Quick reminders:
\[ \limsup_{n \to \infty} a_n = \lim_{N \to \infty} \left( \sup_{n \geq N} a_n \right) \]
\[ a_1, a_2, a_3, a_4, \ldots \]
\[ (A_1 \supset A_2 \supset \ldots \supset A_N \supset A_{N+1} \supset \ldots) \text{ is any decreasing sequence has a limit (either } -\infty \text{, or it is bounded)} \]
\[ \text{Similarly, define } \liminf_{n \to \infty} a_n. \]

Example: If \( a_n \) runs through all rational numbers in \([0, 1]\), find \( \limsup a_n \) & \( \liminf a_n \).

Similar tricks for sets. If \( \{A_n\}_{n=1}^\infty \) is a sequence of sets in \( \mathbb{Q} \), then
\[ \lim A_n = \bigcap_{N=1}^{\infty} \left( \bigcup_{n \geq N} A_n \right). \]

\[ \sup \mathbb{Q} \subseteq \bigcup \mathbb{Q} \]

\[ \bigcap \mathbb{Q} \subseteq \left[ \frac{1}{5}, \frac{2}{5}, \ldots \right] \]
In words: $\bigcap_{n} A_n = \{x: x \in \text{infinitely many of } A_n\}^3$

Exercise: Represent $A_n$ as a $\{0,1\}$-valued function on $\Omega$: $\alpha \in A_n \iff \mathbb{1}_{A_n}(\alpha) = 1$. Then $x \in \bigcap_{n} A_n \iff \bigcap_{n} \mathbb{1}_{A_n}(x) = 1$

Similarly, $\bigcup_{n} A_n = \bigcap_{n} A_n$. In words, $\alpha \in \bigcup_{n} A_n$ if $\alpha$ belongs to all but finitely many sets of $\{A_n\}^3$.

Exercise: Express $\lim_{n} A_n$ in terms of $\mathbb{1}_{A_n}$, $\omega \in \Omega$.

**Algebras of sets**

Collections — sets closed under standard operations.

Algebra generated by a collection of sets: minimal algebra containing them all.

$A \subset 2^{\Omega}$ of subsets $\mathcal{A}$ of $\Omega$ algebra if

$A, B \in \mathcal{A} \Rightarrow \mathcal{A}$ algebra if

$A \cup B, A \cap B, A^c \in \mathcal{A}$
\( \sigma \)-algebra: exactly the same, but countably many operations now allowed. Big difference! Algebra generated by all rational segments in \( \mathbb{R} \) is boring; \( \sigma \)-algebra contains subset of algebraic numbers (\( \mathbb{Q} \cap \mathbb{R} \)).

**Counting algebras:**

- \( A_1 \) is finer than \( A_2 \) (or \( A_2 \) is coarser than \( A_1 \)) if \( A_1 \subseteq A_2 \).
- Finest of them all: \( 2^\mathbb{Q} \) — all subsets of \( \mathbb{Q} \).
- Coarsest: \( \{ \emptyset, \mathbb{Q} \} \).

If \( C \) is some collection of subsets, \( \sigma(C) \) is the coarsest \( \sigma \)-algebra containing \( C \).
\(-\)algebras on countable sets \(\mathcal{D}\) are easy to describe: just partition \(\mathcal{D}\) into smallest pieces of it.

Conversely, any \(\sigma\)-algebra generated by a countable partition is behaving like a \(\sigma\)-algebra on \(\mathbb{N}\).

Exercise: For \(\mathcal{D} = \mathbb{N} \times \mathbb{N}\), describe \(\mathcal{F} = \sigma(\mathcal{A} \times \mathcal{A}, \mathcal{A} \subset \mathbb{N})\), smallest \(\mathcal{A}\) containing 3.
For a σ-algebra \( \mathcal{A} \) on \( \Omega \), a measure is an \( \sigma \)-additive function (valued wherever one can add) \( \mu : \mathcal{A} \to \mathbb{R} \) (typically, \( \mathbb{R} = \mathbb{R} \)).

Additivity: \( \mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B) \)

As, for disjoint \( A, B \), \( \mu(A \cup B) = \mu(A) + \mu(B) \).

\( \sigma \)-additivity: the same true for countable families of disjoint sets:

\[ \mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n) \].

N3: want the sum on the right converge, when.

If \( \mu \) is valued in \( \mathbb{R}_{++} \) and \( \mu(\Omega) = 1 \), \( \mu \) is probability measure: convergence is not an issue.
Example: \( \mu(A) = \frac{|A|}{|Q|} \) for finite \( Q \).

Caution \( \sigma \)-additivity \( \Rightarrow \) measure on countable \( Q \)

is fully characterized by measure on finite sets:

\[ \mu(A) = \sum_{a \in A} \mu(\{a\}) \quad \text{(assume here that} \ A = 2^Q \).

Not so for merely finitely additive measures: there are \( \mu: \mu(a) = 0 \) for any \( a \in Q \), yet \( \mu(Q) = 1 \).

Example: For any ultrafilter \( \mathcal{F} \), set \( \mu(A) = \left\{ \begin{array}{ll}
1 & A \in \mathcal{F} \\
0 & \text{otherwise}
\end{array} \right. \)

Call elements of \( Q \) elementary outcomes; \( A \in \mathcal{F} \)-events.
Fubini Theorem (for countable sets)

If \( \mu \) is a non-negative measure on \( \mathcal{O}_1 \times \mathcal{O}_2 \),
then \( \mu(A_1 \times A_2) = \sum_{a_i \in \mathcal{O}_1} \mu_i(a_i) = \sum_{b \in \mathcal{O}_2} \mu_2(b) \),
where \( \mu_i(a) = \mu(a \times \mathcal{O}_2) = \sum_{b \in \mathcal{O}_2} \mu(a, b) \) i diff for \( \mu_2 \).

1. Reordering elements in an infinite sum can be tricky.

Example: \( \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} = \ln 2 \); can be reordered to any given sum (not if permuted terms are bounded).

But, if summands are positive, we are fine.

- \( S := \sum \mu(a, b) = \lim_{N \to \infty} \sum_{a_i, b \leq N} \mu(a, b) =: S_N \)
- If \( S = \infty \), then \( \sum_{a \in \mathcal{N}} \mu_i(a) \geq S_N \), so also \( \to \infty \).
- If $S < \infty$, then $\sum_{a \leq N} \mu_1(a) = S_N$, so $\sum_{a \leq N} \mu_1(a) \geq S$.

But if it overshoots, then for some $N_1$, $\sum_{a \leq N_1} \mu_1(a) = S + c, c > 0$

and then for some $k_2, l = 1, \ldots, N_1$, $\sum_{b \leq k_2} \mu(a, b) > \sum_{b \leq k_2} \mu(a, b) - \frac{c}{2N}$

So $\sum_{a \leq N} \mu(a, b) \geq (S + c) - \frac{c}{2} > S$, contradiction.

$\sum_{a \leq N} \mu(a, b) \geq (S + c) - \frac{c}{2} > S$, contradiction.
Random variables

A function $X: \Omega \rightarrow \mathbb{E}$ between two measurable spaces is a ($\mathbb{E}$-valued) random variable if $X^{-1}(A) \in \mathcal{F}_\omega$ for any $A \in \mathcal{E}_\mathbb{E}$. Easy case: $\mathbb{E}$ is countable (then we request that $X^{-1}(a) \in \mathcal{F}_\omega$ for any $a \in \mathbb{E}$) or $\mathbb{E} = \mathbb{R}$, with Borel $\sigma$-algebra (generated by intervals).

Given a r.v. $X$ (or a collection of r.v. $X_k$), we define $\sigma$-algebra $\sigma_\omega(\{X_k\}_{k \in K})$ as smallest $\sigma$-algebra such that all $X_k$ are r.v. (equivalently $\sigma$-algebra generated by $\{X_k^{-1}(A)\}, A \in \mathcal{F}_\mathbb{E}_K$)
Example \( \Omega = \{ \omega = \omega_1, \omega_2, \ldots \} \) \( \omega_k \in \{ H, T \} \).
\( \mathcal{F} = \sigma \left( \mathcal{A}_N \right) = \sigma \left( \omega_1, \ldots, \omega_N = w_1, \ldots, w_N \in \{ H, T \}^N \right) \).

Then \( \omega_k \) is \( \{ H, T \} \)-valued r.v., \( \mathcal{A}_N = \sigma (\omega_k, k \in N) \).
(Space of coin tosses).

**Random Variables & Probability**

A probability measure \( P \) on \( (\Omega, \mathcal{F}) \), and an \( E \)-valued random variable \( X \) define a probability measure on \( E \), given by \( P_X (A) = P [ X^{-1} (A) ] \).

Example For \( \Omega \) as above (coin tosses), set
\( P (\omega : \omega_1, \ldots, \omega_N = w_1, \ldots, w_N) = 2^{-N} \) for any sequence \( \omega \) (a.k.a. fair coin). Then \( P (\omega : \text{there are } k \text{ heads among first } n \text{ tosses}) = 2^{-n} \binom{n}{k} \).
Example: Consider the subset $A \subseteq \Omega$ given by $A = \{w: w_1 = H, \text{ and first double } \text{TT}\}$. Find $\mathbb{P}(A)$.

Conditional Probability

If $A, B$ are events with $\mathbb{P}(B) > 0$, we set

\[
\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad [\mathbb{P}(B|B) = 1]
\]

Thus, if $\{B_n\}_n$ is a partition of $\Omega$ (i.e. $\Omega = \bigcup B_n$, $B_k \cap B_l = \emptyset$, $k \neq l$), then

\[
\mathbb{P}(A) = \sum \mathbb{P}(A|B_k) \mathbb{P}(B_k) \quad [\text{"Total prob. law"}]
\]

Def: Events $A, B$ are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \mathbb{P}(B)$

Exercise: $A, B$ are independent iff $A_i, B_i$ are.
Expectation & Conditional Expectation

For a random variable with countable range \( X(\Omega) =: \mathcal{E} \)
we define
\[
\mathbb{E} X = \sum_{a \in \mathcal{E}} a \cdot \mathbb{P}(X = a)
\]
(for countable \( \Omega \),
this is the same as
\[
\sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}(\omega) - \text{ by TBM.}
\]

Good way to think of \( \mathbb{E} \) as a linear functional,
taking values in wherever \( \Omega \) lives.

**Def:** \( X, Y \) are independent if
\[
\mathbb{P}(X = a, Y = b) = \mathbb{P}(X = a) \cdot \mathbb{P}(Y = b)
\]

**Proposition:** \( X, Y \) independent \( \Rightarrow \) \( \mathbb{E} XY = \mathbb{E} X \cdot \mathbb{E} Y \)
\[
\mathbb{E} XY = \sum_{c} c \cdot \mathbb{P}(XY = c) = \sum_{c} \sum_{a, b : ab = c} a b \cdot \mathbb{P}(X = a, Y = b) = \sum_{a, b : ab = c} a b \cdot \mathbb{P}(X = a) \cdot \mathbb{P}(Y = b)
\]
Def For a random variable $X$, and an event $B$, \[ E(X|B) = \sum a \cdot P(X = a | B) = \frac{\sum a \cdot P(X = a, B)}{P(B)} \]

Def For a countable $\sigma$-algebra $\mathcal{A}$, $Y = E(X|\mathcal{A})$ is an $\mathcal{A}$-measurable random variable such that $E(Z|X) = E(Z|Y)$ for any $\mathcal{A}$-measurable $Z$. Or \[ E(X|\mathcal{A})(\omega) = E(X|\mathcal{E}) \text{ for } \mathcal{E} \text{ an elementary partition} \]

$E X$ is just $E(X|\mathcal{A})$ for $\mathcal{A} = \{∅, \Omega\}$.

$E(X|B)$ is $E(X|\mathcal{A})$ on $B$, for $\mathcal{A} = \{∅, B, B^c, \Omega\}$.
If \( X \) is already \( \mathcal{A} \)-measurable, \( \mathbb{E}(X|\mathcal{A}) = X \)

If \( \mathcal{A}_1 \) is finer than \( \mathcal{A}_2 \) (\( \mathcal{A}_2 \subseteq \mathcal{A}_1 \)), then
\[
\mathbb{E}(\mathbb{E}(X|\mathcal{A}_2)|\mathcal{A}_1) = \mathbb{E}(X|\mathcal{A}_2).
\]
\[
\mathbb{E}(\mathbb{E}(X|\mathcal{A}_1)|\mathcal{A}_2) = \mathbb{E}(X|\mathcal{A}_2)
\]

Good way to think about \( \mathbb{E}(X|\mathcal{A}) \) as a projector from space of all r.v. to subspace of \( \mathcal{A} \)-meas. r.v.

Filtrations

Increasing families of \( \sigma \)-algebras are called filtrations (typically they model information flow)

Exercise: Find \( \mathbb{E}((X+Y)^2|\mathcal{A}) \), where \( \mathcal{A} = \mathcal{A}_X \), generated by \( X \)
Moments of a r.v.

\[ M_k := \mathbb{E} X^k. \]  The numbers (if they are finite!) can be combined into exponential generating \emph{func}tions \[ M_x(t) = \sum_{k=0}^{\infty} M_k \frac{t^k}{k!} = \mathbb{E} e^{tX} = \sum \frac{e^t}{a} \cdot P(X=a) \]

For independent \( X, Y \) \[ M_{x+y}(t) = M_x(t) \cdot M_y(t) \]

Of course \[ \mathbb{E} X^k = \left( \frac{d}{dt} \right)^k M_x(t) \big|_{t=0} \]

Taking \( L_x := \log M_x \) gives the \textit{cumulant} \textit{exponential generating function,}

\[ L_x(t) = \sum \frac{C_k(X)}{k!} \frac{t^k}{k!}, \quad C_i = M_i = \mathbb{E} X^i; \]

\textit{Exercise} \[ C_2 = \mathbb{E}(X^2) - (\mathbb{E} X)^2 = \text{Var} X. \]
Convex functions & Jensen inequality

If \( f \) is convex, i.e.
\[
f\left(\frac{x_1 + x_2}{2}\right) \leq \frac{f(x_1) + f(x_2)}{2}
\]
then
\[
f(\mathbb{E}X) \leq \mathbb{E}(f(X)).
\]

Proof. For any \( x \), there is a linear function \( l(\xi) = f(x) + k \cdot (\xi - x) \) such that \( f(\xi) \geq l(\xi) \). Then
\[
\mathbb{E}f(X) \geq f(x) + k \mathbb{E}X.
\]
Choosing \( x = \mathbb{E}X \) gives the desired.