UPPSALA UNIVERSITET

Föreläsningsanteckningar

Sannolikhet & Martingaler

Rami Abou Zahra

Inlämningsdatum
October 22, 2024

Contents

1. σ-algebras & Measure spaces

1.1. σ-algebras.

Definition 1.1 σ-algebra

A collection of subsets Σ of a set S is called a *σ-algebra* if:

- $\varnothing \in \Sigma$
- Is an algebra:
  - Closed under complements such that for $A \in \Sigma \Rightarrow A^c = S \backslash A \in \Sigma$
  - Closed under unions such that $A, B \in \Sigma \Rightarrow A \cup B \in \Sigma$
- Closed under countably infinite unions $A_i \in \Sigma$ for $i \in \mathbb{N}$, then $\bigcup_{i=1}^{\infty} A_i \in \Sigma$

Example:
$\Sigma = \{\varnothing, S\}$ is a σ-algebra on any set S.
Another example is $\mathcal{P}(S)$, which denotes the powerset.
Another example is $S = \mathbb{N}$, then $\Sigma = \{\varnothing, \mathbb{N}, \{2k : k \in \mathbb{N}\}, \{2k+1 : k \in \mathbb{N}\}\}$

Remark:
There exists many equivalent definitions of a σ-algebra. For example, instead of the first axiom of $\varnothing \in \Sigma$, an equivalent definition could be "Σ is non-empty", since then $\exists A \in \Sigma \Rightarrow A^c \in \Sigma \Rightarrow A \cup A^c = S \in \Sigma \Rightarrow (A \cup A^c)^c = \varnothing \in \Sigma$

Remark:
Closed under unions $\Rightarrow$ closed under finite unions since $A_1, \cdots, A_n \in \Sigma \Rightarrow A_1 \cup A_2 \in \Sigma, A_1 \cup A_2 \cup A_3 = \underbrace{(A_1 \cup A_2)}_{\in \Sigma} \cup A_3$, thus by induction $A_1 \cup \cdots \cup A_n \in \Sigma$

This does *not* imply Σ is closed under countable unions.

Counter-example:
Consider $S = [0, 1) \subseteq \mathbb{R}$. Let Σ be all finite unions of disjoint sets on the form $[a, b)$ such that $0 \leq a \leq b < 1$ (if $a = b \Rightarrow \varnothing$).
First and all algebra axioms are fulfilled, but the last one is not since we cvan consider $A_n = \left[\frac{1}{n}, 1\right)$.
Then $\bigcup_{i=2}^{\infty} = (0, 1) \notin \Sigma$

An algebra Σ is an algebra in an algebraic sense.
The symmetric difference $A \triangle B = (A \backslash B) \cup (B \backslash A)$. This behaves like "+" on Σ and intersections behave like multiplication.
Just like one would expect from an algebra, the multiplication is distributive over addition, eg. $C \cap (A \triangle B) = (C \cap A) \triangle (C \cap B)$

1.2. Measures.

Let Σ be a σ-algebra on S, and let μ_0 be a function from Σ_0 to $[0, \infty] = [0, \infty) \cup \{\infty\}$, essentially a function that assigns some value to subsets of Σ.
Intuitively, a measure should increase if we measure something bigger.

Definition 1.2 Additive and σ-additive measures

A measure μ_0 is called *additive* if $\mu_0(A \cup B) = \mu_0(A) + \mu_0(B)$ where A, B are disjoint sets.

A measure μ_0 is called *σ-additive* if this holds for ocuntable unions, i.e if A_n are pairwise disjoint, then $\mu_0\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu_0(A_n)$

Remark:
We say that μ_0 is a measure if μ_0 is σ-additive and $\mu_0(\varnothing) = 0$

Example:
$S = \{1, 2, \cdots, 6\}$, $\Sigma = \mathcal{P}(S)$ and set $\mu_0(A) = \dfrac{1}{6}|A|$. Note here that $\mu_0(S) = 1$

Definition 1.3 Probability measures

All measures that sum up to 1 are called *probability measures*

Example:
$S = \mathbb{N}$, $\Sigma = \mathcal{P}(S)$ and set $\mu_0(A \in \Sigma) = |A|$. Here $\mu_0(S) = \infty$

Example:
$S = \mathbb{N}$, $\Sigma = \mathcal{P}(S)$ and set $\mu_0(A \in \Sigma) = \begin{cases} 0 & \text{if} \quad |A| < \infty \\ \infty & \text{if} \quad |A| = \infty \end{cases}$
This is an example of an additive but not σ-additive measure, since if $A_n = \{n\}$, then $\mu_0\left(\bigcup_{n=1}^{\infty} A_n\right) = \infty$, but $\sum_{n=1}^{\infty} \mu_0(A_n) = -1$

1.3. **Measure spaces.**

Definition 1.4 Measure space triplet

A *measure space* is a triplet (S, Σ, μ) where S is some set, Σ is a σ-algebra over S, and μ is a σ-additive function $\mu : \Sigma \to [0, \infty]$ such that $\mu(\varnothing) = 0$

Definition 1.5 Probability space

If $\mu(S) = 1$, then the triplet is called a *probability space.*

Example: (finite measure space)
Let $S = \{s_1, \cdots, s_k\}$ where $k \in \mathbb{N}$ be a set of outcomes. We also associate probabilities $p_1, \cdots, p_k$ to each $s_1, \cdots, s_k$ such that $\sum_i p_i = 1$. Let $\mu(A) = \sum_{s_i \in A} p_i \ \forall A \subseteq S$. If we let $\Sigma = \mathcal{P}(S)$, then (S, Σ, μ) is a measure and a probability space.

Example: (Lebesgue measure)
Let $S = \mathbb{R}$, $\Sigma = \mathcal{B}(\mathbb{R})$ be the Borel σ-algebra (smallest σ-algebra that makes open sets measureable, note that $\mathcal{B}(\mathbb{R}) \neq \mathcal{P}(\mathbb{R})$) and let μ be something measuring length on finite unions of disjoint open intervals $A = (a_1, b_1) \cup \cdots \cup (a_n, b_n)$ such that $\mu(A) = |b_1 - a_1| + \cdots + |b_n - a_n|$

This μ is called the Lebesgue measure ($\mathcal{L}$)
Restricting S to $[0, 1]$, then we have a probability measure

$$\mu = \mathcal{L}\,|_{[0,1]}\,(A) = \mathcal{L}(A \cap [0,1]) \Rightarrow ([0,1], \mathcal{B}([0,1], \mathcal{L}\,|_{[0,1]})) \quad \text{is a probability measure}$$

This is a formulation of uniform random numbers in $[0, 1]$

1.4. **Properties of measures.**
For a measure space, we have the following properties:
(1) $\mu(A \cup B) \leq \mu(A) + \mu(B)$
(2) $\mu\left(\bigcup A_i\right) \leq \sum \mu(A_i)$
(3) $\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i) - \mu(A_1 \cap A_2) - \cdots - \mu(A_{n-1} \cap A_n) + \mu(A_1 \cap A_2 \cap A_3) \cdots + (-1)^{n+1}\mu(A_1 \cap A_2 \cdots \cap A_n)$

Note that for the first two points, we have previously assumed that A, B were disjoint. This would be the case for "joint" sets.

Bevis 1.1

Consider $\mu(A) = \mu(A \backslash B \cup (A \cap B)) = \mu(A \backslash B) + \mu(A \cap B)$ and proceed. □

Remark:
For point 4, check Math Stackexchange

The idea is if we can consider some set that is measurable, we want to be able to say something about the compositions of those measurable sets so the idea is we include their subsets in the σ-algebra (in the space we set up) as well as keeping it closed in an algebraic sense.

1.5. **Monoticity of measure.**
Let (A_i) be a sequence of increasing sets in Σ such that $\varnothing \subseteq A_1 \subseteq \cdots \subseteq S$. Then:

$$\mu(A_i) = \mu(A_i \backslash A_{i-1} \cup (A_i \cap A_{i-1})) = \mu(A_i \backslash A_{i-1} \cup A_{i-1}) = \mu(A_i \backslash A_{i-1}) + \mu(A_{i-1}) \geq \mu(A_{i-1})$$

Thus, by induction, $\mu(A_1) \leq \mu(A_2) \leq \cdots$ and by monotone convergence the limit $\lim_{i\to\infty} \mu(A_i)$ exists in the extended positive real line.

Writing $A = \bigcup_{i=1}^{\infty} A_i$, we have $\mu(A) = \lim_{i\to\infty} \mu(A_i)$, this because:

$$A = A_1 \cup (A_2 \backslash A_1) \cup (A_3 \backslash A_2) \cup \cdots$$

$$\mu(A) = \mu(A_1) + \mu(A_2 \backslash A_1) + \mu(A_3 \backslash A_2) + \cdots = \lim_{n\to\infty} \sum_{i=1}^{n} \mu(A_i \backslash A_{i-1})$$

where $A_0 = \varnothing = \lim_{n\to\infty} \mu(A_n)$

A similar result holds for decreasing sets, i.e $S \supseteq A_1 \supseteq A_2 \cdots \supseteq \varnothing$
We do the limit as $A = \bigcap_{i=1}^{\infty} A_i$ and by monotone convergence $\mu(A) = \lim_{i\to\infty} \mu(A_i)$ with similar proof.

Remark:
The last set in the decreasing sets does not necessarily have to be the empty set, recall that we are dealing with intersections instead of unions.

1.6. **Generated σ-algebras.**
Given any collection of subsets of $\mathfrak{A}\subseteq \mathcal{P}(S)$, the σ-algebra *generated* by $\mathfrak{A}$ is the smallest σ-algebra that cointains $\mathfrak{A}$ is denoted by $\sigma(\mathfrak{A}) = \bigcap_{\Sigma:\sigma\text{-alg}\&\mathfrak{A}\subseteq\Sigma}$
This is sometimes denoted by $< \mathfrak{A} >$

One can verify that this is indeed a σ-algbera:
(1) $\varnothing$ is contained in all σ-algebras, so $\varnothing$ is contained in all of the intersections
(2) If $A \in \sigma(\mathfrak{A})$, then $A \in \Sigma \ \forall\ \sigma$-algebras, but then $A^c \in \Sigma \ \forall\ \sigma$-algebras $\Rightarrow A^c \in \sigma(\mathfrak{A})$
The rest of the axioms for a σ-algebras are shown in an equivalent manner as in (2)

Example: (Borel σ-algebra)
Let $\mathcal{B}(S) = \sigma$(open subsets of S) (here we mean open in a topological sence since we need S to have a notion of opened-ness).
Since we mean open in a topological sense (which is defined as the complement of a closed set), we could have used the complement of a closed set to denote the open set, but since the complement is in the σ-algebra we may as well had the equivalent definition using the closed set all together.
This leads us to $\mathcal{B}(\mathbb{R}) = \sigma(\{(a, b) : a < b, a, b \in \mathbb{R}\})$. Instead of $\mathbb{R}$, any dense set could have worked as well (such as $\mathbb{Q}$)

Example:
Let $S = \{1, 2, 3 \cdots , 10\}$, and $\mathfrak{A} = \{\{1, 2\} , \{5\}\}$.
In order to generate a σ-algebra, we just need to recursively insert things that work with the axioms. For example, we need the empty set so we chuck in the empty set. We need the complement of the empty set so we chuck in the complement to the empty set. We need the complements to all the sets in $\mathfrak{A}$, so we add those as well, as well as their intersections.
We should then be left with just enough to call it a σ-algebra, and nothing more, hence the smallest σ-algebra:

$$\sigma(\mathfrak{A}) = \{\varnothing, S, \{1, 2\} , \{5\} , \{1, 2, 5\} , \{3, 4, 5, 6, 7, 8, 9, 10\} , \{1, 2, 3, 4, 6, 7, 8, 9, 10\} , \{3, 4, 6, 7, 8, 9, 10\}\}$$

Definition 1.6 π-system

A π-*system* on a set S is a collection of subsets π such that $\varnothing \in \pi$ and if $A, B \in \pi$ then $A \cap B \in \pi$

Sats 1.1

Suppose $\mathfrak{A} \subseteq \mathcal{P}(S)$ is a π-system and suppose that μ_1, μ_2 are measures on $(S, \sigma(\mathfrak{A}))$ such that $\mu_1(A) = \mu_2(A)\ \forall A \in \mathfrak{A}$

$\Rightarrow$ Then $\mu_1 = \mu_2$ on $(S, \sigma(\mathfrak{A}))$

In other words, π-systems uniquely determine a measure.

Example:
Let $S = \mathbb{R}$, $\mathfrak{A} = \{[-\infty, a) : a \in \mathbb{R}\}$, $\sigma(\mathfrak{A}) = \mathcal{B}(\mathbb{R})$
$\mathfrak{A}$ is a π-system and have any measure is uniquely defined on $\mathfrak{A}$.
note that $\mu([-\infty, a))$ is nothing but thte cumulative distribution function of the measure μ (in terms of a).
"Measure up to a point". The following gives justification to construct measures from small collections.

Sats 1.2: Caratheodorys extension theorem

If Σ_0 is an algebra and $\mu_0 : \Sigma \to [0, \infty]$ is a σ-additive, $\exists! \quad \mu$ on $\Sigma = \sigma(\Sigma_0)$ such that $\mu(A) = \mu_0(A) \quad \forall A \in \Sigma_0$

An important consequence is that the Lebesgue measure is unique (only one notion of length on $\mathcal{B}(\mathbb{R})$) defined through sets of the form $A = (a_1, b_1) \cup \cdots \cup (a_n, b_n)$ (disjoint union of open sets)

$$\mathcal{L}(A) = |b_1 - a_1| + \cdots + |b_n - a_n|$$

2. Probability Spaces

Probability spaces are normally denoted by $(\Omega, \mathcal{E}, \mathbb{P})$ where:

- Ω is the space of realisations
- $\mathcal{E}$ is the sets of events
- $\mathbb{P}$ is the probability measure

Example:
$\Omega = \mathbb{R}$, $\mathcal{E} = \mathcal{B}(\mathbb{R})$, $\mathbb{P}(A = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2}) dx$, $A = (a, b)$
This models a normally distributed real number.

2.1. Almost sure events.

We say that an event occurs *almost surely* if $\mathbb{P}(\mathcal{E}) = 1$ (equivalently $\mathbb{P}(\mathcal{E}^c) = 0$)

Proposition:
Let $E_1, \cdots \in \mathcal{E}$ be such that $\mathbb{P}(E_i) = 1 \quad \forall i \in \mathbb{N}$
Then, $\mathbb{P}\left(\bigcap_{i=1}^{\infty} E_i\right) = 1$

Bevis 2.1

Note that since each of them have probability measure 1, their complement must have measure 0 so:

$$\mathbb{P}\left(\bigcup_{i \in \mathbb{N}} E_i^c\right) \leq \sum_{i \in \mathbb{N}} \mathbb{P}(E_i^c) 0$$

However, since:

$$0 \leq \mathbb{P}\left(\bigcup_{i \in \mathbb{N}} E_i^c\right) \leq 0 \Rightarrow \mathbb{P}\left(\bigcup_{i \in \mathbb{N}} E_i^c\right) = 0$$

$$\Rightarrow \mathbb{P}\left(\left(\bigcup_{i \in \mathbb{N}} E_i^c\right)^c\right) = 1$$

But we have de-Morgans law, i.e $\bigcup_{i \in \mathbb{N}} E_i^c = \left(\bigcap_{i \in \mathbb{N}} E_i\right)^c$, which yields:

$$\left(\left(\bigcap_{i \in \mathbb{N}} E_i\right)^c\right)^c = \bigcap_{i \in \mathbb{N}} E_i$$

$\square$

Remark:
This applies only to countable unions. If uncountable, we could consider

$$\Omega = [0, 1], \quad \Sigma = \mathcal{B}([0, 1]), \quad \mathbb{P} = \mathcal{L}\mid_{[0,1]}$$

Then $\mathbb{P}(X = x) = 0$ (where X is some randomly chosen number and x is some fixed number). Taking the complement of this event yields $\mathbb{P}(X \neq x) = 1$ so $\mathbb{P}(X \neq x : x \in \mathbb{Q}) = 1$

2.2. **Liminf and limsup.**
Recall from real analysis:

$$\left.\begin{array}{l}\lim_{n\to\infty}\sup x_n = \lim_{n\to\infty}\sup_{m\geq n} x_n \\ \lim_{n\to\infty}\inf x_n = \lim_{n\to\infty}\inf_{m\geq n} x_n\end{array}\right\} \text{Limits exists in the extended reals and the limit exists iff limsup = liminf}$$

Recall that if $\lim_{n\to\infty}\sup x_n \geq x \Leftrightarrow \exists$ a subsequence $(x_n)_k$ with limit $\geq x$ and the opposite for liminf.

There exists a similar notion for sets.
Let $E_1, \cdots$ be events (sets)

$$\left.\begin{array}{l}\lim_{n\to\infty}\inf E_n = \bigcup_{n\geq 1}\bigcap_{m\geq n} E_n \\ \lim_{n\to\infty}\sup E_n = \bigcap_{n\geq 1}\bigcup_{m\geq n} E_n\end{array}\right\}$$

Some intuition here is definitely necessary.
For the first one, we are taking intersections of less and less sets (increasing sequence of sets), then finally unions. Think of this as events that eventually will appear

For the second one, it is decreasing (because of the intersection outside), all points will occur infinitely often.

Lemma 2.1: Fatous Lemma

Let $E_1, \cdots$ be events, then:

$$\mathbb{P}\left(\liminf_n E_n\right) \leq \liminf_n \mathbb{P}(E_n)$$

Bevis 2.2: Fatous Lemma

Let $F_n = \bigcap_{m\geq n} E_m$, i.e $E_n = \bigcup_{n\in\mathbb{N}} F_n$.
Here F_n is an increasing sequence of sets, which implies $F_n \in E_m$ $\forall m \geq n$, so $\mathbb{P}(F_n) \leq \mathbb{P}(E_m)$ $\forall m \geq n$
However, this also implies that $\mathbb{P}(F_n) \leq \inf_{m\geq n}\mathbb{P}(E_m)$

F_n is increasing $\Rightarrow$ probabilities are incfreasing $\Rightarrow \lim_{n\to\infty}\mathbb{P}(F_n)$ exists

$$\Rightarrow P\left(\bigcup_n^\infty F_n\right) = P\left(\liminf_n E_n\right)$$

$$\Rightarrow \lim_{n\to\infty}\mathbb{P}(F_n) \leq \lim_{n\to\infty}\inf\mathbb{P}(E_n) \qquad \text{by } \mathbb{P}(F_n) \leq \inf_{m\geq n}\mathbb{P}(E_m)$$

This yields finally $\mathbb{P}(\lim_n\inf E_n) \leq \lim_n\inf\mathbb{P}(E_n)$, which is what we wanted to prove. □

Note:
The reverse Fatous lemma can be proved by flipping everything (signs, inequalities, infimum to supremum etc.)

Lemma 2.2: Borel-Cantelli Lemma

Let $E_1, \cdots$ be a sequence of events such that $\sum_{n=1}^\infty \mathbb{P}(E_n) < \infty$

Then $\mathbb{P}(\lim_n\sup E_n) = 0 = \mathbb{P}(\text{"infinitely many } E_n \text{ occur"})$

Bevis 2.3: Borel-Cantelli Lemma

Recall what the limsup is, i.e $\lim_n \sup E_n = \bigcap_{n\in\mathbb{N}} \underbrace{\bigcup_{m\geq n} E_m}_{G_n}$

Note here that G_n is a decreasing sequence of sets, so $\lim_{n\to\infty} \sup E_n \subseteq G_n \quad \forall m \in \mathbb{N}$ and $\mathbb{P}(\lim_{n\to\infty \sup E_n}) \leq \mathbb{P}(G_m) \quad \forall m \in \mathbb{N}$

In particular, this is bounded above by:

$$\sum_{k=m}^{\infty} \mathbb{P}(E_k) \leq \mathbb{P}(\lim_{n\to\infty} \sup E_n)$$

But $\sum_{k=m}^{\infty} \mathbb{P}(E_k) \to 0$ as $m \to \infty$ since $\sum \mathbb{P}(E_n) < \infty$, so $\mathbb{P}(\lim_n \sup E_n) = 0$ $\square$

Example: (Coin toss)
Let E_n be the event that the first n coin toss in a sequence of tosses is heads. We have $\mathbb{P}(E_n) = 2^{-n}$ (assuming a fair coin) and $\sum_{n=1}^{\infty} \mathbb{P}(E_n) = 1 < \infty$ (since $\sum_{n=1}^{\infty} 2^{-n} \to 1$)
Thus, by the Borel-Cantelli lemma, $\mathbb{P}(\lim_n \sup E_n) = 0$ (finintely many values in which E_n occurs $\Rightarrow$ the run of heads will end almost surely)

Definition 3.7 Measurable functions

Let (S, Σ, μ) be a measure space. We say that $f : S \to \mathbb{R}$ is *measurable* if all pre-images of all Borel sets are in Σ:

$$f^{-1}(A) \in \Sigma \Rightarrow A \in \mathcal{B}(\mathbb{R})$$

Note:

$$f^{-1}(A) = \{s \in S : f(s) \in A\} \in \Sigma \quad \forall A \in \mathcal{B}(\mathbb{R})$$

- $m\Sigma$ are all measurable functions with respect to Σ
- $(m\Sigma)^+$ are all non-negative measurable functions with respect to Σ
- $b\Sigma$ are all bounded measurable functions with respect to Σ

Remark:
This can be generalized as functions $f : S \to T$ where (T, Σ', ν) is a measure space.

Lemma 3.1

We have:
(1) $f^{-1}(A^c) = (f^{-1}(A))^c$
(2) $f^{-1}\left(\bigcup_i A_i\right) = \bigcup_i f^{-1}(A_i)$
(3) $f^{-1}\left(\bigcap_i A_i\right) = \bigcap_i f^{-1}(A_i)$

Bevis 3.1

We shall only prove number 2, but the rest is proved in a similar manner:

$$\text{If } x \in f^{-1}\left(\bigcup_i A_i\right) \Leftrightarrow f(x) \in \bigcup_i A_i \Leftrightarrow \exists i \text{ s.t } f(x) \in A_i$$
$$\Leftrightarrow x \in f^{-1}(A_i) \Leftrightarrow x \in \bigcup_i f^{-1}(A_i)$$

□

Proposition:
If $f : S \to \mathbb{R}$ is continuous then it must also be measurable with respect to the Borel σ-algebra $\mathcal{B}(\mathbb{R})$

Bevis 3.2

Follows from topology, since pre-images of any open set is open as well as some help using the following: □

Proposition:

If $C \subseteq \mathcal{P}(\mathbb{R})$ is a collection such that $\sigma(C) = \overbrace{\mathcal{B}(\mathbb{R})}^{\text{codomain}}$, then $f : S \to \mathbb{R}$ is measurable with respect to $\underbrace{\mathcal{B}(S)}_{\text{domain}}$ if and only if $f^{-1}(A) \in \mathcal{B}(S) \quad \forall A \in C$

Examples:
In order to check whether $f : S \to \mathbb{R}$ is measurable, it suffices to check one of these:

- $f^{-1}(A) \in \Sigma \quad \forall A$ where A is an open set
- $f^{-1}((a, b)) \in \Sigma \quad \forall a < b \in \mathbb{R}$
- $f^{-1}([-\infty, a)) \in \Sigma \quad \forall a \in \mathbb{R}$
- Anything that generates $\mathcal{B}(\mathbb{R})$

Lemma 3.2

If $f_1, f_2 : S \to \mathbb{R}$ are measurable functions, then $f_1 + f_2$ is measurable

Bevis 3.3

We want to show that addition is measurable, we can do this by considering $(f_1+f_2)^{-1}((x,\infty)) \in \Sigma$ given that f_1, f_2 are measurable of course.

Since they are individually measurable, this means that the pre-images of this open set is in Σ, i.e

$$f_1^{-1}((x,\infty)), f_2^{-1}((x,\infty)) \in \Sigma$$

We use the fact that $x < f_1(s) + f_2(s) \Leftrightarrow \exists q \in \mathbb{Q}$ such that $q < f_1(s)$ and $x - q < f_2(s)$ (this reminds of the construction of the Dedekind sets, which is justified since there must be a rational number between x and $f_1 + f_2$ since we can just decrease the denominator to make a DIY ε)

$$\Rightarrow (f_1+f_2)^{-1}((x,\infty)) = \underbrace{\bigcup_{q\in\mathbb{Q}} \underbrace{\left(\underbrace{\overbrace{f_1^{-1}((q,\infty))}^{\in\Sigma}}_{s \text{ s.t } q < f_s(s)} \cap \underbrace{\overbrace{f_2^{-1}((q,\infty))}^{\in\Sigma}}_{x-q<f_2(s)} \right)}_{\in\Sigma}}_{\in\Sigma}$$

$\square$

Remark:

$\underbrace{f_1 \circ f_2}_{\text{"multiplication"}}$ is measurable by a similar proof. In fact, any infinite linear combination is measurable.

Lemma 3.3

Compositions of measurable functions is measurable

Bevis 3.4

$$(f_1 \circ f_2)^{-1}(A) = \underbrace{f_2^{-1} \circ \underbrace{f_1^{-1}(A)}_{\substack{\text{measurable} \\ \in\Sigma}}}_{\substack{\text{measurable} \\ \in\Sigma}}$$

$\square$

Lemma 3.4

If $f_n : S \to \mathbb{R}$ is a sequence of measurable functions $\forall n \in \mathbb{N}$, then

- $\inf_n f_n$
- $\sup f_n$
- $\lim_n \inf f_n$
- $\lim_n \sup f_n$

are measurable. Moreover, the event that it exists is measurable, i.e

$$\left\{ s \in S : \lim_{n\to\infty} f_n(s) \text{ exists and is finite} \right\} \in \Sigma$$

Bevis 3.5

Note that $(\inf_n f_n)^{-1}([x,\infty)) = \left\{ s \in S : \underbrace{\inf_n f_n(s) \in [x,\infty)}_{\Leftrightarrow \inf_n f_n(s) \geq x} \right\}$

Then all events have to be $\geq x$, i.e intersection:

$$\bigcap_{n\in\mathbb{N}} \underbrace{\{s \in S : f_n(s) \geq x\}}_{=f_n^{-1}([x,\infty))\in\Sigma} \in \Sigma$$

This can be concluded naturally since f_n is measurable, and hence $\inf_n f_n$ is measurable. Similar reasoning shows that $\sup_n f_n$ is measurable.

Note that $\lim_{n\to\infty} \inf f_n(s) = \sup_{n\in\mathbb{N}} \inf_{m\geq n} f_n(s)$ which is just a composition of measurable functions which we have shown is measurable $\Rightarrow \lim_n \inf f_n$ is measurable. Similar reasoning shows that $\lim_n \sup f_n$ is measurable.

The last statement in Lemma 3.4 can be decomposed into the following:

$$\left\{s \in S : \lim_{n\to\infty} f_n(s) \text{ exists and is finite}\right\} \in \Sigma =$$

$$\left\{s \in S : \liminf_n f_n(s) > -\infty\right\} \cap \left\{s \in S : \limsup_n f_n(s) < \infty\right\} \cap \left\{s \in S : \liminf_n f_n(s) = \limsup_n f_n(s)\right\}$$

This is measurable since all of the 3 sets are measurable (pre-images of open sets). Think of it in the following way:

- $> -\infty \Rightarrow (-\infty, \infty]$ which is an open set
- $< \infty \Rightarrow [-\infty, \infty)$ which is an open set
- $= \Rightarrow \{0\}$ which is an open set

Since compositions of intersections are measurable, the proof is complete. $\square$

Definition 3.8 Random Variable

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.
A measurable funcftion $X : \mathcal{F} \to \mathbb{R}$ is called a *random variable.*

Example:
Let $\Omega = \{1, \cdots, 6\}$, $\mathcal{F} = \mathcal{P}(\Omega)$, $\mathbb{P} = \frac{1}{6}|A|$ (rolling a die)

Define $X(\omega) = \begin{cases} 1 & \omega \in \{1,3,5\} \quad \text{(odd)} \\ 0 & \omega \in \{2,4,6\} \quad \text{(even)} \end{cases}$

This is a random variable. One can verify this by checking pre-images of open sets of the range of the random variable $\{\varnothing, \{1,0\}, \{1\}, \{0\}\}$
By taking the discrete topology (collection of all subsets of S) $\mathcal{P}(S)$, this sneaky random variable is actually a random variable.
$Y(\omega) = \omega$ is also a random variable here since $\forall A \in \mathcal{F} = \mathcal{P}(\Omega) \subseteq \mathbb{R}$
Note that we have 2 distinct spaces, Ω could have not been a subset of $\mathbb{R}$, so Y would not have been a random variable since then $\mathcal{P}(\Omega) \not\subseteq \mathbb{R}$

Random variables "collapse" the space S due to their inherent injectivity. One way to measure this collapse is thinking of Borel σ-algebras in terms of this random variable. I.e, the smallest σ-algebra such that X is measurable

$$\overbrace{\sigma}^{\text{ensures }\sigma\text{-alg}}\left(\left\{ \underbrace{X^{-1}(A)}_{\text{ensure measurability}} : A \in \mathcal{B}(\mathbb{R}) \right\}\right) = \sigma(X)$$

In particular, $(\Omega, \sigma(X), \mathbb{P})$ is sufficient for X to be measurable (with respect to this space).

Example:

$$X(\omega) = \begin{cases} 1 & \omega \in \{1,3,5\} \quad \text{(odd)} \\ 0 & \omega \in \{2,4,6\} \quad \text{(even)} \end{cases} \qquad\qquad Y(\omega) = \omega$$

In this example,

$$\sigma(Y) = \sigma(\{Y^{-1}(A) : A \in \mathcal{B}(\mathbb{R})\}) = \sigma\overbrace{(\{\{1\}, \cdots, \{6\}\})}^{\mathcal{P}(\Omega)=\mathcal{F}}$$
$$\sigma(X) = \{\text{pre-images of neighborhoods of 1 \& 0}\} = \{\varnothing, \Omega, \{1,3,5\}, \{2,4,6\}\}$$

The last one may be difficult to grasp, but think of it like constructing the following set

$$\{\{\text{neither } 0 \vee 1 = \varnothing\}, \{\text{both } 0 \vee 1\}, \{\text{pre-image of } 1\}, \{\text{pre-image of } 0\}\}$$

This yields the smallest σ-algebra that contains these but still is $\neq \mathcal{F}$

Knowing nothing about a measurable/probability space is not possible, we always know things such as $\mu(\Omega) = 1$ and $\mu(\varnothing) = 0$. We could say that "if we know nothing, then we know those 2 things"
Conversely, if we know that $\mathcal{P}(\Omega)$ or $\mathcal{F}$ is a σ-algebra, then we know everything (we know the probability of every event happening).
For a constant random variable X, we know nothing (pre-images is $\{\varnothing, \Omega\}$). We can encode information using this.

4. Laws & Distribution functions

Definition 4.9 Law

Let X be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. A *law* $\mathcal{L}_X(A)$ captures probability of X in A, where $A \in \mathcal{B}(\mathbb{R})$

$$\mathcal{L}_X(A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\})$$

Remark:
This is the pull-back measure on $\mathbb{R}$. It is uniquely characterized by something known as the cummulative distribution function

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

4.1. Properties of distribution functions.

- Non-decreasing, i.e $F_X(t) \geq F_X(s)$ if $t \geq s$
- $\lim_{t\to-\infty} F_X(t) = 0$ $\lim_{t\to\infty} F_X(t) = 1$
- Right continuous, i.e $\lim_{t\searrow a} F_X(t) = F_X(a)$

Conversally, any F satisfying the above gives rise to a probability measure $\mathcal{L}$ such that $\mathcal{L}((\infty, \sqcup])$ is just the value at t

5. Independence

In this little mini-chapter, we fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$

Definition 5.10 Independence

Let $E_1, \cdots, \in \mathcal{F}$ be events (finitely many or countably infinite).
We say these are *independent* if for any combination of the following:

$$\mathbb{P}(E_{i_1} \cap \cdots \cap E_{i_k}) = \prod_{j=1}^{k} \mathbb{P}(E_{e_j}) \quad \forall i_1 < \cdots < i_k \wedge k$$

Example:
Consider throw of die as before. Let $E_1 =$ "number ≤ 2" $= \{1, 2\}$.
Here $\mathbb{P}(E_1) = \dfrac{2}{6} = \dfrac{1}{3}$. Let E_2 be "number we roll is even" $= \{2, 4, 6\}$
$\mathbb{P}(E_2) = \dfrac{3}{6} = \dfrac{1}{2}$. Then $\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(\{2\}) = \dfrac{1}{6} = \dfrac{1}{2} \cdot \dfrac{1}{3} = \mathbb{P}(E_1)\mathbb{P}(E_2) \Rightarrow$ independence.

Example:
Let $E_3 =$ "number " $\leq 3 = \{1, 2, 3\}$, $\mathbb{P}(E_3) = \dfrac{1}{2}$.
Then $\mathbb{P}(E_2 \cap E_3) = \mathbb{P}(\{2\}) = \dfrac{1}{6} \neq \dfrac{1}{2} \cdot \dfrac{1}{2} = \mathbb{P}(E_2)\mathbb{P}(E_3) \Rightarrow E_2 \wedge E_3$ are independent.

Definition 5.11 Independent random variable

Let $X_1, \cdots$ be finite (or countably infinite) random variables. We say that these are *independent* if all events that can happen with these random variables are independent, i.e for any choice $i_1 < \cdots < i_k$ and Borel sets $A_1, \cdots, A_k \in \mathcal{B}(\mathbb{R})$

$$\mathbb{P}(\{X_{i_1} \in A_1\} \cap \{X_{i_2} \in A_2\} \cap \cdots \cap \{X_{i_k} \in A_k\}) = \prod_{j=1}^{k} \mathbb{P}(\{X_{ij} \in A_j\})$$

A little remark may be that it looks kinda like a π-system. Can we define this using σ-algebras?

Definition 5.12

The sub σ-algebras $\mathcal{G}_1, \mathcal{G}_2, \cdots$ (finitely many or countably infinite) are said to be independent if

$$\mathbb{P}(G_{i_1} \cap \cdots \cap G_{i_k}) = \prod_{j=1}^{k} \mathbb{P}(G_{i_j}) \quad \forall G_{i_j} \in \mathcal{G}_{i_j}$$

Remark:
Note that independence of events and random vairables are special cases

- Let $E_1, \cdots$ be events. We let $\mathcal{G}$ be the trivial ones generated by an event, i.e $\mathcal{G}_{i_j} = \{\varnothing, \Omega, E_i, E_i^c\}$
- Let $X_1, \cdots$ be random variables and let $\mathcal{G}_i = \sigma(X_i)$

Lemma 5.1

Let $\mathcal{G}, \mathcal{H}$ be the sub σ-algebras and $\mathcal{I}, \mathcal{J}$ be 2 π-systems (i.e invariant under intersections) such that $\sigma(\mathcal{I}) = \mathcal{G}$ and $\sigma(\mathcal{J}) = \mathcal{H}$

Thus, $\mathcal{G}, \mathcal{H}$ are independent $\Leftrightarrow \mathbb{P}(I \cap J) = \mathbb{P}(I)\mathbb{P}(J) \quad \forall I \in \mathcal{I}, J \in \mathcal{J}$

Remark:

- $\{E_i\}$ are π-systems
- Events on the form $\{X_i \leq t\}$ are π-systems from $\sigma(X_i)$

To verify these claims, it suffices to check

$$\mathbb{P}(X_{i_1} \leq x_1 \quad \& \quad X_{i_2} \leq x_2 \cdots) = \prod_{j=1}^{k} \mathbb{P}(X_{i_j} \leq x_j)$$

Sats 5.3: Second Borell-Cantelli Theorem

Assume $E_1, \cdots$ are independent events and $\sum_{j=1}^{\infty} \mathbb{P}(E_j) = \infty$. Then

$$\mathbb{P}(\lim_{n\to\infty} \sup E_n) = \mathbb{P}(E_n \quad \text{happens } \infty \text{ often}) = 1$$

Bevis 5.1

Recall, $\lim_{n\to\infty} E_n = \bigcap_{n\in\mathbb{N}} \bigcup_{m\geq n} E_m$. Our strategy will be to prove that the complement is 0 (this is a good strategy whenever we want to prove that the probability of something is 1), since we can find the upper bound to be infinitely small.

By de-Morgan, complement is given by

$$\bigcup_{n\in\mathbb{N}} \bigcap_{m\geq n} E_m^c$$

Now:

$$\begin{aligned}\mathbb{P}(\bigcap_{m\geq n}^{M} E_m^c) \geq \mathbb{P}(\bigcap_{n\leq m\geq M} E_m^c) &= \prod_{m=n}^{M} \mathbb{P}(E_m^c)\\ &= \prod_{m=n}^{M} 1-\mathbb{P}(E_m)\end{aligned}$$

(Complements preserves independence). We approximate using exponents since

$$1-x \leq e^{-x} \Leftrightarrow \ln(t) < t-1 \leq \prod_{m=n}^{M} e^{-\mathbb{P}(E_m)} = \exp\left\{-\sum_{m=n}^{M} \mathbb{P}(E_m)\right\}$$

Since $\mathbb{P}(E_m) \to \infty$ as $M \geq n \to 0$

Taking $M \gg n$, we get

$$\mathbb{P}(\bigcap_{m>n} E_m^c) < \varepsilon \quad \forall\varepsilon > 0 \Rightarrow \mathbb{P}(\bigcap_{m>n} E_m^c) = 0$$

$$\Rightarrow \left.\underbrace{\bigcup_{n\in\mathbb{N}} \underbrace{\bigcap_{m\geq n} E_m^c}_{\mathbb{P}=0}}_{\mathbb{P}=0}\right\} \text{countable unions of something with } \mathbb{P}=0 \text{ yields } 0$$

$$\Rightarrow \text{Complement is } 1$$

$\square$

As long as the sum tends to 0 such that the sum diverges, it will happen infinitely often.

Example:
Let E_k = "draw card 1 on kth draw". Assuming independence, we have

$$\mathbb{P}(E_k) = \frac{1}{k} \quad \wedge \quad \sum_{k=1}^{\infty} \mathbb{P}(E_k) = \sum_{k=1}^{\infty} \frac{1}{k}$$

By the second Borell-Cantelli theorem, $\mathbb{P}(\text{"draw 1 } \infty \text{ often"}) = 1$

Remark:
For $E_1, \cdots$ independent events, the Borell-Cantelli theoerems give the following dichotomy:

$$\mathbb{P}(\lim_{n\to\infty} \sup E_n) = \mathbb{P}(E_n \text{ happens } \infty \text{ often}) = \begin{cases} 1 & \text{if } \sum_i \mathbb{P}(E_i) = \infty \\ 0 & \text{if} \sum_i \mathbb{P}(E_i) < \infty \end{cases}$$

This is a special case of Kolmogorovs 0-1 Law:

Definition 5.13 Tail σ-algebra

Let $X_1, \cdots$ be a sequence of random variables. Set $\mathcal{T}_n = \sigma(X_{n+1}, X_{n+2}, \cdots)$ and $\mathcal{T} = \bigcap_{n\in\mathbb{N}} \mathcal{T}_n$ What remains is knowledge at ∞. We say $\mathcal{T}$ is a tail *σ-algebra*

Example:
Some typical events $E \in \mathcal{T}$ are $\{\lim_n X_n \text{ exists}\}$ or something like $\{\sum X_n \text{ converges}\}$.
The point here is that all information is talking about what happens at ∞

Sats 5.4: Kolmogorovs 0-1 Law

Let $X_1, \cdots$ be independent random variables. Then $\forall T \in \mathcal{T}$ (tail σ-algebras generated by X_i), we have either $\mathbb{P}(T) = 0$ or $\mathbb{P}(T) = 1$ almost surely.

Bevis 5.2: (Sketch) Kolmogorovs 0-1 Law

(1) Define $\mathcal{X}_n = \sigma(X_1, \cdots, X_n)$. From this, we can say that the $\mathcal{X}_n$ and $\mathcal{T}_n$ are indeoendent (one stops at n, the other one continues at $n+1$) $\forall n \in \mathbb{N}$
(2) $\mathcal{T} \subseteq \mathcal{T}_\setminus \quad \forall n$, so $\mathcal{T}$ is independent of $\mathcal{X}_n \quad \forall n$
(3) $\mathcal{X}_\infty = \sigma(X_1, \cdots)$ and $\mathcal{T}$ must be independent since $\mathcal{T}$ is independent for all $\mathcal{X}_n$ and $\bigcup \mathcal{X}_n$ are a π-system (which generates $\mathcal{X}_\infty$)
(4) $\mathcal{T} \subseteq \mathcal{X}_\infty$ (knowledge in tail is contained in ∞ for $\mathcal{X}$)

So for any event $F \in \mathcal{T}$ we know

$$\mathbb{P}(F \cap F) = \mathbb{P}(F) = \mathbb{P}(F)\mathbb{P}(F)$$

$\Rightarrow \mathbb{P}(F)$ must solve $x = x^2$, i.e $x \in \{0, 1\}$ $\square$

Corollary:

Let ξ be a $\mathcal{T}$ measurable random variable (random thing that only depends on tail, eg. $x = \begin{cases} 1 & \text{if } \lim_{n\to\infty} \text{ exists} \\ 0 & \text{else} \end{cases}$), then $\exists C \in [-\infty, \infty]$ such that $\mathbb{P}(\xi = C) = 1$, i.e ξ is almost surely constant.

6. Integration

Let (S, Σ, μ) be a measure-space. Given some measurable function $f \in m\Sigma$ where $f : S \to \mathbb{R}$, our goal is to define the integral of this function with respect to our measure μ
We do it in 3 steps, and then the idea is to reduce integration problems to these 3 cases and proceed. Later we will develop some tools to ease in this conversion.

(1) Define integration of an indicator function

$$I_A(s) = \begin{cases} 1 & \text{if } s \in A \\ 0 & \text{else} \end{cases} \qquad \int I_A(s) d\mu = \mu(A) \quad \forall A \in \Sigma$$

(2) Define the integral for finite linear combinations of characteristic functions $f(s) = \sum_{k=1}^{n} a_k I_{A_k}(s)$ where $a_k \in \mathbb{R}$ (sometimes we say they are non-negative, but this is not always the case) and $A_k \in \Sigma$
These functions are called *step-functions*. We set the integral:

$$\int f(s) d\mu = \sum_{k=1}^{n} a_k \int I_{A_k} d\mu = \sum_{k=1}^{n} a_k \mu(A_k)$$

Note that if A_k overlap, they are double-counted. This is the desired functionality

(3) For $f \in m\Sigma^+$, we define:

$$\int f d\mu = \sup \left\{ \int g d\mu : g \leq f \;\; f, g \text{ are non-negative step-functions} \right\}$$

(4) We extend this to all measurable functions f by defining the positive & negative parts of a functions by

$$f^+(s) = \begin{cases} f(s) & \text{if } f(s) > 0 \\ 0 & \text{else} \end{cases}$$
$$f^-(s) = \begin{cases} -f(s) & \text{if } f(s) < 0 \\ 0 & \text{else} \end{cases}$$

Note here that both are non-negative and $f = f^+ - f^-$. All operations utilized preserve measurability, thus we define

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

6.1. Properties of the integral.

(1) **Linearity:** $\int af + bg d\mu = a \int f d\mu + b \int g d\mu \quad \forall a, b \in \mathbb{R}$ and $f, g \in m\Sigma$

(2) **Monotonicity:** If $f \leq g \quad \forall s$, then $\int f d\mu \leq \int g d\mu$ (note here that the $\forall$ is really just a "for almost all" sign, since $\mu(\varnothing) = 0$)

(3) **Triangle-inequality:**

$$\left| \int f d\mu \right| = \left| \int f^+ d\mu - \int f^- d\mu \right| \leq \left| \int f^+ d\mu \right| + \left| \int f^- d\mu \right| = \int |f| \, d\mu$$

Remark:
If μ is a Lebesgue measure on $\mathbb{R}$ (in general in $\mathbb{R}^n$), then $\int d\mu$ is standard Lebesgue integration. If both the Riemann and Lebesgue integral exist (and are equal) then it is just standard Riemann integration

Example:
Let μ be the counting measure on integers. Consider $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$
Then $\int f d\mu$ for any $f \in m\mathcal{B}(\mathbb{R}) = \sum_{k=1}^{\infty} f(k)\mu(\{k\}) = \sum_{k=1}^{\infty} f(k)$
Note, this sees very little of the space and does not work with Riemann integration.

We can restrict the domain just like we would with "regular integration". Let $A \in \Sigma$, then

$$\int_A f d\mu = \int f I_A d\mu$$

Definition 6.14 Integrable

We say that $f \in m\Sigma$ is *integrable* with respect to the measure μ if $\int f^+ d\mu$ and $\int f^- d\mu$ exist and are finite. Otherwise, we say that the integral is *undefined*

The class of integrable functions is denoted $L^1(S, \Sigma, \mu)$

Note: If $f(s) = \pm\infty$ and $f \in L^1$, we must have $\mu(\{s\}) = 0$: Moreoever,

$$\mu(\{s : f(s) = \pm\infty\}) = 0$$

Lemma 6.1

If f is a non-negative integrable function and $\int f d\mu = 0$, we claim that the measure of all points s such that f is positive is zero for almost every s (i.e $\mu(\{s : f(s) = \pm\infty\}) = 0$)

Bevis 6.1

The proof strategy here will be to dissect this using unions/intersections and their boundedness. Note that

$$\bigcup_{n\in\mathbb{N}} \left\{ s : f(s) > \frac{1}{n} \right\} = \{s : f(s) > 0\}$$

If $\mu(A_n) = 0 \quad \forall n$, then the function is almost surely 0
If this is not the case, then $\exists k \in \mathbb{N}$ such that $\mu(A_k) > 0$, but then we can take a function that is bounded above by f like $\frac{1}{k} I_k \leq f$.
By monotonicity,

$$\int f(s) d\mu \geq \int \frac{1}{k} I_{A_k} d\mu = \frac{1}{k}\mu(A_k) > 0$$

By assumption, $\int f d\mu = 0$ which is a contradiction, so $\mu(I_{A_k}) = 0$ $\square$

One question that we shall try to explore now is what happens with limits of sequences of functions and their integrals?
Well, first we may see them as a sequence of expectation of random variables. By intution one might jump to the conclusion that

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

But this is not the case, for example if $f_n = I_{[n,n+1)}$ and μ is the Lebesgue measure we have

$$\int f_n(s) d\mu = \int_{[n,n+1)} (x) dx = 1$$

In particular, $\lim_{n\to\infty} \int f_n(s) d\mu = 1$
However, for a fixed x we have $\lim_{n\to\infty} f_n(x) = 0$ and

$$\int \lim_{n\to\infty} f_n d\mu = 0 d\mu = 0$$

There are circumstances where this equality holds however, and the goal is somewhat to discover *when* this is.

Sats 6.5: Monotone Convergance Theorem

Let f_n be a sequence of non-negative measurable funcitons $f_n \in m\Sigma^+$ such that $f_n \to f$ pointwise (i.e $f_n(x)$ is non-decreasing in n and $f_n(x) \to f(x)$ as $n \to \infty$). Then

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

We can always approximate measurable functions using sequences of step functions using

$$\alpha^{(r)}(x) = \begin{cases} 0 \quad \text{if } x = 0 \\ (i-1)2^{-r} \quad \text{if } (i-r)2^{-r} \leq x \leq i2^{-r} \leq r \\ r \quad \text{for } x > r \end{cases}$$

Note here that we do not mean $i \in \mathbb{C}$
This is basically discretisation of $y = x$ up to r. It is non-decreasing (monotonly) in r, and always less than y = x.
By setting $f^{(r)}(x) = \alpha^{(r)}(f(x)) \quad f \in m\Sigma^+$, we get a function that has all the properties that we wished for (i.e $f_n \to f$ and f_n are step functions).
We can use this to base our proofs on since we can then start to construct proofs in the following way:

(1) Prove the property for indictor functions
(2) Extend to linear combination of step functions (by proven linearity)
(3) Extend to $f \in m\Sigma^+$ by monotonicity (exchange limits)
(4) Extend to $f \in m\Sigma$ by splitting $f = f^+ - f^-$

Lemma 6.2

Suppose $f, g \in L^1$ and $\in m\Sigma$ and $f = g$ for almost all $s \in S$. Then

$$\int f d\mu = \int g d\mu$$

Bevis 6.2: (Sketch)

Consider $f - g$. We want to show that $\int f - g d\mu = 0$
If f, g are indictor functions, this is trivially true (since they must have the same step-functions)
$\Rightarrow$ Take finite linear combinations and consider monotonicity through step functions (step of 0 is 0) and split into f^+ and f^- $\square$

Lemma 6.3: Fatous Lemma v2

Suppose f_n is a sequence of non-negative measurable functions ($\in m\Sigma^+$). Then

$$\mu(\lim_{n\to\infty} \inf f_n) \leq \lim_{n\to\infty} \inf \mu(f_n)$$

Bevis 6.3

Consider $g_k = \inf_{n \geq k} f_n$ (increasing). Then, $\lim_{k\to\infty} g_k$ exists and is equal to $\lim_{k\to\infty} \inf f_n$
Since g_k is monotonly increasing, we have that $g_k \to \lim_{n\to\infty} \inf f_n$, we can use the Monotone Conversion Theorem (MCT) to take the limit out

$$\mu(\lim_{k\to\infty} g_k) = \mu(\lim_{n\to\infty} \inf f_n) = \lim_{k\to\infty} \mu(g_k)$$

We have $g_k \leq f_n \quad \forall n \geq k$, we can use the monotonicity

$$\mu(g_k) \leq \mu(f_n) \quad \forall n \geq k \Rightarrow \mu(g_k) \leq \inf_{n \geq k} \mu(f_n)$$

Substitution yields

$$\mu(\lim \inf f_n) \leq \lim_{k\to\infty} \inf_{n \geq k} \mu(f_n) = \lim_{n\to\infty} \inf \mu(f_n)$$

$\square$

Corollary:
We can say something about lim sup, suppose $f_n \leq g$ for $f_n, g \in m\Sigma^+$, then

$$\mu(\lim_{n\to\infty} \sup f_n) \geq \lim_{n\to\infty} \sup \mu(f_n)$$

Bevis 6.4

Apply Fatous lemma with some $h_n = g - f_n$. The sign flips things arround and yields what we want $\square$

Sats 6.6: Dominated Convergance Theorem

Let f_n be a sequence of measurable functions & assume $|f_n| \leq g$ for some $g \in L^1$. If $f_n \to f$ pointwise, then:

- $\mu(|f_n - f|) = \int |f_n - f| \, d\mu \to 0$
- $\mu(f_n) = \int f_n d\mu \to \int f d\mu = \mu(f)$

Bevis 6.5: Dominated Convergance Theorem

We have

$$|f_n - f| \leq |f_n| + |f| \leq 2g$$

Since $|f_n| \leq g$ and $|f_n| \to |f_n|$ pointwise.
By reverse Fatous lemma

$$\lim_{n\to\infty} \sup \mu\left(|f_n - f|\right) \leq \mu\left(\lim_{n\to\infty \sup} |f_n - f|\right) \Rightarrow \lim_{n\to\infty} \sup \mu\left(|f_n - f|\right) = 0$$

It goes to 0 $\forall n$, so $\lim_{n\to\infty} \inf \mu\left(|f_n - f|\right) \leq 0$, so $\lim_{n\to\infty} \mu\left(|f_n - f|\right) = 0$
So $|\mu(f_n) - \mu(f)|$, by linearity of the integral:

$$|\mu(f_n - f)| \leq \mu\left(|f_n - f|\right) \to 0 \text{ as } n \to \infty$$
$$\Rightarrow \mu(f_n) \to \mu(f)$$

$\square$

Lemma 6.4: Scheffes lemma

Suppose f_n, f are non-negative and $f_n \to f$ for almost any $s \in S$, then

$$\mu(f_n) \to \mu(f) \text{ as } n \to \infty \Leftrightarrow \mu(|f_n - f|) \to 0 \text{ as } n \to \infty$$

Recall how we defined $\int f d\mu$. We started with indicator functions, then linear combinations of them, then we defined for non-negative functions by taking supremum. Then for negative functions, we looked at their positive parts and their negative part.
THe idea now is we want to measure one function with respect to another function.

6.2. **Modifying Measures.**
Let $f \in \Sigma^+$. We consider the restricted integral

$$\int_A f d\mu = \int f I_A d\mu = \lambda(A) = \lambda_{f,\mu}(A)$$

where $A \in \Sigma$. This is a measure (also denoted $\mu(f;a)$):

- $\lambda(A) \geq 0 \quad \forall A \in \Sigma$
- σ-additivity follows from linearity of integral, disjoint $A_n \in \Sigma$
-

$$\lambda\left(\bigcup_{n=1}^{\infty} A_n\right) = \int f A_{\bigcup_{n=1}^{\infty} A_n} d\mu = \int f\left(\sum_{n=1}^{\infty} I_{A_n}\right) d\mu = \int f \lim_{N\to\infty} \underbrace{\sum_{n=1}^{N} I_{A_n}}_{\leq 1} d\mu = \int \lim_{N\to\infty} \sum_{n=1}^{N} f I_{A_n} d\mu$$

By MCT, we can take the limit out

$$\lim_{N\to\infty} \int \sum_{n=1}^{N} f I_{A_n} d\mu = \lim_{N\to\infty} \sum_{n=1}^{N} \underbrace{\int f I_{A_n} d\mu}_{\lambda(A_n)} \Rightarrow \lim \sum \lambda(A_n) = \sum_{n=1}^{\infty} \lambda(A_n)$$

- $\lambda(\varnothing) = 0$

$\Rightarrow \lambda$ is a measure, with density f with respect to μ
We write this as $f = \dfrac{d\lambda}{d\mu}$

Definition 6.15 σ-finite measure

We say a measure μ is σ-finite if we can split it into finite measures

$$\exists A_n \in S \text{ s.t } S = \bigcup_{n=1}^{\infty} A_n, \mu(A_n) < \infty \quad \forall n \in \mathbb{N}$$

Example is a Lebesgue measure

Sats 6.7: Radon-Nikodyn theore

If μ, λ are σ-finite measures and one dominates the other such that $\mu(A) = 0 \Rightarrow \lambda(A) = 0$, then $\exists$ a density function $f = \dfrac{d\lambda}{d\mu}$ such that $\lambda(A) = \int_A f d\mu \quad \forall A \in \Sigma$

Example:
Let μ be the Lebesgue measure and $f = \dfrac{1}{\sqrt{2\pi}} e^{-x^2/2}$, then $\lambda(A) = \int_A f d\mu$ is the measure associated with normal distributed random variables on $\mathbb{R}$

7. Expectations

The are integrals with respect to a probability measure. An expected value is ”a value we expect the random variable to take”

Let $(\Omega, \mathcal{F}, \mathbb{R})$ and X be a random variable. Then $\mathbb{E}(X) = \int X(\omega) d\mathbb{P}(omega) = \int X d\mathbb{P}$.
If it exists, then X is integrable ($\int |X| \, d\mathbb{P} < \infty$)

Example: (*Die roll*)

$$X(\omega) = X(1)I_1(\omega) + \cdots + X(6)I_6,\ \mathbb{E}(X) = \int X d\mathbb{P} = \frac{1}{6}$$

$$\mathbb{E}(X) = \int X d\mathbb{P} = \frac{1}{6}X(1) + \cdots + \frac{1}{6}X(6)$$

A nice thing to remember is all the integral theorems that were previously stated have now become expectation theorems:

- $0 \leq X_n \to X \Rightarrow \mathbb{E}(X_n) \to \mathbb{E}(X)$ (MCT)
- $|X_n| \leq Y \quad \mathbb{E}(Y) < \infty \Rightarrow \mathbb{E}(X_n) \to \mathbb{E}(\lim_{n\to\infty} X_n)$ (DCP)
- $X_n \to X \Rightarrow \mathbb{E}(X) = \mathbb{E}(\lim_{n\to\infty} X_n) \leq \lim_{n\to\infty} \inf \mathbb{E}(X_n)$ (Fatou)
- $\mathbb{E}(X; A) = \mu(X; A) = \mathbb{E}(X I_A) = \int_A X d\mathbb{P}$

We are now going to look at estimating size/somethign large using expectation.

Sats 7.8: Markovs Inequality

Let Z be a random variable with values in some set G ($Z : \Omega \to G$) and let $g : G \to [0, \infty]$ be a non-decreasing function in $G \subseteq \mathbb{R}$ and measurable. Then

$$\mathbb{E}(g(Z)) \geq \mathbb{E}(g(Z) I_{Z\geq C}) = \mathbb{E}(g(Z); Z \geq C)$$

Bevis 7.1: Markovs Inequality

Since g is non-decreasing, we have $\geq \mathbb{E}(g(C); Z \geq C) = g(C)(I_{Z\geq C} \cdot \text{ const.}) \leq g(C)\mathbb{P}(Z \geq C)$

We can now estimate $\mathbb{P}(Z \geq C)$ using expectation, since $\mathbb{E}(g(Z)) \geq g(C)\mathbb{P}(Z \geq C)$ we get $\mathbb{P}(Z \geq C) \leq \dfrac{\mathbb{E}(g(Z))}{g(C)}$ $\square$

Obviously the special case occurs when $g(X) = X$, since the inequality becomes $\mathbb{P}(Z \geq C) \leq \dfrac{\mathbb{E}(Z)}{C}$ for non-negative Z

Example:
Let $Z : \Omega \to \mathbb{N}$, then

$$\mathbb{P}(Z \neq 0) = P(Z \geq 1) \leq \frac{\mathbb{E}(Z)}{1} = \mathbb{E}(Z)$$

Important special case:

$$g(X) = e^{\theta X}, \quad \theta > 0$$

Then, $\mathbb{P}(Z \geq C) = \mathbb{P}(e^{\theta Z} \geq e^{\theta C}) \leq \dfrac{\overbrace{\mathbb{E}(e^{\theta Z})}^{\text{mgf.}}}{e^{\theta C}}$

7.1. **Jensens inequality.**

A function is called convex on I if function value of an average $f(px + qy) \leq pf(x) + qf(y) \quad \forall p, q \in [0, 1]$ such that $p + q = 1$ and $x, y \in I$
In laymans terms, a straight line from $x \to y$ is above the function, then $f(px+qy) \leq$ straightline$(px+qy)$

Examples:
$x \mapsto C$, $x \mapsto cx$, $x \mapsto e^x$, $x \mapsto x^n$ for $n \geq 1$

Definition 7.16 Jensens inequality

Let $f : I \to \mathbb{R}$ be a convex function and $x : \Omega \to I$ be a random variable.
Then $\mathbb{E}(f(X)) \geq f(\mathbb{E}(X))$

Bevis 7.2: Jensens inequality

We start off by rewriting the convexity condition: $\underbrace{f(v) - f(u)}_{v-u} \leq \underbrace{f(w) - f(u)}_{w-v}$ for $u < v < w$.

So, by monotonicity, left and right derivatives exist (but not always equal). We get $f(x) > f(x) \geq f(w) + m(x - v)$ for any m between left and right derivative. Substituting this into the above, we get

$$\mathbb{E}(f(x)) \geq \underbrace{\mathbb{E}(f(\underbrace{\mathbb{E}(x)}_{\text{const.}})))}_{f(\mathbb{E}(x))} + \underbrace{m(x - \mathbb{E}(x))}_{\mathbb{E}(m(x-\mathbb{E}(x)))=m(\mathbb{E}(x))-m(\mathbb{E}(x))} = f(\mathbb{E}(x))$$

$\square$

7.2. L^p**-norms.**
These norms tell us how well behaved our function is.
For $p \geq 1$, we define $||X||_p = \mathbb{E}(|X|^p)^{1/p}$, this defines a norm for $p \geq 1$.

$L^p(\Omega, \mathcal{F}, \mathbb{P})$ is a collection of all functions such that the p-norm is finite ($\forall X$ such that $||X||_p < \infty$). Let $f(X) = X^{r/p}$, we want this to be convex so for $r \geq p \geq 1$, then f is convex and we can use Jensens inequality and up with

$$\mathbb{E}(|Y|^{r/p}) \geq \mathbb{E}(|Y|)^{r(p)} \Rightarrow Y = |X|^p$$
$$\Rightarrow \mathbb{E}(|X|^r) \geq \mathbb{E}(|X|^p)^{r/p} \Rightarrow \mathbb{E}(|X|^r)^{1/r} \geq \mathbb{E}(|X|^p)^{1/p}$$
$$\Rightarrow ||X||_r \geq ||X||_p \qquad \text{where } r > p$$

If r-norm $< \infty \Rightarrow p$-norm $< \infty$, so $L^r(\Omega, \mathcal{F}, \mathbb{P}) \subseteq L^p(\Omega, \mathcal{F}, \mathbb{P})$

Be advised! L^2 is important!

Definition 7.17 Cauchy-Schwarz inequality

If X, Y are random variables and $\in L^2(\Omega, \mathcal{F}, \mathbb{P})$, then $X \cdot Y$ is integrable (also in L^2) and we can bound the product:

$$|\mathbb{E}(XY)| \leq \mathbb{E}(|XY|) \leq ||X||_2 \, ||Y||_2 = \sqrt{\mathbb{E}(|X|)^2 \mathbb{E}(|Y|)^2}$$

This implies we got an inner product on 2 random variables by $\langle X, Y \rangle = \mathbb{E}(|XY|)$

Bevis 7.3: Cauchy-Schwarz inequality

Truncating X, Y by defining $X_n = \min\{|X|, n\}$ and similarly for Y_n. These are bounded and non-negative random variables
For $a \in \mathbb{R}$, we look at
$E((aX_n + Y_n)^2) \geq 0$ (since we are squaring it), but we can also write this on the form $a^2\mathbb{E}(X_n^2) + \mathbb{E}(Y_n^2) + 2a\mathbb{E}(X_nY_n) \geq 0$
Lets consider this as a polynomial in a:

$$f(a), \quad f \geq 0 \Rightarrow (2\mathbb{E}(X_nY_n))^2 - 4\mathbb{E}(X_n^2)\mathbb{E}(Y_n^2) < 0$$
$$\Rightarrow \mathbb{E}(X_nY_n)^2 \leq \mathbb{E}(X_n^2)\mathbb{E}(Y_n^2) \Rightarrow \mathbb{E}(X_nY_n) \leq \sqrt{\mathbb{E}(X_n^2)\mathbb{E}(Y_n^2)}$$

This is the Cauchy-Schwarz inequality for X_n, Y_n. To finish the proof, we take limits in n and use monotonicity and MCT to get:

$$\mathbb{E}(|XY|) \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}$$

$\square$

The proof idea we used was to truncate (bound) and then take the limit.

Corollary:

$$||X+Y||_2 \leq ||X||_2 + ||Y||_2$$

Bevis 7.4

The trick here is to take the p:th power, in this case 2:

$$||X+Y||_2^2 = \mathbb{E}((X+Y)^2) = \mathbb{E}(X^2) + \mathbb{E}(Y) + 2\mathbb{E}(XY) \leq \mathbb{E}(X^2) + \mathbb{E}(Y^2) + 2\,||X||_2\,||Y||_2$$

□

Since what we have shown satisfies the triangle inequality, means we got a norm.

Definition 7.18 Covariance and variance

Let X, Y be random variables with $m_X = \mathbb{E}(X)$ and $m_Y = \mathbb{E}(Y)$
We set $\mathrm{Cov}\,(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ and $\mathrm{Var}\,(X) = \mathrm{Cov}\,(X,X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$

Note that $\mathrm{Cov}\,(X,Y) = \mathbb{E}((X - m_X)(Y - m_Y))$

Properties:

- $\mathrm{Var}\,(x) \geq 0$
- X, Y independent $\Rightarrow \mathrm{Cov}\,(X,Y) = 0$
- $|\mathrm{Cov}\,(X,Y)| \leq \sqrt{\mathrm{Var}\,(X)\,\mathrm{Var}\,(Y)}$

Normalisiing yields the correlation:

$$\frac{\mathrm{Cov}\,(X,Y)}{\sqrt{\mathrm{Var}\,(x)}\mathrm{Var}\,(Y)} = \mathrm{Corr}\,(X,Y)$$

Note that the denominator is a bound for the numerator, hence $|\mathrm{Corr}\,(X,Y)| \leq 1$ and equality if and only if there is an almost sure linear relation between X, Y.

This generalises to something known as the Hölder inequality:
Assume $X \in L^p, \quad Y \in L^q$ with $\dfrac{1}{p} + \dfrac{1}{q} = 1$ (so $p, q \geq 1$), then

$$|\mathbb{E}(XY)| \leq \mathbb{E}(|XY|) \leq ||X||_p + ||Y||_q$$

holds for measure spaces.

Corollary: *(Minkowskis inequality)*

$$||X+Y||_p \leq ||X||_p + ||Y||_p$$

Bevis 7.5

Want to use Hölders inequality and previous trick of truncating. $||X+Y||_p^p = \mathbb{E}((X+Y)^p)$, wlog, $p \geq 1$ (case $p=1$ is trivial)

$$\begin{aligned}\mathbb{E}(|X+Y|\,|X+Y|^{p-1}) &\leq \mathbb{E}(|X|\,|X+Y|^{p-1}) + \mathbb{E}(|Y|\,|X+Y|^{p-1}) \\ &\leq ||X||_p \left|\left||X+Y|^{p-1}\right|\right|_q + ||Y||_p \left|\left||X+Y|^{p-q}\right|\right|_q\end{aligned}$$

By Hölder inequality and the requirement on p, q yielding $q = \dfrac{p}{p-1}$, we collect some terms and get:

$$\begin{gathered}(||X||_p + ||Y||_p) \left|\left||X+Y|^{p-1}\right|\right|_q = (||X||_p + ||Y||_p) E\left(|X+Y|^p\right)^{1/q} \\ \Rightarrow (||X+Y||_p)^{p-p/q} \leq ||X||_p + ||Y||_p\end{gathered}$$

But $p - \frac{p}{q} = 1$ and thus the claim follows. $\square$

Thus $||\cdot||_p$ is a norm.

Definition 7.19 Completeness of L^p

$L^p(\Omega, \mathcal{F}, \mathbb{P})$ is complete (not bounded), i.e Cauchy sequences with respect to the p-norm converge in the space.

8. Densities

Suppose X is a random variable with law $\Lambda_X(A) = \mathbb{P}(X \in A)$

Sats 8.9

For every Borell measurable function f, we get that $\mathbb{E}(f(X)) = \int_{\mathbb{R}} f(X) d\Lambda_X$
Recall $\mathbb{E}(f(X)) = \int_{\Omega} f(X) d\mathbb{P}$

Here $\mathbb{R}$ is the tangent space of X
Recall the proof strategy for integrals, we start by considering indicator functions.

Bevis 8.1

$f(x) = I_A$ where $A \in \mathcal{B}(\mathbb{R})$, then

$$\int_{\Omega} I_A(X) d\mathbb{P} = \mathbb{E}(I_{X \in f}) = \mathbb{P}(X \in A)$$

The left hand side is complete, lets check the right hand side:

$$\int_{\mathbb{R}} I_A(X) d\Lambda_X = \int_A 1 d\Lambda_X = \Lambda_X(A)\mathbb{P}(X \in A)$$

We see that it is true for indicator functions, this can be extended to finitely many linear combinations of step functions and then using MCT with $f \in m\Sigma^+$, then $f = f^+ - f^- \Rightarrow f \in m\Sigma$ $\square$

Remark:
If X has a density, we get

$$\mathbb{E}(f(X)) = \int_{\mathbb{R}} f(X)\varphi(X) dX$$

Where φ is the density of X. The proof of this is similar to the previous proof.

Remark:
Density here is the same as in the Radon-Nikodyn theorem.

Recall:
X, Y are independent if $\mathbb{P}(\{X \in A\} \cap \{Y \in B\}) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B) \quad \forall A, B \in \Sigma$
Through independence, we can split expectation

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$$

Bevis 8.2

The idea here is to use the step function trick.
We estimate X, Y by increasing step functions $\alpha^{(r)}(X)$ and similarly for Y
Each function is a linear combination of indicators which which $I_A(X) = \begin{cases} 1 & X \in A \\ 0 & \text{else} \end{cases}$ (and similarly for Y). We get:

$$\mathbb{E}(I_A(X)I_B(Y)) = \mathbb{P}(\{X \in A\} \cap \{Y \in B\}) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B) = \mathbb{E}(I_A(X))\mathbb{E}(I_B(Y))$$

It has now been proved for indicators, we extend this by linearity and MCT $\square$

Corollary:
Two independent random variables X, Y have 0 covariance and their variance behaves linear.

Bevis 8.3

Just plug and chugg in definitions:

$$\mathrm{Cov}\,(X,Y) = \mathbb{E}(XY)_{\mathbb{E}}(X)\mathbb{E}(Y) \Rightarrow \mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$$

For the variance:

$$\mathbb{E}((X+Y)^2) = \mathbb{E}(X^2) + \mathbb{E}(Y^2) + 2\mathbb{E}(XY)$$

$$(\mathbb{E}(X+Y))^2 = (\mathbb{E}(X) + \mathbb{E}(Y))^2 = (\mathbb{E}(X))^2 + (\mathbb{E}(Y))^2 + 2\mathbb{E}(X)\mathbb{E}(Y)$$

$$\Rightarrow \mathbb{E}((X+Y)^2) - (\mathbb{E}(X+Y))^2$$

$$= \mathbb{E}(X^2) - (\mathbb{E}(X))^2 + \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2 + 2\mathbb{E}(XY) - 2\mathbb{E}(XY) = \mathrm{Var}\,(X) + \mathrm{Var}\,(Y)$$

□

Remark:
If X, Y are independent, then $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, but the converse is not true!

Sats 8.10: (Weak) Strong law of large numbers

Let $X_1, \cdots$ be a sequence of independent random variables where $\mathbb{E}(X_i) = 0$ and $\mathbb{E}(|X_i|^4) < \infty$ (or bounded by some finite constant). Then:

$$\frac{X_1 + \cdots + X_n}{n} \to 0 \quad \text{as } n \to \infty$$

Bevis 8.4: (Weak) Strong law of large numbers

There is some clue in that we have the 4th power requirement. We shall use the inequalities from above. Let $S_n = \sum_{i=1}^n X_i$, and $\mathbb{E}(S_n^4)\mathbb{E}\left((\sum_{i=1}^n X_i)^4\right)$
Using the binomial theorem, we get:

$$\begin{gathered}\mathbb{E}(X_1^4) + \mathbb{E}(X_2^4) + \cdots \\ +4\mathbb{E}(X_1X_2^3) + 4\mathbb{E}(X_1X_3^3) + \cdots \\ 6\mathbb{E}(X_1^2X_2^2) + 6\mathbb{E}\cdots \\ +12\mathbb{E}(X_1^2X_2X_3) + \cdots \\ +24\mathbb{E}(X_1X_2X_3X_4)\end{gathered}$$

Remember that we have independence, so for terms on the form $\mathbb{E}(X_1X_2^3)$, we can rewrite this as $\mathbb{E}(X_1)\mathbb{E}(X_2^3)$. However, we do not know if the square is independent. This yields

$$\mathbb{E}(X_1^4) + \cdots + \cdots + \mathbb{E}(X_n^4) + R$$

Where R here is the remainder, which is on the form $\mathbb{E}(X_1)\mathbb{E}(\cdots)$, so all that survives are:

$$\left.\mathbb{E}(S_n^4) = \underbrace{\sum \mathbb{E}(X_i^4)}_{\text{bounded}} + \underbrace{\sum_{i\neq j} \mathbb{E}(X_i^2X_j^2)}_{\text{looks like C.S}}\right\} \Rightarrow \mathbb{E}(X_i^2X_j^2) \leq \sqrt{\underbrace{\mathbb{E}(X_i^4)}_{\text{bounded}}\,\underbrace{\mathbb{E}(X_j^4)}_{\text{bounded}}} \leq \sqrt{k^2} = k$$

So $\mathbb{E}(S_n^4) \leq nk + n\underbrace{(n-1)}_{i\neq j}k \leq 2nk$, and so

$$\mathbb{E}\left(\left(\frac{S_n}{n}\right)^4\right) \leq \frac{1}{n^4}2n^2k = \frac{2k}{n^2}$$

Borell-Cantelli type beat:

$$\mathbb{E}\left(\sum_{n=1}^{\infty}\left(\frac{S_n}{n}\right)^4\right) \leq \sum_{k=1}^{\infty}\frac{2k}{n^2} < \infty$$

If the expectation is finite, then with probability 1 $\sum\left(\frac{S_n}{n}\right)^4$ is finite $\Rightarrow \frac{S_n}{n} \to 0$ almost surely.

An alternative method is to use Markovs inequality $\mathbb{P}\left(\frac{S_n}{n} \leq \frac{1}{n}\right)$ $\square$

Note:
For i.i.d with $\mathbb{E}(X_i) = m$, $\frac{\sum^n X_i}{n}$ converges to m almost surely, provided the 4th moment is bounded. This can be proved by considering $Y = X_i - \mathbb{E}(X_i)$

Definition 8.20 Chebychevs inequality

$$\mathbb{P}(|X - \mu| \geq C) \leq \frac{\mathbb{E}(|X - \mu|^2)}{C^2} = \frac{\text{Var}\,(X)}{C^2}$$

Applying Chebychevs inequality to S_n, we note $\mathbb{E}(S_n/n) = \frac{\sum \mathbb{E}(X_i)}{n} = \frac{n\mathbb{E}(X_1)}{n} = \mathbb{E}(X_1) = \mu$
Variation is given by $\frac{1}{n^2}\text{Var}\,(S_n)$, independence yields

$$\frac{1}{n^2} n\text{Var}\,(X_1) = \frac{\text{Var}\,(X_1)}{n} = \frac{\sigma^2}{n}$$

So $\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq C\right)$. Chebychevs yields $\leq \frac{\sigma^2}{C^2 n} \to 0$ as $n \to \infty$.
Note however, it is not summable, otherwise we could have applied Borell-Cantelli, so $\frac{S_n}{n} \to \mu$ in probability.

9. Conditional Expectations

Example:
Consider the throw of a die. The outcomes are $\Omega = \{1, 2, \cdots, 6\}$. Let $X(\omega) = \omega$.
We have $\mathbb{P}(X \leq 3) = \frac{3}{6} = \frac{1}{2}$. Suppose we are given knowledge that outcome is odd and or even. Recall that the conditional *probability* of these events is given by:

$$\mathbb{P}(X \leq 3 \mid \text{ odd outcome}) = \frac{\mathbb{P}(X \leq 3, \text{ odd outcome})}{\mathbb{P}(\text{odd outcome})} = \frac{2/6}{3/6} = \frac{2}{3}$$

Conversely:

$$\mathbb{P}(X \leq 3 \mid \text{even outcome}) = \frac{\mathbb{P}(X \leq 3, \text{ even outcome})}{\mathbb{P}(\text{even outcome})} = \frac{1/6}{3/6} = \frac{1}{3}$$

The conditional expecation is in this case given by:

$$\mathbb{E}(X \mid X \text{ odd}) = \frac{1+3+5}{3} = 3 \qquad \mathbb{E}(X \mid X \text{ even}) = \frac{2+4+6}{3} = 4$$

The division by 3 is from the probability of the outcomes (3 outcomes in each case).

9.1. **Conditional expecations with respect to σ-algebras.**

Definition 9.21 Conditional expectation wrt. σ-algebra

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{G} \subseteq \mathcal{F}$ a sub σ-algebra. Let X be an integrable random variable.
$\exists$ a random variable $Y(\omega)$ which satisfies:

- Y is $\mathcal{G}$ measurable
- Y is integrable
- Reduce it to any element in $\mathcal{G}$, then $\mathbb{E}(Y) = \mathbb{E}(X)$:

$$\forall g \in \mathcal{G} \quad \int_g Y d\mathbb{P} = \int_g X d\mathbb{P}$$

Moreover, Y is unique (almost surely), since any Y' also satisfying this must satisfy that $\mathbb{P}(Y = Y') = 1$

We call this Y the *conditional expectation of* X *conditionen on* $\mathcal{G}$, and we write this as $\mathbb{E}(X \mid \mathcal{G}) = Y \Leftrightarrow$ random variable.

If the σ-algebra is $\sigma(Z)$ on $\sigma(Z_1, \cdots, Z_n)$ (generated by Z), we write $\mathbb{E}(X \mid Z) = \mathbb{E}(X \mid \sigma(Z))$

Example:
In the die example we have $\mathcal{F} = \mathcal{P}(\Omega)$. Both the even and the odd case is:

$$\mathcal{G} = \{\varnothing, \Omega, \{1,2,3\}, \{2,4,6\}\}$$

$\mathcal{G}$-measurability implies Y is a constant $\{1,3,5\}$ and $\{2,4,6\}$ (can only take 1 value for smallest piece of the σ-algebra by pre-image), so $Y(\omega) = a$ if $\omega \in \{1,3,5\}$ and $Y(\omega) = b$ if $\omega \in \{2,4,6\}$
Since the last requirement in the definition ($\forall g \in \mathcal{G}$) we have:

$$\int_{\varnothing} Y d\mathbb{P} = \int_{\varnothing} X d\mathbb{P}$$

This does not tell us anything, but it is worth noting. We continue:

$$\underbrace{Y_{\{1,3,5\}} Y d\mathbb{P}}_{=\int_{\{1,3,5\}} a d\mathbb{P}} = \int_{\{1,3,5\}} A = 1 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} = \frac{9}{6}$$

$$= a\mathbb{P}(\{1,3,5\}) = \frac{a}{2} \Rightarrow \frac{9}{6} = \frac{a}{2} \Rightarrow a = 3$$

$$\underbrace{\int_{\{2,4,6\}} Y d\mathbb{P}}_{b\mathbb{P}(\{2,4,6\})} = \int_{\{2,4,6\}} X d\mathbb{P} = 2 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 2 = \frac{b}{2} \Rightarrow b = 4$$

Obviously we have to verify for all the $g \in \mathcal{G}$:

$$\underbrace{\int_{\Omega} Y d\mathbb{P}}_{\frac{1}{2} \cdot 3 + \frac{1}{2} \cdot 4} = \int_{\Omega} X d\mathbb{P} = \frac{1+2+3+4+5+6}{6} = \frac{7}{6}$$

If $\mathcal{G}$ is a trivial σ-algebra, then $\mathbb{E}(X \mid \mathcal{G}) = \mathbb{E}(X)$ since $\int Y = \int X$ almost surely.

Philosophically, we may interpret this as a knowledge of a system.
We want to now investiage sequences of random variables without the iid constraint.

Lemma 9.1

$\mathbb{E}(X \mid \mathcal{G})$ is unqiue

Bevis 9.1

Assume we have Y, Y' satisfying 1)$\to$3) and $\mathbb{P}(Y = Y') \neq 1$. Then $\mathbb{P}(Y > Y') > 0$ (wlog, assume this). Note:

$$\{Y > Y'\} = \bigcup_{n \in \mathbb{N}} \left\{ Y \geq Y' + \frac{1}{n} \right\}$$

and for some n we have $\mathbb{P}(Y \geq Y' + \frac{1}{n}) > 0$. Y, Y' are $\mathcal{G}$-measurable $\Rightarrow Y - Y'$ is $\mathcal{G}$-measurable $\Rightarrow \left\{ Y \geq Y' 0 \frac{1}{n} \right\} = \left\{ Y - Y' \geq \frac{1}{n} \right\} \in \mathcal{G}$

So we can compare integrals and by condition 3)

$$\int_G Y d\mathbb{P} = \int_G X d\mathbb{P} = \int_G Y' d\mathbb{P} \Rightarrow \int_G Y d\mathbb{P} - \int_G Y' d\mathbb{P} \quad \forall G$$
$$= \int_G Y - Y' d\mathbb{P} = \int_{\left\{Y - Y' \geq \frac{1}{n}\right\}} \geq \frac{1}{n}\mathbb{P}(\left\{Y - Y' \geq \frac{1}{n}\right\}) \neq 0$$

Buy this is a contradiction (since 0 cannot be > 0) $\square$

10. Product Measures & Product Spaces

Given 2 measure spaces (S_1, Σ_1, μ_1) and (S_2, Σ_2, μ_2), we want to build a "canonical" measure on $S = S_1 \times S_2$
First we build the product Σ-algebra.

10.1. **Product Σ-algebra.**

The notation we shall use is $\Sigma = \Sigma_1 \times \Sigma_2 = \sigma\left(\bigcup_{A\in\Sigma_1} \times S_2 \cup \bigcup_{B\in S_2} S_1 \times B\right)$
Remark:
It is generated by a π-system on the form $\{A_1 \times A_2 : A_1 \in \Sigma_1, A_2 \in \Sigma_2\}$

If f is a bounded measurable function on (S, Σ), then we have the projection

$$\begin{aligned} S_1 \to \mathbb{R} &\qquad s_1 \mapsto f(s_1, s_2) \quad \text{fix } s_2 \\ S_2 \to \mathbb{R} &\qquad s_2 \mapsto f(s_1, s_2) \quad \text{fix } s_1 \end{aligned}$$

Where f is measurable with respect to Σ_1 and Σ_2 resp.

Bevis 10.1

It holds for indicator functions on the form $I_{A_1\times A_2}(s_1, s_2) = \begin{cases} 1 & \text{if } (s_1, s_2) \in A_1 \times A_2 \\ 0 & \text{else} \end{cases}$

Then we can extend this □

10.2. **Product Measures.**

The goal is to define a measure that works with projections.
Assume we are given 2 measures μ_1, μ_2 on (S_1, Σ_2) and (S_2, Σ_2)

$$\mathcal{I}_1^f(s_2) = \int_{S_2} f(s_1, s_2) d\mu_2 \qquad \mathcal{I}_2^f(s_2) = \int_{S_1} f(s_1, s_2) d\mu_1$$

Lemma 10.1

If f was bounded and measurable, then $\mathcal{I}_1, \mathcal{I}_2$ are bounded and measurable

Bevis 10.2

We use indicators!

$$f = I_{A_1\times A_2}, \quad \mathcal{I}_1^f(s_1) = \int_{S_2} I_{A_1\times A_2}(s_1, s_2) d\mu_2 = \int_{S_2} I_{A_1}(s_1) I_{A_2}(s_2) d\mu_2$$
$$= I_{A_1}(s_1) \int_{S_2} I_{A_2}(s_2) d\mu_2 = I_{A_1} \int_{A_2} d\mu_2 = I_{A_1}(s_1)\mu_2(A_2)$$

Which is bounded and measurable. Similarly proceed for $\mathcal{I}_2^f$ and for an arbitrary f □

Now for $F \in \Sigma$ and $f = I_F(s_1, s_2)$. Define measure of f by

$$\mu(F) = \int_{S_1} \mathcal{I}_1^f d\mu_1 = \int_{S_1}\int_{S_2} f(s_1, s_2) d\mu_2 d\mu_1 = \int_{S_2} \mathcal{I}_2^f(s_2) d\mu_2 = \int_{S_2}\int_{S_1} f(s_1, s_2) d\mu_1 d\mu_2$$

Sats 10.11: Fubinis

The measure of f for $F \in \Sigma$ and $f = I_F(s_1, s_2)$ is well defined and you may indeed swap orders of integrals. In fact, what we end up with is:

$$\int_{S_1}\int_{S_2} f d\mu_2 d\mu_1 = \int_{S_2}\int_{S_1} f d\mu_1 d\mu_2 = \int_S f d\mu$$

For all non-negative (MCT) integrable (DCT) f

Bevis 10.3: Fubinis Theorem

We consider $f = I_{A_1 \times A_2}$ and

$$\int_{S_1}\int_{S_2} I_{A_1\times A_2}(s_1,s_2) d\mu_2 d\mu_1 = \int_{S_1}\int_{S_2} I_{A_1}(s_1) I_{A_2}(s_2) d\mu_2 d\mu_1 = \int_{S_1} I_{A_1}(s_1) d\mu_1 \int_{S_2} I_{A_2}(s_2) d\mu_2$$

By symmetry of multiplication:

$$= \underbrace{\int_{S_2} I_{A_2}(s_2) d\mu_2}_{\mu_2(A_2)} \underbrace{\int_{S_1} I_{A_1}(s_1) d\mu_1}_{\mu_1(A_1)} = \int_{S_2}\int_{S_1} f d\mu_1 d\mu_2$$

For general (non-indicators), we approximate by step functions □

Generally, $\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$. In fact, μ is uniquely defined by this relationship since

$$\{A_1 \times A_2 : A_1 \in \Sigma_1, A_2 \in \Sigma_2\}$$

is defined by a π-system.
This construction defined $\mu = \mu_1 \times \mu_2$

Remark:
Fubini extends to σ-finite measures, but does not necessarily hold for non-σ-finite measures.

Example:
$(s_1, \mu_1) = [0, 1]$, and $\mu_1 =$ Lebesgue on $[0, 1]$ (this is σ-finite)
$(s_2, \mu_2) = [0, 1]$ and $\mu_2 =$ counting measure (this is not σ-finite)
Lets check if $\mu = \mu_1 \times \mu_2$ will still hold with Fubini:
Let $f(s_1, s_2) = \begin{cases} 1 & \text{if } s_1 = s_2 \\ 0 & \text{else} \end{cases}$, but:

$$\int_{S_1}\int_{S_2} f d\mu_2 d\mu_1 = \int_{S_1} 1 d\mu_1 = 1 \quad \text{since counting measure on } \{1\} \text{ is } 1$$

$$\int_{S_2}\int_{S_1} f d\mu_1 d\mu_2 = \int_{S_2} 0 d\mu_2 = 0$$

Since $0 \neq 1$ Fubinis theorem does *not* hold.

Remark:
We can iterate construction to define product measures on the form $\mu = \mu_1 \times \cdots \times \mu_n$, but in fact, construction holds for countable products

$$\mu(A_1 \times \cdots \times A_k \times S_{k+1} \times \cdots) = \mu(A_1)\mu(A_2)\cdots\mu(S_{k+1})$$

Example:
The d-dimensional Lebesgue measure $\mathcal{L}^d$ can be defined by $\underbrace{\mathcal{L}^1 \times \cdots \times \mathcal{L}^1}_{d\text{-times}}$

10.3. **Application.**
We can construct a formula for expectation of X. Suppose we have X be a non-negative random variable on $(\Omega, \mathcal{F}, \mathbb{P})$, then

$$\int\!\!\int_{\Omega} \underbrace{I(X \geq x)d\mathbb{P}}_{\text{expectation of indicator}} \quad dx = \int_0^{\infty} \mathbb{P}(X \geq x)dx$$

By Fubinis theorem:

$$\underbrace{\int_{\Omega}\int_0^{\infty} I(X \geq x)dxd\mathbb{P}}_{=X(\omega)} = \int_{\Omega} X(\omega)d\mathbb{P} = \mathbb{E}(X)$$

We shall consider the special case of product probability measure on $\mathbb{R} \times \mathbb{R}$ with densities f_X, f_Y (componentwise) and $f_{X,Y}$ is the joint density.

$$\mathbb{P}((X,Y) \in A) = \int\!\!\int_A f_{X,Y}(x,y)dxdy \quad A \in \mathcal{B}(\mathbb{R}^2)$$

We define conditional density through:

$$f_{X|Y}(x \mid y) = \begin{cases} \dfrac{f_{X,Y}(x,y)}{f_Y(y)} & \text{if } f_Y(y) \neq 0 \\ 0 \quad \text{else} \end{cases}$$

Here $f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x,y)dx$

For fixed y,

$$\int_{\mathbb{R}} f_{X|Y}(x \mid y) = \int_{\mathbb{R}} f_{X,y}(x,y)/f_Y(y)dx = \frac{1}{f_Y(y)}\int_{\mathbb{R}} f_{X,Y}(x,y)dx = \frac{f_Y(y)}{f_Y(y)} = 1$$

when $f_Y = 0$ its 0.
So $f_{X|Y(x|y)}$ is a denisty (gives rise to a probability measure). A good guess of conditional expectation given some point Y:

$$g(y) = \int_R x f_{X|Y}(x \mid y)dx = \text{ conditional expectation of } x \text{ w.r.t some } y$$

We now want to show that $g(y)$ satisfies conditions of conditional expectation. By inspection, g is $\sigma(Y)$ integrable. We need to show 3), so let $A \in \Sigma_2$, want to show

$$\int_{\{Y \in A\}} Xd\mathbb{P} = \int_{\{Y \in A\}} g(Y)d\mathbb{P}$$

LHS:

$$= \int_{\Omega} I_{Y \in A} X d\mathbb{P}$$

We use the joint denisty to express as integral over $\mathbb{R} \times \mathbb{R}$:

$$\int\!\!\int_{R \times R} I_{Y \in A} x f_{X,Y}(x,y)dxdy$$

By Fubinis theorem:

$$= \int_{\mathbb{R}} x \int_{\mathbb{R}} I_{Y \in A} f_{X,Y}(x,y)dydx$$

For the RHS, get rid of interval:

$$\int_{\Omega} I_{Y \in A} g(Y)d\mathbb{P}$$

Use the density:

$$\int\!\!\int_{\mathbb{R} \times \mathbb{R}} I_{Y \in A} g(y) f_{X,Y}(x,y)dxdy$$

By Fubinis theorem:

$$\int_{\mathbb{R}}\int_A g(y) f_{X,Y}(x,y)dydx$$

By definition of density $f_{X,Y}(x,y) = f_{X|Y}(x \mid y)f_Y(y)$ for a set of full measure (this case Lebesgue).
LHS becomes:

$$\int_{\mathbb{R}} x \int_A f_{X|Y}(x \mid y)f_Y(y)dydx$$

By Fubinis theorem:

$$\int_A f_Y(y) \underbrace{\int_{\mathbb{R}} x f_{X|Y}(x \mid y) dx}_{=g(y)} dy = \int_A f_Y(y) g(y) dy = \int_A \left(\int_{\mathbb{R}} f_{X,Y}(x,y) dx \right) g(y) dy$$

By using Fubinis theorem on the RHS, we finally get RHS = LHS and since A was arbitrarily chosen we have checked 3).

Example:
Consider random variables $X, Y \sim U([0,1] \times [0,1])$ only on the lower triangle $f_{X,Y}(x,y) = 2I_{\{x \geq y\}}$. We get

$$f_Y(y) = \int_0^1 2I_{\{x\geq y\}} dx = \int_y^1 2dx = 2 - 2y$$

$$f_{X|Y}(x \mid y) = \frac{2I_{\{x\geq y\}}}{2-2y} = \frac{I_{\{x\geq y\}}}{1-y} \qquad g(y) = \int_0^1 x \frac{I_{\{x\geq y\}}}{1-y}$$

$$= \frac{1}{1-y} \int_y^1 x dx = \frac{1}{2} \frac{(1-y^2)}{1-y} = \frac{1+y}{2} = \mathbb{E}(X \mid Y = y)$$

Example:
Let $X_1, \cdots, X_n$ be independent identically distributed random variables and let $S_n = \sum_{i=1}^n X_i$. What is $\mathbb{E}(X_1 \mid S_n)$? Well let $A \in \sigma(S_n)$, we must have

$$\int_A \mathbb{E}(X_1 \mid S_n) + \mathbb{E}(X_2 \mid S_n) + \cdots + \mathbb{E}(X_n \mid S_n) d\mathbb{P}$$
$$= \int_A X_1 + \cdots + X_n d\mathbb{P} = \int_A S_n d\mathbb{P}$$

So

$$n \int_A \mathbb{E}(X_1 \mid S_n) d\mathbb{P} = \int_A S_n d\mathbb{P} \quad \forall A$$

In particular, we can take our Ω

$$\Rightarrow \mathbb{E}(X_1 \mid S_n) = \frac{S_n}{n}$$

Is it possible to derive conditional probablity in terms of conditional epectation?

$$\mathbb{P}(A \mid \mathcal{G}) = \mathbb{E}(I_A \mid \mathcal{G})$$

Where $A \in \Sigma$. Here $\mathbb{P}(A \mid \mathcal{G})$ is a random variable. This is unique since the expectation is unique (up to a null set) and satisfies

$$\mathbb{P}\left(\bigcup_{i=1}^\infty A_i \mid \mathcal{G} \right) = \sum_{i=1}^\infty \mathbb{P}(A_i \mid \mathcal{G})$$

If $A_i \neq A_j$ for $i \neq j$

If $\mathcal{G} = \sigma(B)$ where $B \in \Sigma$ (σ-algebra condition on some event), then $\mathbb{P}(A \mid \mathcal{G})$ is a random variable and $\mathcal{G} = \{\varnothing, \Omega, B, B^c\}$. By measurability of $\mathcal{G}$, it is constant on B, B^c:

$$\mathbb{P}(A \mid \mathcal{G})(\omega) = \begin{cases} a & \omega \in B \\ b & \omega \in B^c \end{cases}$$

we get:

$$a\mathbb{P}(B) = \int_B I_A d\mathbb{P} = \int_\Omega I_A I_B d\mathbb{P} = \int_\Omega I_{A\cap B} d\mathbb{P} = \overbrace{\mathbb{P}(A \cap B)}^{=\mathbb{P}(A|B)}$$

This tells us $a = \dfrac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}$. Similarly for $b = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \mathbb{P}(B \mid A)$

10.4. **Independent random variables.**

If $X_1, \cdots, X_n$ independent. $\mathbb{E}(h(X_1, \cdots, X_n))$ where h is a function involving one or more X_i, what is $\mathbb{E}(h(X_1, \cdots, X_n) \mid X_1) = g(X_1)$ where $g = \mathbb{E}(h(x, X_2, \cdots, X_n))$.
This follows from Fubinis theroem, since $\mathcal{G} = \sigma(X_1)$, we only need to check that it satisfies conditions and events of the form $\{x_1 \in A\}$ for $A \in \Sigma$

$$\int_{\{x_1 \in A\}} h(X_1, \cdots, X_n) d\mathbb{P}$$

Recall that $\mathbb{P}^1 \times P^1$ and $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)$, so we can split up $\mathbb{P}$ into all components:

$$\int_{\mathbb{R}} I_{\{x_1 \in A\}} \underbrace{\iint \cdots \int}_{\mathbb{R}^{n-1}} h(x_1, \cdots, X_n) d(\Lambda_2 \times \cdots \times \Lambda_n) d\Lambda_1$$

The idea is to express using laws and then by independence use Fubinis theorem to be able to use indicator functions

$$= \int_{\{x_1 \in A\}} g(X_1) d\Lambda_1$$

So 3) holds, and 1-2) are immediate.

Example:
Let $X_1, \cdots, X_n$ be independent random variables and $\overline{X} = \dfrac{1}{n} \sum X_i$. What is $\mathbb{E}(\overline{X} \mid X_1)$, we can plug $h = \overline{X}$ and $g = \overline{X}$ with fixing X_1 we get $\mathbb{E}(\overline{X} \mid X_1) = g(X_1) = \dfrac{1}{n}\mathbb{E}(x + X_2 + \cdots + X_n)$, by linearity

$$\frac{x + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n)}{n} = \frac{X_1}{n} + \frac{\sum \mathbb{E}(X_i)}{n}$$

11. Martingales

Definition 11.22 Martingale

Let $X_1, \cdots,$ be a sequence of integrable random variables, we say that the sequence is a martingale if $\mathbb{E}(X_{n+1} \mid X_n, X_{n-1}, \cdots) = X_n$ almost surely

On average we have no change. The expectation on the $n+1$:th outcome given the knowledge of previous outcomes is the same as the last outcome.

Sats 11.12: Properties of martingales

(1) $\mathbb{E}(\mathbb{E}(X \mid \mathcal{G})) = \mathbb{E}(X)$
(2) If X is $\mathcal{G}$-measurable, then $\mathbb{E}(X \mid \mathcal{G}) = X$ almost surely
(3) Linearity condition still holds: $\mathbb{E}(aX + bY \mid \mathcal{G}) = a\mathbb{E}(X \mid \mathcal{G}) + b\mathbb{E}(Y \mid \mathcal{G})$ almost surely
(4) Positivity: If $X \geq 0$ almost surely, then $\mathbb{E}(X \mid \mathcal{G}) \geq 0$ almost surely irregardless of $\mathcal{G}$

Bevis 11.1: Properties of martingales

(1) $\mathbb{E}(\mathbb{E}(X \mid \mathcal{G})) \Rightarrow \int_\Omega \mathbb{E}(X \mid \mathcal{G}) d\mathbb{P} = \int_\Omega X d\mathbb{P}$
(2) X satisfies condition of conditional expectation (measurable on any subset and equality)
(3) Follows by linearity of $\int \cdots d\mathbb{P}$
(4) Suppose $Y = \mathbb{E}(X \mid \mathcal{G}) > 0$, for positive measure set A. Then $\exists n$ such that $\mathbb{P}(Y \leq \dfrac{-1}{n}) > 0$. This is a $\mathcal{G}$-measurable set and $\int_{A_n} Y d\mathbb{P} \leq \dfrac{-1}{n} \underbrace{\mathbb{P}(A_n)}_{\geq 0} = \int_{A_n} X d\mathbb{P} \geq 0$ which is a contradiction.

□

Of course, the results for integration are extended into results in expectation:

Sats 11.13

(1) If $X_1, \cdots$ are a sequence of non-negative random variables and $X_n \to X$, then $\mathbb{E}(X_n \mid \mathcal{G}) \to \mathbb{E}(X \mid \mathcal{G})$ (MCT)
(2) If $X_1, \cdots$ satisfy $|X_i| \leq Z$ (where Z is integrable and positive and $X_n \to X$), then $\mathbb{E}(X_n \mid \mathcal{G}) \to \mathbb{E}(X \mid \mathcal{G})$ (DCT)
(3) Fatous: If $X_1, \cdots$ are non-negative, then we can also take the $\lim_{n\to\infty} \inf$ out:
$$\mathbb{E}(\lim_{n\to\infty} \inf X_n \mid \mathcal{G}) \leq \lim_{n\to\infty} \inf \mathbb{E}(X_n \mid \mathcal{G})$$

We get an analogy of Jensens inequality:

Sats 11.14: Jensens inequality for martingales

Let $g : I \to \mathbb{R}$ be convex on $I \subseteq \mathbb{R}$ and assume $X : \Omega \to I$ is integrable (as well as $g(X)$), then:
$$\mathbb{E}(g(X) \mid \mathcal{G}) \geq g(\mathbb{E}(X \mid \mathcal{G}))$$
Almost surely

Dealing with random variables we must remember the 0 probability set can produce strange results, hence the "almost surely" remark when dealing with random variables.

Simplification rules:
(1) $\mathbb{E}(\mathbb{E}(X \mid \mathcal{G}) \mid \mathcal{H})$ where $\mathcal{H} \subseteq \mathcal{G}$
(2) $\mathbb{E}(ZX \mid \mathcal{G}) = Z\mathbb{E}(X \mid \mathcal{G})$ if Z is $\mathcal{G}$ measurable
(3) $\mathbb{E}(X \mid \sigma(\mathcal{G}, \mathcal{H})) = \mathbb{E}(X \mid \mathcal{G})$ is $\mathcal{H}$ is independent of $\mathcal{G}$ and X

Special cases:
- $\mathbb{E}(X \mid \mathcal{G}) = \mathbb{E}(X)$ if X is independent of $\mathcal{G}$
- $\mathbb{E}(X \mid \mathcal{G}) = X$ if X is $\mathcal{G}$-measurable
- $\mathbb{E}(\mathbb{E}(X \mid \mathcal{G})) = \mathbb{E}(X)$

The proofs of this follows by checking the constraints

The previous paragraphs were a "basic intro" to martingales, lets delve into the deeper and a bit more rigourous definitions.

11.1. **Stochastic Process.**

Definition 11.23 Discrete Stochastic Process

A sequence of random variables $X_0, X_1, \cdots$ is called a *discrete random proces*

Definition 11.24 Filtration

A *filtration* is a sequence of σ-algebras $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}$. We write $\mathcal{F}_\infty = \sigma\left(\bigcup_{i=0}^{\infty} \mathcal{F}_\rangle\right) \subseteq \mathcal{F}$

Definition 11.25 Adapted Process

We say (X_n) is adapted to the filtration $(\mathcal{F}_i)$ if X_n is $\mathcal{F}_n$-measurable $\forall n$

Definition 11.26 Martingale

A martingale is a stochastic process adapted to a filtration $(\mathcal{F}_n)$ such that
$$\mathbb{E}(X_n \mid \mathcal{F}_{\setminus -\infty}) = X_{n-1}$$

Equivalently, we can express in increments:
$$\mathbb{E}(X_n - X_{n-1} \mid \mathcal{F}_{n-1}) = \mathbb{E}(X_n \mid \mathcal{F}_{n-1}) - \underbrace{\mathbb{E}(X_{n-1} \mid \mathcal{F}_{n-1})}_{=X_{n-1}} = 0$$

Definition 11.27 Super-martingale

A super-martingale is a martingale such that $\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) \leq X_{n-1}$

Definition 11.28 Sub-martingale

Is a martingale such that $\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) \geq X_{n-1}$

Example:
Consider the standard random walk on $\mathbb{Z}$. We move with probablity $\frac{1}{2}$ either to the left or to the right.
Let $Y_1, \cdots,$ be iidrv wiht $\mathbb{P}(Y_i = 1) = \mathbb{P}(Y_i = -1) = \frac{1}{2}$
$X_0 = 0$, $X_n = \sum_i^n Y_i = X_{n-1} + Y_n$
Let $\mathcal{F}_n = \sigma(X_0, X_1, \cdots, X_n)$. Then (X_n) is adapted to $(\mathcal{F}_n)$ (measurable with respect to the σ-algebra, but it contains (X_n))
Is it a martingale?
$$\begin{aligned}\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) = \mathbb{E}(X_{n-1} + Y_n \mid \mathcal{F}_{n-1}) &= \underbrace{\mathbb{E}(X_{n-1} \mid \mathcal{F}_{n-1})}_{\mathcal{F}_{n-1}\text{-measurable}} + \underbrace{\mathbb{E}(Y_n \mid \mathcal{F}_{n-1})}_{\text{indep.}} = \mathbb{E}(Y_n) \\ &= X_{n-1} \cdot 1\frac{1}{2}1 \cdot \frac{1}{2} = 0 + X_{n-1}\end{aligned}$$
Hence a martingale. Note, no need of iid, only independence is required!

Example:
Let $Y_1, \cdots$ be independent random variables with $\mathbb{E}(Y_i) = 1$. Let $X_0 = 1$ and $X_n = X_0 \cdot \prod_{k=1}^n Y_k$. Again, X_n is adapted to $(\mathcal{F}_n)$ where $\mathcal{F}_n = \sigma(X_0, X_1, \cdots, X_n)$.
We check the martingale condition:
$$\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) = \mathbb{E}(X_0 \prod_{k=1}^n Y_k \mid \mathcal{F}_{n-1}) = \mathbb{E}(X_{n-1}Y_n \mid \mathcal{F}_{n-1}) = X_{n-1}\mathbb{E}(Y_n \mid \mathcal{F}_{n-1}) = X_{n-1}\mathbb{E}(Y_n) = X_{n-1}$$

Example:
Let X be a $\mathcal{F}$-measruable random variable. Let $\mathcal{F}_1, \cdots$ be a filtration. Then the conditional expecation with respect to previous filtration is still a martingale.
Let $X_n = \mathbb{E}(X \mid \mathcal{F}_n)$, $\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) = \mathbb{E}(\mathbb{E}(X \mid \mathcal{F}_n) \mid \mathcal{F}_{n-1})$ where $\mathcal{F}_{n-1}$ is a coarser σ-algebra $= \mathbb{E}(X \mid \mathcal{F}_{n-1}) = X_{n-1}$

Remark:
Let $m < n$. For every martingale, we have:
$$\begin{aligned}\mathbb{E}(X_n \mid \mathcal{F}_m) &= \mathbb{E}(\mathbb{E}(\cdots \mathbb{E}(\underbrace{\mathbb{E}(X_n \mid \mathcal{F}_{\setminus -\infty})}_{=X_{n-1}} \mid \mathcal{F}_{n-2}) \cdots) \mid \mathcal{F}_m) \\ &= \mathbb{E}(\mathbb{E}(\cdots \underbrace{\mathbb{E}(X_{n-1} \mid \mathcal{F}_{n-2})}_{X_{n-2}} \cdots \mid \mathcal{F}_{m+1}) \mid \mathcal{F}_m) \\ &\vdots \\ &= \mathbb{E}(X_{m+1} \mid \mathcal{F}_m) = X_m\end{aligned}$$

Definition 11.29 Pre-visible process

A pre-visible process is a seuqnce $C_1, \cdots$ of random variables such that C_n is $\mathcal{F}_{n-1}$-measurable $\forall n$

Let C_n be a pre-visible process. The martingale transform of X by C is

$$(C \cdot X)_n = \sum_{k=1}^{n} C_k(X_k - X_{k-1})$$

In particular, if $C_k = 1\ \forall k$, then

$$(C \cdot X)_n = X_n - X_0$$

Definition 11.30

If C is a bounded pre-visible process with $|C_n(\omega)| \leq K$ for all n and $\omega \in \Omega$, then $(C \cdot X)_n$ is a martingale if X_n is a martingale.

If C is also non-negative, then $(C \cdot X)_n$ is a sub/super-martingale whenever X_n is.

Bevis 11.2

We have

$$\begin{aligned} &\mathbb{E}((C \cdot X)_n - (C \cdot X)_{n-1} \mid \mathcal{F}_{n-1}) \\ &= \mathbb{E}(C_n(X_n - X_{n-1}) \mid \mathcal{F}_{n-1}) \\ &= C_n(\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) - \mathbb{E}(X_{n-1} \mid \mathcal{F}_{n-1})) \\ &= C_n(\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) - X_{n-1}) = \begin{cases} = 0 & \text{if } X_n \text{ is a martingale} \\ \geq 0 & \text{if } C_n \geq 0 \text{ and } X_n \text{ is a sub-martingale} \\ \leq 0 & \text{if } C_n \geq 0 \text{ and } X_n \text{ is a super-martingale} \end{cases} \end{aligned}$$

□

12. Stopping times

A stopping time is a random variable T with values in $\{0, 1, 2, \cdots, \infty\}$ and the property that $\{T \leq n\} = \{\omega \in \Omega : T(\omega) \leq n\} \in \mathcal{F}_n$ for all n.
Equivalently, $\{T = n\} = \mathcal{F}_n \quad \forall n$. This follows from

$$\{T \leq n\} = \{T \leq n-1\} \cup \{T = n\}$$

Examples:

- All constants are stopping times
- "First occurence", i.e $T = \min\{n : X_n = 0\}$ for an adapted process X_n
- If S, T are stopping times, then so are:
  - $\min\{S, T\} = S \wedge T$ "either stopped"
  - $\max\{S, T\} = S \vee T$ "both stopped"
- "Counting": For example, set $N_n =$ number of indices $k \leq n$ with $X_k = 0$, $T = \min\{n : N_n = 10\}$

Example:
The following are generally *not* stopping times:

$$T = \max\{n : N_n = 0\}$$

Since we cannot determine whether $N_k = 0$ for $k > n$

Also:

$$T = \min\left\{n : X_n = \sup_k X_k\right\}$$

Since $\sup_k X_k$ is not measurable with respect to $\mathcal{F}_n$. Could be larger Later

13. Stopped Processes

Let X_n be an adapted process and T a stopping time with respect to a given filtration. The stopped process X^T is

$$X_n^T(\omega) = X_{n \wedge T(\omega)}(\omega) = \begin{cases} X_{T(\omega)}(\omega) & \text{if } n \geq T(\omega) \\ X_n(\omega) & \text{if } n < T(\omega) \end{cases}$$

Sats 13.15

If X_n is a martingale/super-martingale/sub-martingale, then so is X_n^T. In particular, for every n

$$\mathbb{E}(X_{T \wedge n}) \leq \mathbb{E}(X_0) \qquad \text{super-martingale}$$
$$\mathbb{E}(X_{T \wedge n}) \geq \mathbb{E}(X_0) \qquad \text{sub-martingale}$$

Bevis 13.1

Note that $C_n^T = \underbrace{I_{\{n \leq 1\}} = 1 - I_{\{T \leq n-1\}}}_{\text{not yet stopped at time } n-1}$ is pre-visible.

We have

$$(C^T \cdot X)_n = \sum_{k=1}^{n} C_k^T (X_k - X_{k-1})$$
$$= \sum_{k=1}^{n} I_{\{k \leq T\}}(X_k - X_{k-1}) = \sum_{k=1}^{T \wedge n} (X_k - X_{k-1})$$
$$= X_{T \wedge n} - X_0$$

So $X_{T \wedge n}$ is a martingale.
Since $\mathbb{E}(\mathbb{E}(X \mid \mathcal{F})) = \mathbb{E}(X)$, the second conclusion follows $\square$

So for every fixed n, $\mathbb{E}(X_{T\wedge n}) = \mathbb{E}(X_0)$. Question is, is it true that $\mathbb{E}(X_T) = \mathbb{E}(X_0)$?
In general, **no!**

Example:

Consider the martingale $X_0 = 1$, $X_n = \begin{cases} 2X_{n-1} & \text{prob. } 1/2 \\ 0 & \text{prob.} 1/2 \end{cases}$
Let $T = \min\{n : X_n = 0\}$. Clearly $\mathbb{E}(X_T) = 0 \neq \mathbb{E}(X_0)$

Example:

Consider the simple random walk $X_0 = 0$ and $X_n = \begin{cases} X_{n-1} + 1 & \text{prob.} 1/2 \\ X_{n-1} = -1 & \text{prob.} 1/2 \end{cases}$
Define $T = \min\{n : X_n = 1\}$. This is a stopping time and one can show that $T < \infty$ almost surely.
Hence $\mathbb{E}(X_T) = 1 \neq \mathbb{E}(X_0)$
However, under simpler conditions $\mathbb{E}(X_T) = \mathbb{E}(X_0)$

Sats 13.16: Doobs Optional Stopping Theorem

Let T be a stopping time and let X be either a super-martingale or a sub-martingale. Suppose one of the following hold:

- T is bounded almost surely
- X_n is bounded and $T < \infty$ for $\omega \in \Omega$
- $\mathbb{E}(T) < \infty$ and $|X_n(\omega) - X_{n-1}(\omega)|$ for all n and almost every $\omega \in \Omega$

Then
$$\begin{aligned} \mathbb{E}(X_T) &\leq \mathbb{E}(X_0) && \text{super-martingale} \\ \mathbb{E}(X_T) &= \mathbb{E}(X_0) && \text{martingale} \\ \mathbb{E}(X_T) &\geq \mathbb{E}(X_0) && \text{sub-martingale} \end{aligned}$$

Bevis 13.2: Doobs Optional Stopping Theorem

- If T is bounded by some N almost surely, we have $T \wedge N = T$ and so:
$$\begin{aligned} \mathbb{E}(X_T) &\leq \mathbb{E}(X_0) && \text{super-martingale} \\ \mathbb{E}(X_T) &= \mathbb{E}(X_0) && \text{martingale} \\ \mathbb{E}(X_T) &\geq \mathbb{E}(X_0) && \text{sub-martingale} \end{aligned}$$
- We have $\mathbb{E}(X_{T\wedge n}) \leq, =, \geq \mathbb{E}(X_0)$ for fixed n. Since X is bounded, we can use DCT
$$\mathbb{E}(X_0) = \lim_{n\to\infty} \mathbb{E}(X_{T\wedge n}) = \mathbb{E}(\lim_{n\to\infty} X_{T\wedge n}) - \mathbb{E}(X_T)$$
- We have
$$|X_{T\wedge n} - X_0| = \left|\sum_{k=1}^{T\wedge n}(X_k - X_{k-1})\right| \leq \sum_{k=1}^{T\wedge n}|X_k - X_{k-1}| \leq K(T\wedge n) \leq KT$$
So $\mathbb{E}(KT) = K\mathbb{E}(T) < \infty$ and we can apply DCT on above

□

The following lemma is useful to show that $\mathbb{E}(T) < \infty$ for specific stopping times:

Lemma 13.1

Suppose there exists $\varepsilon > 0$ and a positive integer N such that $\mathbb{P}(T \leq n + N \mid \mathcal{F}_n) \geq \varepsilon$ for all n.
Then $\mathbb{E}(T) < \infty$

I.e, the probablity of stopping at any point within the next N steps is at least $\varepsilon > 0$

Bevis 13.3

We have

$$\begin{aligned} \mathbb{P}(T > N) &\leq 1 - \varepsilon \qquad \text{first } N \text{ step} \\ \mathbb{P}(T > 2N \mid T > n) &\leq 1 - \varepsilon \qquad \text{steps } N+1, \cdots, 2N \\ \mathbb{P}(T > 3N \mid T > 2N) &\leq 1 - \varepsilon \end{aligned}$$

So:

$$\begin{aligned} \mathbb{E}(T) &\leq N\varepsilon + 2N\varepsilon(1-\varepsilon) + 3N\varepsilon(1-\varepsilon) + \cdots \\ &= N\varepsilon(1 + 2(1-\varepsilon) + e(1-\varepsilon)^2 + \cdots) \\ &N\varepsilon\frac{1}{(1-(1-\varepsilon))^2} = \frac{N}{\varepsilon} < \infty \end{aligned}$$

□

Example:
Consider the simple random walk that we have considered in previous examples.
Take $T = \min\{n : |X_n| = a\}$, then $\mathbb{E}(T) < \infty$. It follows by taking $N = a$ and $\varepsilon = \frac{1}{2}a$

More generally, we can consider $T = \min\{n : X_n \geq a \vee X_n \leq -b\}$
Since $|X_k - X_{k-1}| = 1$, the third (or second item) of Doobs optional stopping theorem applies.
This allows us to answer questions such as:

- What is the probablity that we reach a before $-b$?
- What is the expected time for one of the two to happen?

We get:

$$\mathbb{E}(X_T) = \mathbb{E}(X_0) = 0 \quad \text{from DOST}$$

$$\Leftrightarrow \left.\begin{array}{l} a\mathbb{P}(X_T = a) + (-b)\mathbb{P}(X_T = -b) = 0 \\ \mathbb{P}(X_T = a) + \mathbb{P}(X_T = -b) = 1 \end{array}\right\}$$

$$\Rightarrow \mathbb{P}(X_T = a) = \frac{b}{a+b}$$

$$\Rightarrow \mathbb{P}(X_T = -b) = \frac{a}{a+b}$$

Now lets look at X_n^2:

$$\mathbb{E}(X_n^2 \mid \mathcal{F}_{n-1}) = \frac{1}{2}(X_{n-1}+1)^2 + \frac{1}{2}(X_{n-1}-1)^2 = X_{n-1}^2 + 1$$

It follows that

$$\mathbb{E}(X_n^2 - n \mid \mathcal{F}_{n-1}) = X_{n-1}^2 + 1 - n = X_{n-1}^2 - (n-1)$$

Hence $Y_n = X_n^2 - n$ is a martingale!
The 2nd and 3rd of DOST then apply and

$$\mathbb{E}(Y_T) = \mathbb{E}(Y_0) = 0$$

$$Y_T = X_T^2 - T = \text{ either } a^2 - T \vee b^2 - T$$

$$\Rightarrow \mathbb{E}(X_T^2) = \mathbb{E}(T) \wedge \frac{a^2 b}{a+b} + \frac{b^2 a}{a+b} = \mathbb{E}(T)$$

So we find

$$\mathbb{E}(T) = \frac{ab(a+b)}{a+b} = ab$$