n is a mg then $\mathbb{E}[X_0] = \mathbb{E}[X_{n\wedge N}] = \mathbb{E}[X_n]$

   Proof. Note that $n \wedge N$ is a bounded stopping time (hence the name). Let X_n be a sub mg. And we have that:

   $$X_{n\wedge N} - X_0 = \sum_{i=1}^{n} \mathbf{1}\{N > i\}(X_i - X_{i-1})$$

   is a sub mg. Taking expectation, we have that $\mathbb{E}[X_{n\wedge N}] \geq \mathbb{E}[X_0]$. Now:

   $$X_n - X_{n\wedge N}s = \sum_{i=1}^{n} \mathbf{1}\{N \leq i-1\}(X_i - X_{i-1})$$

   Taking expectation, $\mathbb{E}[X_n] \geq \mathbb{E}[X_{n\wedge N}]$. For sup mg, take $-X_n$ as the sub mg. □

2. Bounded Stopping Times and UI Sub mg

   Lemma 13.14. *Let N be a stopping time. If X_n is a u.i. submartingale then $X_{n\wedge N}$ is u.i.*

   Proof.

   $$\begin{aligned}\mathbb{E}[|X_{n\wedge N}|\mathbf{1}\{|X_{n\wedge N}| \geq M\}] \leq \mathbb{E}[|X_n|\mathbf{1}\{|X_n \geq M\}\mathbf{1}\{N > n\}] \\ + \mathbb{E}[|X_N|\mathbf{1}\{|X_N \geq M\}]\end{aligned}$$

   The first term tends to 0 since X_n are u.i. Since $X_{N\wedge n}$ is also a sub mg and $\sup \mathbb{E}\left[X_{N\wedge n}^+\right] \leq \sup_n \mathbb{E}[X_n^+] < \infty$, $X_{n\wedge N} \to X_N$ and $X_N \in L^1$. Therefore the second term goes to zero as $M \to \infty$. □

3. Unbounded Stopping Times and UI Sub mg

   Theorem 13.8. *If X_n is u.i. sub mg then for any stopping time N:*

   $$\mathbb{E}[X_0] \leq \mathbb{E}[X_N] \leq \mathbb{E}[X_\infty]$$

   Proof. Since X_n is a u.i. submg, then $X_{n\wedge N}$ is a u.i. sub mg.. Therefore, $X_n \to X_\infty$ and $X_{n\wedge N} \to X_N$ in L^1. The result follows from the Bounded Stopping Time result. □

4. Unbounded Stopping Times and Sup mg

Theorem 13.9. *If $X_n \geq 0$ sup mg, and N is a stopping time, then* $\mathbb{E}[X_0] \geq \mathbb{E}[X_N]$

Proof. Note that $X_{n\wedge N}$ is a non-negative sup mg and so it converges to X_N. And by Fatou's lemma:

$$\mathbb{E}[X_0] = \mathbb{E}[X_{N\wedge 0}] \geq \mathbb{E}[X_{N\wedge n}]$$

□

5. Optional Stopping Time

Theorem 13.10. *If $M \geq L$ are stopping times and $Y_{M\wedge n}$ is a u.i. sub mg then $\mathbb{E}[Y_L] \leq \mathbb{E}[Y_M]$ and $Y_L \leq \mathbb{E}[Y_M | \mathcal{F}_L]$.*

Proof. Note that the first result follows from unbounded stopping times where L is N and $M \wedge n$ take sthe place of n. Let $A \in \mathcal{F}_L$. Then:

$$\mathbb{E}[Y_L \mathbf{1}\{A\}] \leq \mathbb{E}[Y_M \mathbf{1}\{A\}] = \mathbb{E}[\mathbb{E}[Y_M | \mathcal{F}_L] \mathbf{1}\{A\}]$$

Taking $A_\epsilon = \{Y_L - \mathbb{E}[Y_M | \mathcal{F}_L] > \epsilon > 0\}$, we have that $\mathbb{P}[A_\epsilon] = 0$. □

6. Application: Wald's Equations

Lemma 13.15. *Suppose ξ_i are i.i.d. and τ is a stopping time.*

(a) If $\mathbb{E}[|\xi_i|] < \infty$ and $\mathbb{E}[\tau] < \infty$ then $\mathbb{E}[|S_\tau|] < \infty$ and

$$\mathbb{E}[S_\tau] = \mathbb{E}[\tau]\,\mathbb{E}[\xi_i]$$

(b) If $\mathbb{E}[\xi_i] = 0$, $\mathbb{E}[\xi_i^2] = \sigma^2 < \infty$, and $\mathbb{E}[\tau] < \infty$ then $\mathbb{E}[S_\tau^2] = \sigma^2 \mathbb{E}[\tau]$

(c) If $\mathbb{E}[\exp \theta \xi_i] =: \exp(-\psi(\theta)) < \infty$ then for $\tau < \infty$,

$$\mathbb{E}[\theta S_\tau - \tau\phi(\theta)] = 1$$

Proof. For the first identity, we note that $S_{\tau\wedge n} - \mu(\tau \wedge n)$ is a mg. Letting $T_n = \sum_{i=1}^n |\xi_i|$ and $\mathbb{E}[|\xi_i|] = \bar{\mu}$, $T_{\tau\wedge n} - \bar{\mu}(\tau\wedge n)$ is a mg. Then for bounded stopping time $\tau \wedge n$:

$$\mathbb{E}[T_{\tau\wedge n}] = \bar{\mu}\mathbb{E}[\tau \wedge n]$$

By monotone convergence theorem: $\mathbb{E}[T_\tau] = \bar{\mu}\mathbb{E}[\tau] < \infty$. Since $|S_{\tau\wedge n}| \leq T_\tau$, $S_{\tau\wedge n}$ are uniformly integrable and so the result follows from **Corollary 13.3**.

For the second identity, note that $S_{\tau\wedge n}^2 - \sigma^2(\tau \wedge n)$ is a mg, and so:

$$\mathbb{E}[S_{\tau\wedge n}] = \sigma^2 \mathbb{E}[\tau \wedge n]$$

We must show that $S_{\tau\wedge n}$ is u.i., which we can do by proving bounded increments. (If this is possible of course).

The third follows since $\tau \leq C < \infty$ from the bounded stopping time result. □

7. Application: Bounded Increments

Lemma 13.16. *Suppose X_n is a sub mg and N is a stopping time such that $\mathbb{E}[N] < \infty$. If $\mathbb{E}[|X_{n+1} - X_n| \mathcal{F}_n] \le B$ a.s. then $X_{N\wedge n}$ is u.i. and so $\mathbb{E}[X_N] \ge \mathbb{E}[X_0]$.*

Proof.

$$\begin{aligned}
\mathbb{E}[|X_{N\wedge n}|] &\le \mathbb{E}[|X_0|] + \sum_{i=1}^{\infty} \mathbb{E}[\mathbf{1}\{N > i\} |X_i - X_{i-1}|] \\
&\le \mathbb{E}[|X_0|] + \sum_{i=1}^{\infty} \mathbb{E}[\mathbf{1}\{N > i\} \mathbb{E}[|X_i - X_{i-1}| \mathcal{F}_{i-1}]] \\
&\le \mathbb{E}[|X_0|] + B \sum_{i=1}^{\infty} \mathbb{P}[N > i] \\
&\le \mathbb{E}[|X_0|] + B\mathbb{E}[N] < \infty
\end{aligned}$$

Note that we have just shown that:

$$|X_{N\wedge n}| \le |X_0| + \sum_{i=1}^{\infty} \mathbf{1}\{N > i\} |X_i - X_{i-1}|$$

and the right hand side is integrable. Therefore $X_{N\wedge n}$ is dominated and so it is u.i. Therefore it converges in L^1 and since $\mathbb{E}[X_{n\wedge N}]$ is an increasing sequence:

$$\mathbb{E}[X_N] \ge \mathbb{E}[X_0]$$

□

13.7 Maximal Inequalities

1. Doob's Maximal Inequality

Theorem 13.11. *Let X_m be a sub mg. Let $Y_n = \max_{0\le m\le n} X_m^+$. Let $\lambda > 0$ then:*

$$\lambda \mathbb{P}[Y_n \ge \lambda] \le \mathbb{E}\left[X_n^+ \mathbf{1}\{Y_n \ge \lambda\}\right] \le \mathbb{E}\left[X_n^+\right]$$

Proof. Let $N = \inf\{n : X_m \ge \lambda\}$. Then $N \wedge n$ is a bounded stopping time. Moreover:

$$\{Y_n \ge \lambda\} = \{N \le n\}$$

Therefore:

$$\lambda \mathbb{P}[Y_n \ge \lambda] \le \mathbb{E}\left[X_N^+ \mathbf{1}\{N \le n\}\right] \le \mathbb{E}\left[X_n^+ \mathbf{1}\{N \le n\}\right]$$

□

2. Kolmogorov's Maximal Inequality

Corollary 13.6. *Let $X_i \in L^2$ be independent, mean zero, and S_n be its partial sums. Then:*

$$\lambda^2 \mathbb{P}\left[\max_{0 \leq m \leq n} |S_n| \geq \lambda\right] \leq \mathbb{V}\left[S_n^2\right]$$

Proof. Consider S_n^2, which is a sub mg. By Doob's Maximal Inequality:

$$\lambda^2 \mathbb{P}\left[\max_{0 \leq m \leq n} |S_m| > \lambda\right] \leq \mathbb{E}\left[S_n^2\right]$$

□

13.8 Backwards/Reverse Martingales

1. Backwards Martingale

   Definition 13.5. *Let $\mathcal{F}_n$ for $n \leq 0$ be a filtration (that is, $\mathcal{F}_n \subset \mathcal{F}_{n+1}$. Let X_n, $n \leq 0$ be adapted to $\mathcal{F}_n$. Then X_n is a backwards mg if $\mathbb{E}\left[X_{n+1} \middle| \mathcal{F}_n\right] = X_n$.*

2. Convergence of Backwards mg

   Theorem 13.12. *$X_{-\infty} = \lim_{n \to -\infty} X_n$ exists a.s. and in L^1.*

   Proof. Note that $X_n = \mathbb{E}\left[X_0 \middle| \mathcal{F}_n\right]$. Therefore, X_n are a generated family of conditionals, and so are u.i.. Therefore, they converge a.s. and in L^1. □

3. Decreasing Sequence of Filtrations

   Theorem 13.13. *If $X_{-\infty} = \lim_{n \to -\infty} X_n$ and $\mathcal{F}_{-\infty} = \bigcap_n \mathcal{F}_{-n}$ then $X_{-\infty} = \mathbb{E}\left[X_0 \middle| \mathcal{F}_{-\infty}\right]$*

   Proof. Let $A \in \mathcal{F}_{-\infty}$. Then for any n, $A \in \mathcal{F}_n$. So by L^1 convergence:

   $$\mathbb{E}\left[X_0 \mathbf{1}\{A\}\right] = \mathbb{E}\left[X_n \mathbf{1}\{A\}\right] \to \mathbb{E}\left[X_{-\infty} \mathbf{1}\{A\}\right]$$

   By the definition of conditional expectation, the result follows. □

4. Reversed Levy's 0-1 Law

   Corollary 13.7. *Let $\mathcal{F}_n$, $n \leq 0$ be a filtration and $\mathcal{F}_{-\infty} = \bigcap_n \mathcal{F}_n$. If $A \in \mathcal{F}_{-\infty}$ then $\mathbb{P}\left[A | \mathcal{F}_n\right] \to \mathbf{1}\{A\}$*

5. Backwards Dominated Convergence

   Corollary 13.8. *Suppose $|Y_n| \leq Z$ for all Y_n, $Y_n \to Y$ a.s. and $\mathbb{E}[Z] < \infty$. Let $\mathcal{F}_{-\infty} = \bigcap_n \mathcal{F}_n$ where $\mathcal{F}_n, n \leq 0$ is a filtration. Then $\mathbb{E}\left[Y_n \middle| \mathcal{F}_n\right] \to \mathbb{E}\left[Y \middle| \mathcal{F}_{-\infty}\right]$.*

Proof. Let $W_M = \sup\{|Y_n - Y_m| : n, m \geq M\}$. Then $\mathbb{E}[W_M] \leq 2\mathbb{E}[Z] < \infty$. Consider

$$\mathbb{E}[|Y_n - Y||\mathcal{F}_n] \leq 2\mathbb{E}[W_M|\mathcal{F}_n] \to 2\mathbb{E}[W_M|\mathcal{F}_{-\infty}]$$

By monotone convergence theorem, as $M \to \infty$, $W_M \to 0$. So the result follows. □

Note 13.1. *Convergence of* $Y_n \to Y$ *in* L^1 *implies* $\mathbb{E}[Y_n|\mathcal{F}_n] \to \mathbb{E}[Y|\mathcal{F}_{-\infty}]$ *in* L^1 *similarly.*

Part V

Canonical Processes

14 Exchangeable Sequences

1. Finite Permutation. Permutable.

   Definition 14.1. *A finite permutation of $\mathbb{N}$ is a function $\pi : \mathbb{N} \to \mathbb{N}$ such that $\pi(i) \neq i$ for only finitely many i. Note for $\omega \in S^{\mathbb{N}}$, we denote $(\pi\omega)_i = \omega_{\pi(i)}$. An event A is permutable if $\{\omega : \pi\omega \in A\} = \pi - 1A = A$ for all π.*

2. Exchangeable σ-algebras

   Theorem 14.1. *The collection of all n-permutable A, denoted $\mathcal{V}_n$, for all $\pi \in [n]$ is a σ-algebra. Moreover, $\mathcal{V}_{n+1} \subset \mathcal{V}_n$. Finally, the collection of all permutable event, $\mathcal{V}$, is a σ-algebra.*

   Proof. Let $S^{\mathbb{N}} \in \mathcal{V}_n$. If $A \in \mathcal{V}_n$ then for any $\pi \in [n]$, $\pi^{-1}A = A$. So $\pi^{-1}A^c = (\pi^{-1}A)^c = A^c$, so $A^c \in \mathcal{V}_n$. Finally, if $A_i \in \mathcal{V}_n$ then:

   $$\pi^{-1}\bigcup_i A_i = \bigcup \pi^{-1}A_i = \bigcup A_i$$

   Hence, $\mathcal{V}_n$ is a σ-algebra. Since $[n] \subset [n+1]$, $\mathcal{V}_{n+1} \subset \mathcal{V}_n$. Finally, $\bigcap \mathcal{V}_n = \mathcal{V}$ so the final result follows. □

3. Exchangeable SLLN

   Theorem 14.2. *Let X_n be an exchangeable sequence such that $\mathbb{E}[|X_0|] < \infty$. Let $\theta_{-n} = \frac{1}{n}\sum_{i=1}^n X_i$. Then:*

   (a) $\{\theta_n : n \leq 0\}$ is a reverse martingale with respect to $\mathcal{V}_n$, $n \leq 0$.

   (b) $\lim_{n\to\infty} \theta_n = \mathbb{E}[X_0|\mathcal{V}]$

   Proof. Note that $\mathcal{V}_n = \sigma(\theta_n, \theta_{n-1}, \ldots)$. We show that θ_n is a reverse martingale. First:

   $$\mathbb{E}[\theta_{n+1}|\mathcal{V}_n] = \mathbb{E}\left[\left.\frac{\theta_n|n| - X_n}{|n| - 1}\right| \mathcal{V}_n\right]$$

   Note that

   $$\mathbb{E}[X_n|S_n] = \frac{1}{|n|}\sum_{i=1}^n \mathbb{E}[X_i|S_n] = \theta_n$$

   Therefore, this is a backwards martingale. Hence, the second result follows. □

4. Application: i.i.d. SLLN

(a) Trivial σ-algebras

Lemma 14.1. *If $Z \in L^1$ and $\mathcal{H}$ is trivial then $\mathbb{E}[Z|\mathcal{H}] = \mathbb{E}[Z]$ a.s.*

Proof. Suppose $Z \in \mathcal{H}$. Then $\{Z \leq q\}$ for $q \in \mathbb{Q}$ have probably either 0 or 1. Therefore, Z is a.s. a constant. Since $\mathbb{E}[Z|\mathcal{H}] \in \mathcal{H}$, it is a.s. a constant and since $\mathbb{E}[Z] = \mathbb{E}[\mathbb{E}[Z|\mathcal{H}]]$ the result follows. □

(b) Hewitt-Savage 0-1 Law

Lemma 14.2. *If $\{X_n\}$ are i.i.d. then the exchangeable σ-algebra they generate, $\mathcal{E}$, is a trivial σ-algebra.*

(c) SLLN

Corollary 14.1. *Let $X_1, \ldots$ be i.i.d. and $\mathbb{E}[|X_1|] < \infty$. Then:*

$$\frac{S_n}{n} \to \mathbb{E}[X_1] \sim a.s.$$

Proof. Use Exchangeable SLLN and apply Hewitt-Savage 0-1 Law. □

15 Renewal Processes

15.1 Renewal Processes and Convergence

1. Renewal Process. Ordinary Renewal Process. Delayed Renewal Process. Renewal/Occurrence. Renewal Counting Process. Arithmetic. Non-arithmetic. Renewal Measure.

   Definition 15.1.

   (a) A renewal process is an increasing sequence of non-negative random variables, S_i, whose differences are i.i.d.

   (b) An ordinary renewal process is a renewal process for which $S_0 = 0$

   (c) A delayed renewal process is a renewal process for which S_0 is a non-negative random variable.

   (d) A renewal/occurrence is each S_i

   (e) A Renewal Counting Process is $N(t) := \max\{n : S_N \leq t\}$

   (f) A renewal process is arithmetic if $S_i - S_{i-1}$ are supported on $h\mathbb{Z}$

   (g) A renewal process is non-arithmetic if it is not arithmetic

   (h) A renewal measure, u, is defined by $u(k) = \mathbb{P}[S_n = k, n \geq 0] = \sum_{n=0}^{\infty} \mathbb{P}[S_n = k]$

2. Feller Erdos Pollard

   Theorem 15.1. *Let u be the renewal measure of an ordinary, arithmetic renewal process whose inter-occurrence time has mean $0 < \mu < \infty$ and is not supported on a proper sub-group of $\mathbb{Z}$. Then:*

$$\lim_{k \to \infty} u(k) = \frac{1}{\mu}$$

15.2 Renewal Equation

1. Renewal Equations

   Definition 15.2. *Let f_k be the probabilities of inter occurrence times of a renewal process. Let $z(m)$ and $b(m)$ be bounded sequences that satisfy the renewal equation:*

$$z(m) = b(m) + \sum_{k=1}^{m} f_k z(m-k)$$

   A second form is $z(m) = b(m) + \mathbb{E}[z(m - X_1)]$ where $z(m-k) = 0$ when $m - k < 0$.

2. Solution to the Renewal Equation

   Lemma 15.1. *The solution to the renewal equation, for an ordinary renewal process, is:*

$$z(m) = \sum_{k=0}^{\infty} b(m-k)u(k)$$

   where $b(m-k) = 0$ for $m < k$.

Proof. Using the second form, we have that:

$$z(m) = b(m) + \mathbb{E}[b(m - X_1)] + \mathbb{E}[b(m - X_1 - X_2)] + \cdots = \sum_{i=0}^{\infty} b(m - S_i)$$

Computing the form of the right hand side:

$$\begin{aligned} z(m) &= \sum_{i=0}^{\infty} b(m - S_i) \\ &= \sum_{i=0}^{\infty} \sum_{k=0}^{m} b(m-k)\mathbb{P}[S_i = k] \\ &= \sum_{k=0}^{m} b(m-k) \sum_{i=1}^{\infty} \mathbb{P}[S_i = k] \\ &= \sum_{k=0}^{m} b(m-k)u(k) \end{aligned}$$

□

3. Key Renewal Theorem

Theorem 15.2. *If $z(m)$ is the solution the renewal equation and $b(m)$ is absolutely summable, then*

$$\lim_{m \to \infty} z(m) = \frac{1}{\mu} \sum_{k=1}^{\infty} b(k)$$

Proof. Note that $z(m) = \sum_{k=0}^{m} b(m-k)u(k) = \sum_{k=0}^{m} b(k)u(m-k)$. Also $|b(k)u(m-k)| \leq |b(k)|$ and since $|b(k)|$ has finite sum over all k, by dominated convergence theorem:

$$\lim_{m} z(m) = \lim_{m} \sum_{k=0}^{\infty} b(k)u(m-k) = \sum_{k=0}^{\infty} b(k)(\lim_{m} u(m-k))$$

□

4. Applications

(a) Residual Lifetime

Example 15.1. *Let $R(m)$ be the residual life time at time m. Then there are two cases $m \geq X_1$ and $0 \leq m < X_1$. In the first case:*

$$\{R(m) = r; m \geq X_1\} = \bigcup_{k=1}^{m} \{R(m-k) = r\} \cap \{X_1 = k\}$$

In the second case:

$$\{R(m) = r; m < X_1\} = \{X_1 = r + m\}$$

Therefore:

$$\begin{aligned} z(m) :&= \mathbb{P}\left[R(m) = r\right] \\ &= \mathbb{P}\left[X_1 = m + r\right] + \sum_{k=1}^{m} \mathbb{P}\left[R(m-k) = r\right] \mathbb{P}\left[X_1 = k\right] \\ &= f_{k+r} + \sum_{k=1}^{m} z(m-k) f_k \end{aligned}$$

By the Key Renewal Theorem:

$$\lim_m z(m) = \frac{1}{\mu} \sum_{m=0}^{\infty} f_{k+r} = \frac{\mathbb{P}\left[X_1 \geq r\right]}{\mu}$$

(b) Age

Example 15.2. *Let $A(m)$ be the age of the renewal at time m. Then, again, there are two cases, $m \geq X_1$ and $0 \leq m < X_1$. In the first case:*

$$\{A(m) = r; m \geq X_1\} = \bigcup_{k=1}^{m} \{A(m-k) = r\} \cap \{\cap \{X_1 = k\}$$

In the second case:

$$\{A(m) = r; m < X_1\} = \{m < X_1\} \cap \{m = r\}$$

Therefore:

$$\begin{aligned} z(m) :&= \mathbb{P}\left[A(m) = r\right] \\ &= \mathbb{P}\left[r < X_1\right] \mathbf{1}\{m = r\} + \sum_{k=1}^{m} z(m-k) \mathbb{P}\left[X_1 = k\right] \end{aligned}$$

Since r is fixed, $\sum_{m=0}^{\infty} \mathbb{P}\left[r < X_1\right] \mathbf{1}\{m = r\} = \mathbb{P}\left[r < X_1\right]$. So by the Key Renewal Theorem:

$$\lim_m z(m) = \frac{\mathbb{P}\left[X_1 > r\right]}{\mu}$$

(c) Lifetime

Example 15.3. *Let $L(m)$ be the lifetime of a renewal at time m. That is $L(m) = A(m) + R(m)$. There are two cases, $m \geq X_1$ and $0 \leq m < X_1$. In the first case:*

$$\{L(m) = r; m \geq X_1\} = \bigcup_{k=0}^{m} \{L(m-k) = r\} \cap \{X_1 = k\}$$

In the second case:

$$\{L(m) = r; m < X_1\} = \{X_1 = r\} \cap \{m < r\}$$

Therefore:

$$\begin{aligned} z(m) :&= \mathbb{P}\left[L(m) = r\right] \\ &= \mathbb{P}\left[X_1 = r\right] \boldsymbol{1}\left\{m < r\right\} + \sum_{k=0}^{m} z(m-k)\mathbb{P}\left[X_1 = k\right] \end{aligned}$$

The Key Renewal Theorem implies:

$$z(m) = \frac{1}{\mu}\mathbb{P}\left[X_1 = r\right] \sum_{m=0}^{\infty} \boldsymbol{1}\left\{m < r\right\} = \frac{r\mathbb{P}\left[X_1 = r\right]}{\mu}$$