1 Elements

Measure: Given a σ-algebra $\mathcal{F}$ of subsets of Ω, a measure is a non-negative, countably additive set function $m : \mathcal{F} \to [0, \infty]$ such that $m(\phi) = 0$.

- A measure is finite if $m(\Sigma) < \infty$. It is a probability measure if $m(\Sigma) = 1$.
- A pre-measure $\nu : \mathcal{A} \to [0, \infty]$ defined on the algebra $\mathcal{A}$ is a countably additive non-negative set function, assigning $m(\phi) = 0$.
- A measure μ on $(\Omega, \mathcal{F})$ is σ-finite if there exists $\bigcup_{n\in\mathbb{N}} E_n = \Omega$ with $\mu(E_n) < \infty$.

Measurable Function: A measurable function $X : (\Omega, \mathcal{F}, \mu) \to \mathcal{G}$ where $\mathcal{G}$ is a metric space is $\mathcal{F}$-measurable if for every Borel set $B \in \mathcal{B}(\mathcal{G})$, $X^{-1}(B) \in \mathcal{F}$.

- The Borel sets of $\mathcal{G}$ comprise the smallest sigma-algebra generated by the open sets in $\mathcal{G}$ (we have open sets because it is a metric space and hence a topological space).
- This measurable function provides an **induced measure** $\mathcal{G}$, given by $\mu_X = \mu \circ X^{-1}$.
- This measure, in turn, induces a **distribution function** $F_X(t) = \mu_X((-\infty, t])$.
- Question: Can we go the other way? Given a non-decreasing, right-continuous function $F : \mathbb{R} \mapsto \mathbb{R}$, can I construct μ_F such that $\mu_F(I) = F(b) - F(a)$ for every $I = (a, b]$?

  The **Caratheodory-Hahn Extension Theorem** says: yes.

Sigma-Algebra: A σ-algebra $\mathcal{F} \subset \mathcal{P}(\Omega)$, where Ω is a set of events, is closed under countable unions and complements.

- The sigma-algebra generated by some set is the smallest sigma-algebra containing that set (i.e. the intersection of all sigma-algebras containing that set).
- The sigma-algebra generated by a random variable $X : \Omega \to \mathbb{R}$ is the sigma-algebra generated by the set of pre-images of X, i.e. $\sigma(X) = X^{-1}(\mathcal{B}(\mathbb{R}))$.
- A **filtration** $\mathcal{F}_n$ is a increasing sequence of sigma-algebras, usually generated as $\mathcal{F}_n = \sigma(X_0, X_1, ...)$

Independence: Two events A, B are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

- A set of events are independent if, for every subset, the probability of the intersections is the product of the events.
- Two random variables X, Y are independent if, for all $B_1, B_2 \in \mathbb{R}$, $\{X \in B_1\}$ is independent of $\{Y \in B_2\}$.

- A sequence of events is independent if every finite subset is independent. Likewise with a sequence of random variables.

Types of Convergence: Consider $(X_n)_{n\in\mathbb{N}}$ a sequence of random variables.

- **Pointwise**: for all $\omega \in \Omega$, $\lim_n X_n(\omega) = X(\omega)$.
- **Almost everywhere**: for a.e. $\omega \in \Omega$, $\lim_n X_n(\omega) = X(\omega)$.
- **In probability**: for all $\epsilon > 0$, we have $\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \epsilon) = 0$.
- **Complete**: for all $\epsilon > 0$, we have $\sum_{n\in\mathbb{N}} \mathbb{P}(|X_n - X| \geq \epsilon) < \infty$.
- **In Distribution**: Let F be the distribution of X and F_n be the distribution of X_n. For all t a continuity point of F, we have $\lim_{n\to\infty} F_n(t) = F(t)$. Equivalently, we have:
$$\lim_{n\to\infty} \mu_n(X_n \leq t) = \mu(X \leq t)$$
- **Vague/Weak**: For all Φ continuous and bounded, $\lim_{n\to\infty} \mathbb{E}(\Phi(X_n)) = \mathbb{E}(\Phi(X))$.
- **Lp**: $\lim_{n\to\infty} \mathbb{E}(\|X_n - X\|^p) = 0$

Some notable facts.

- Convergence almost everywhere implies convergence in probability.

  If a sequence converges in probability, it converges completely along a subsequence.
- Complete convergence implies convergence almost everywhere by Borel-Cantelli.
- Convergence in Lp implies convergence almost everywhere along a subsequence.
- Convergence in probability is metrizable (i.e. there exists a metric on random variables which is equivalent), but convergence almost everywhere is not.

2 Limits

2.1 Inferior and Superior

Recall the definitions of limit superior and inferior for sequences. These are important concepts because the limit does not always exist; in order to show it exists, rigorously, we need to show that the limit inferior equals the limit superior.

$$\liminf_{n\to\infty} x_n = \lim_{k\to\infty} \inf_{n\geq k} x_n = \sup_{k\in\mathbb{N}} \inf_{n\geq k} x_n$$

$$\limsup_{n\to\infty} x_n = \lim_{k\to\infty} \sup_{n\geq k} x_n = \inf_{k\in\mathbb{N}} \sup_{n\geq k} x_n$$

We define the definition of limit superior and inferior for sets similarly.

$$\liminf_{n\to\infty} E_n = \{E_n, ev.\} = \bigcup_{k\in\mathbb{N}} \bigcap_{n\geq k} E_n$$

$$\limsup_{n\to\infty} E_n = \{E_n, i.o.\} = \bigcap_{k\in\mathbb{N}} \bigcup_{n\geq k} E_n = \{\sum_{n\in\mathbb{N}} \mathbf{1}_{E_n} = \infty\}$$

The same idea is behind all of these. The inner process is, in the case of the limit inferior, an increasing sequence (the infimum of a smaller and smaller thing), and we take the supremum (or union) over this.

2.2 Major Limit Theorems

Take probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of **non-negative** random variables. Let $(E_n)_{n\in\mathbb{N}}$ be a sequence of events.

Monotone Convergence Theorem: If $0 \leq X_1 \leq X_2 \leq ...$ pointwise, and $X_n \to X$ almost everywhere, then:

$$\lim_{n\to\infty} \int_\Omega X_n = \int_\Omega X$$

Fatou's Lemma: No additional hypotheses.

$$\int_\Omega \liminf_{n\to\infty} X_n \leq \liminf_{n\to\infty} \int_\Omega X_n$$

Dominated Convergence Theorem: Suppose that there exists an **integrable** Y with $|X_n| \leq Y$ pointwise for all n. Then, if $X_n \to X$ almost everywhere, then we may interchange limit and integral.

$$\lim_{n\to\infty} \int_\Omega X_n = \int_\Omega X$$

3 Foundations of Measure Theory

We have defined a measure, but we have yet to actually construct the most important measure of all: the **Lebesgue measure** on the Borel sets of the real line! We also want a process for obtaining the **Lebesgue-Stieltjes measure** given a non-decreasing right continuous (i.e. distribution) function.

The process is as follows:

1. We reduce the problem of finding a measure to the problem of finding an outer measure.
2. We construct an outer measure in terms of a very minimal set function $\nu : \xi \to [0, \infty]$.
3. If this set function ν is a pre-measure (with ξ an algebra), then ν and μ^* agree on ξ; and if ν is σ-finite, then μ is the unique measure on $\sigma(\xi)$ that restricts to ν on ξ.

First, we establish two definitions

- **Outer Measure**: A set function $\mu^* : \mathcal{P}(\Omega) \mapsto [0, \infty]$ which assigns the empty set zero measure, preserves inequality between subsets, and is countably subadditive.
- **Pre-measure**: $\nu : \xi \to [0, \infty]$ is a pre-meaasure if it assigns the empty set zero measure and it is countably additive.

Caratheodory Characterization: Let $\mu^* : \mathcal{P}(\Omega) \mapsto [0, \infty]$ be an outer measure and consider the family $\mathcal{M}$ of subsets E of Ω which satisfy:

$$\mu^*(A) \geq \mu^*(A \cap E) + \mu^*(A \cap E^C)$$

Then $\mathcal{M}$ is a sigma-algebra and $\mu^*|_{\mathcal{M}}$ is a measure.

Caratheodory Construction of an Outer Measure: Take $\nu : \xi \to [0, \infty]$ which only satisfies $\phi, \Omega \in \xi$ and $\nu(\phi) = 0$. Then the following is an outer measure on $\mathcal{P}(\Omega)$:

$$\mu^* = \inf \left\{ \sum_{n \in \mathbb{N}} \nu(E_n) : (E_n)_{n \in \mathbb{N}} \in \Omega, A \subset \bigcup_{n \in \mathbb{N}} E_n \right\}$$

Hahn Extension Theorem: Take the same setup as before. Assume ξ is an algebra and ν is a pre-measure (countably additive, assigns zero measure to the empty set). Then, if we define $\mu = \mu^*|_{\sigma(\xi)}$, we have the following:

1. $\sigma(\xi) \subset \mathcal{M}$ and $\mu|_\xi = \mu^*|_\xi = \nu$.
2. For any measure ρ on $\sigma(\xi)$ which satisfies $\rho|_\xi = \nu$, we have the following (with equality when $\mu(A) < \infty$),
$$\rho(A) \leq \mu(A) \text{ for all } A \in \sigma(\xi)$$
3. If ν is σ-finite, then μ is the unique measure of $\sigma(\xi)$ with $\mu|_\xi = \nu$, i.e. the unique extension of the pre-measure ν on ξ to a measuer $\sigma(\xi)$.

4 Major Classical Results

4.1 Zero-One Laws

Borel-Cantelli Lemmata: Given a sequence of events $(E_n)_{n\in\mathbb{N}}$, we have:

- If $\sum_{n\in\mathbb{N}} \mathbb{P}(E_n) = \infty$, then $\mathbb{P}(E_n, i.o.) = 0$.
- If $\sum_{n\in\mathbb{N}} \mathbb{P}(E_n) = \infty$ and $(E_n)_{n\in\mathbb{N}}$ is independent, then $\mathbb{P}(E_n, i.o.) = 1$.

Kolmogorov Zero-One Law: Take a sequence of random variables $(X_n)_{n\in\mathbb{N}}$. Consider the tail sigma-algebra, defined as follows.

$$\tau = \bigcap_{n\in\mathbb{N}} \tau^n = \bigcap_{n\in\mathbb{N}} \sigma(X_n, X_{n+1}, ...)$$

If $(X_n)_{n\in\mathbb{N}}$ is independent, then for all $A \in \tau$, $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$.

In other words, modulo sets of measure zero, $\tau = \{\Sigma, \phi\}$. The tail sigma-algebra is trivial.

Hewitt-Savage Zero-One Law: Define $\xi_n \subset \mathcal{F}_n$ as the collection of sets that are invariant under permutation of the first n coordinates, i.e. for all $(X_1 \in B_1, ..., X_n \in B_n) \in \xi_n$ and $\sigma \in S_n$, we have $\pi(A) = (X_{\pi(1)} \in B_{\pi(1)}, ..., X_{\pi(n)} \in B_{\pi(n)}) \in \xi_n$. Let $\xi = \bigcap_{n\in\mathbb{N}} \xi_n$.

If $X_1, X_2, ...$ is independent, then ξ is trivial (analogous to the Kolmogorov zero-one law).

4.2 Classical Concentration of Measure

Weak Law of Large Numbers: If $(X_n)_{n\in\mathbb{N}}$ are independent and identically distributed such that $\lim_{n\to\infty} n\mathbb{P}(|X_1| > n) = 0$, then $\frac{1}{n}S_n - \mu_n \to 0$ in probability, where $\mu_n = \mathbb{E}(X_n \cdot \mathbf{1}_{|X_1|\leq n})$

- The weak law can apply in cases where we have infinite expectation. Consider X_k such that $\mathbb{P}(X_1 = \pm n) = \frac{c}{n^2 \log(n)}$. The idea is that $\mathbb{P}(X_n \geq k)$ is on the order of $1/(n\log(n))$, the series of which does not converge. So $\mathbb{E}(X_n) = \infty$). However, it does satisfy $\lim_{n\to\infty} n\mathbb{P}(|X_1| > n) = 0$. So WLLN applies and SLLN does not; in fact, we know convergence almost everywhere cannot happen because by Borel-Cantelli, $\mathbb{P}(|S_n| > n/2, i.o.) = 1$. So the limsup and liminf of the sample mean is plus or minus infinity in this case.

(Etemadi) Strong Law of Large Numbers: Given $(X_n)_{n\in\mathbb{N}}$ pairwise independent, identically distributed, with $\mathbb{E}(|X_1|) < \infty$ we have:

$$\frac{1}{n}\sum_{i=1}^{n} X_n \to 0 \text{ , a.e.}$$

- We can use this to prove a certain weak convergence of an observed distribution to a true distribution.
- Recall Kolmogorov's zero-one law. Recall that $\{\lim_n S_n/n \text{ exists in } \mathbb{R}\}$ is in the tail sigma-algebra (it is not affected by changing finitely many elements in the sequence). So this told us that this event is in a trivial tail sigma algebra. The SLLN assures us that it is in a set of full measure rather than zero measure.

Central Limit Theorem: Given independent and identically distributed $(X_n)_{n\in\mathbb{N}}$ with finite variance σ^2 and expectation m, we have the following convergence in distribution.

$$\frac{\sum_{i=1}^n (X_i - m)}{\sigma\sqrt{n}} \to \mathcal{N}(0,1)$$

- Standard proof uses equivalence of vague convergence and convergence in distribution.

4.3 Markov Chains

When do Markov Chains possess an invariant measure:

- If Markov chain is **irreducible** and **recurrent**, then it has an invariant measure (converse is false; take the simple symmetric random walk on the number line). It is given by $(\gamma_i^k)_{i\in S}$. Note $\gamma_k^k = 1$. This is minimal and unique up to scalar multiplication. Note also that $m_k = \mathbb{E}_k(T_k) = \sum_i \gamma_i^k$.

$$\gamma_i^k = \mathbb{E}_k(\sum_{n=1}^{T_k - 1} \mathbf{1}_{\{X_n = i\}})$$

- If Markov chain is irreducible, then it is positive recurrent if and only if it has an invariant probability distribution. The distribution is given by $(\gamma_i^k/m_k)_{i\in S}$; this is well-defined because $m_k < \infty$!
- If the Markov chain is irreducible, then it is clearly also closed (you can only stay within this communicating class). So if it is **finite** then, by class properties theorem, it is **recurrent**. In fact, it is positive recurrent because there are only finitely many γ_i^k so they add up to a finite m_k. So there is an **invariant probability measure**!

Convergence to Equilibrium: If $(X_n)_{n\in\mathbb{N}}$ is $MC(P,\lambda)$ irreducible, recurrent, and possesses an invariant distribution π. Then the following limit exists for $j \in S$ the state space:

$$\lim_{n\to\infty} \mathbb{P}(X_n = j) = \pi_j$$

In particular, $\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j$. It does not matter my starting state i.

4.4 $\mathbb{L}^p$ Spaces

Let $\mathbb{L}^p$ be the set of random variables such that the following quantity is finite.

$$\|X\|_p = \Big(\int_\Omega |X|^p d\mathbb{P} \Big)^{1/p}$$

Minkowksi Inequality: General triangle inequality

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

Jensen's Inequality: For f convex and X integrable, we have:

$$f(\mathbb{E}(X))) \leq \mathbb{E}(f(X))$$

Lyapunov Inequality: For $0 < p < q \leq \infty$, $\mathbb{L}^q \subset \mathbb{L}^p$ since:

$$\|X\|_p \leq \|X\|_q$$

Hölder's Inequality: For $\frac{1}{p} + \frac{1}{q} = 1$

$$\|XY\|_1 \leq \|X\|_q \|Y\|_p$$

Cauchy-Schwarz Inequality: Special case of Hölder for $p = q = 2$. Special because $\mathbb{L}^2$ is famous a Hilbert space and the the left-hand side is the inner product.

$$\langle XY \rangle \leq \|X\|_2 \|Y\|_2$$