21-721 Probability
Spring 2010

Prof. Agoston Pisztora
notes by Brendan Sullivan

May 6, 2010

Contents

0 Introduction

Any claim marked with (***) is meant to be proven as an exercise.

1 Measure Theory

1.1 σ-Fields

Let $\Omega \neq \emptyset$ be some set (of all possible outcomes).

Definition 1.1. *A collection $\mathcal{F}$ of subsets of Ω is called a σ-field provided*

1. $\Omega \in \mathcal{F}$
2. $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
3. $A_1, A_2, \dots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

The pair $(\Omega, \mathcal{F})$ is called a measurable space *and the sets $A \in \mathcal{F}$ are called ($\mathcal{F}$-observable) events.*

Remark 1.2. If (3) is required for only finite unions, then $\mathcal{F}$ is merely called an algebra (field). Using DeMorgan's Laws, one can conclude that countable (resp. finite) intersections stay within a σ-field (resp. algebra).

Example 1.3.

1. $\mathcal{F} = \mathcal{P}(\Omega)$ is a (maximal) σ-field. $\mathcal{F} = \{\emptyset, \Omega\}$ is a (trivial) σ-field.
2. Let $\mathcal{B}$ be an arbitrary collection of subsets $\mathcal{B} \subseteq \mathcal{P}(\Omega)$. Then

$$\sigma(\mathcal{B}) := \bigcap_{\substack{\hat{\mathcal{F}}:\sigma\text{-field} \\ \mathcal{B} \subseteq \hat{\mathcal{F}}}} \hat{\mathcal{F}}$$

is the σ-field *generated* by $\mathcal{B}$. (The intersection is nonempty since $\hat{\mathcal{F}} = \mathcal{P}(\Omega)$ is always valid.) It is, indeed, a σ-field since an arbitrary intersection of σ-fields is also a σ-field (***), and by construction it is the *smallest* σ-field containing $\mathcal{B}$.

3. Let Ω be a topological space where τ is the collection of open sets. Then $\sigma(\tau)$ is called the Borel σ-field.

4. Let $Z = (A_i)_{i \in I}$ be a countable partition of Ω; i.e. I is countable and $\Omega = \underline{\bigcup}_{i \in I} A_i$ (disjoint union). Then $\sigma(Z)$ is an *atomic* σ-field with atoms $A_1, A_2, \ldots$, and it can be written as (***)

$$\sigma(Z) = \left\{ \bigcup_{j \in J} A_j : J \subseteq I \right\}$$

Note: if Ω is countable then *every* σ-field is of this form.

5. Time evolution of a random system with state space $(S, \mathfrak{S})$ which is, itself, a measurable space. Set

$$\Omega = S^{\mathbb{N}} = \{\omega = (\omega_0, \omega_1, \ldots) : \omega_i \in S\}$$

to be the set of all *trajectories* in S. The map $X_n : \Omega \to S$ defined by $\omega \mapsto \omega_n$ indicates the state of the system at time n. For $A \in \mathfrak{S}$, write

$$\{X_n \in A\} = \{\omega \in \Omega : X_n(\omega) \in A\} = X_n^{-1}(A)$$

as the event that at time n, the system was in A. Then

$$\mathcal{B}_n := \{\{X_n \in A\} : A \in \mathfrak{S}\}$$

is the collection of "at time n observable events". In fact, $\mathcal{B}_n$ is automatically a σ-field. To see why, we show:

   (a) $\Omega = \{X_n \in S\} \in \mathcal{B}_n$

   (b) If $B \in \Omega$ then $\exists A \in \mathfrak{S}$ such that $B = \{X_n \in A\}$, but then $B^c = \{X_n \in A^c\} \in \mathcal{B}_n$, too.

   (c) Similarly, if $B_1, \cdots \in \mathcal{B}_n$ then $\exists A_1, \cdots \in \mathfrak{S}$ such that $B_n = \{X_n \in A_k\}$ then $\bigcup_{i \geq 1} = \{X_n \in \bigcup_i A_i\} \in \mathcal{B}_n$.

We set

$$\mathcal{F}_n := \sigma\left(\bigcup_{k \leq n} \mathcal{B}_k\right)$$

to be the "up to time n observable events". Similarly, we set

$$\mathcal{F} := \sigma\left(\bigcup_{n \geq 0} \mathcal{B}_n\right)$$

to be the σ-field of all observable events. Note that, a priori, the expressions in the parentheses $\sigma(\cdot)$ above are *not necessarily* σ-fields themselves.

One can show that (***)

$$\mathcal{F}_n = \sigma\{\bigcap_{i=0}^{n}\{X_i \in A_i\} : A_i \in \mathfrak{S}\}$$

and

$$F = \sigma\left(\bigcap_{i\geq 0}\{X_i \in A_i\} : A_i \in \mathfrak{S}\right)$$

We also set

$$\mathcal{F}_n^\star := \sigma\left(\bigcup_{k\geq n} \mathcal{B}_k\right)$$

to be the σ-algebra of "after time n observable events" and

$$\mathcal{F}^\star := \bigcap_{n\geq 0} \mathcal{F}_n^\star$$

to be the "tail field" or asymptotic field. Is $\mathcal{F}^\star$ trivial? NO!

Example 1.4.

$$\{X_n \in A \text{ i.o.}\} = \bigcap_{n\geq 0}\bigcup_{k\geq n}\{X_k \in A\} \in \mathcal{F}^\star$$

where "i.o." means "infinitely often". To see why, set

$$C_n = \bigcup_{k\geq n}\{X_k \in A\} \in \mathcal{F}_n^\star$$

WWTS $\bigcap_{n\geq 0} C_n \in \mathcal{F}_k^\star$ for each fixed k (and the claim above follows directly). Notice that $C_{n+1} \subseteq C_n$ so

$$\bigcap_{n\geq 0} = \bigcap_{n\geq k} C_n \in \mathcal{F}_k^\star$$

and we're done.

1.2 Dynkin Systems

Definition 1.5. *A collection $\mathcal{D} \subseteq \mathcal{P}(\Omega)$ is called a* Dynkin system *provided*

1. $\Omega \in \mathcal{D}$
2. $A \in \mathcal{D} \Rightarrow A^c \in \mathcal{D}$
3. $A_1, A_2, \ldots$ *disjoint and* $A_i \in \mathcal{D} \Rightarrow \bigcup_i A_i \in \mathcal{D}$

Note that these conditions are less restrictive than a σ-field.

Remark 1.6. If $A \subseteq B$ with $A, B \in \mathcal{D}$ then $B \setminus A = (B^c \dot{\cup} A)^c \in \mathcal{D}$, as well. If $\mathcal{D}$ is $\cap$-closed then $\mathcal{D}$ is a σ-field. To see why, observe that:

1. If $A, B \in \mathcal{D}$ then

$$A \cup B = \underbrace{(A \setminus (A \cap B))}_{\in \mathcal{D}} \dot{\cup} \underbrace{(A \cap B)}_{\in \mathcal{D}} \dot{\cup} \underbrace{(B \setminus (A \cap B))}_{\in \mathcal{D}} \in \mathcal{D}$$

2. If $A_1, \cdots \in \mathcal{D}$, set $B_k = \bigcup_{i \leq k} A_i$. Then

$$\bigcup_{i=1}^{\infty} A_i = \dot{\bigcup}_{n \geq 1} \underbrace{(B_n \setminus B_{n-1})}_{\in \mathcal{D}} \in \mathcal{D}$$

Also, note that if $\mathcal{B} \subseteq \mathcal{P}(\Omega)$ then

$$\mathcal{D}(\mathcal{B}) := \bigcap_{\substack{\hat{\mathcal{D}}:\text{D.S.} \\ \mathcal{B} \subseteq \hat{\mathcal{D}}}} \hat{\mathcal{D}}$$

is a Dynkin System (the smallest DS containing $\mathcal{B}$, by construction).

Theorem 1.7 (Dynkin's π-λ Theorem)**.** *Let $\mathcal{B} \subseteq \mathcal{P}(\Omega)$ and assume $\mathcal{B}$ is $\cap$-closed. Then $\mathcal{D}(\mathcal{B})$ is also $\cap$-closed and therefore a σ-field; furthermore, $\mathcal{D}(\mathcal{B}) = \sigma(\mathcal{B})$.*

Note: $\mathcal{D}(\mathcal{B}) \subseteq \sigma(\mathcal{B})$ always, but if $\mathcal{D}(\mathcal{B})$ is itself a σ-field then it cannot be strictly smaller than $\sigma(\mathcal{B})$, by definition. Thus, it suffices to show $\mathcal{D}(\mathcal{B})$ is $\cap$-closed.

Proof. First, let $B \in \mathcal{B}$ and set

$$\mathcal{D}_B := \{A \in \mathcal{D}(\mathcal{B}) : A \cap B \in \mathcal{D}(\mathcal{B})\}$$

Observe that $\mathcal{D}_B$ is, in fact, a DS (***) containing $\mathcal{B}$ and thus $\mathcal{D}_B \supseteq \mathcal{D}(\mathcal{B})$. The reverse containment is trivial, so we have $\mathcal{D}_B = \mathcal{D}(\mathcal{B})$.

Second, let $A \in \mathcal{D}(\mathcal{B})$ and set

$$\mathcal{D}_A := \{C \in \mathcal{D}(\mathcal{B}) : A \cap C \in \mathcal{D}(\mathcal{B})\}$$

Observe that $\mathcal{D}_A$ is, in fact, a DS (***) containing $\mathcal{B}$ (by the first part of this proof), and so $\mathcal{D}_A \supseteq \mathcal{D}(\mathcal{B})$. Again, the reverse is trivial, so $\mathcal{D}_A = \mathcal{D}(\mathcal{B})$. ∎

1.3 Probability Measures

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

Definition 1.8. *A* probability distribution *(prob. measure) P is a positive measure on Ω with total mass 1; that is, $P : \mathcal{F} \to [0, 1]$ with $A \mapsto P[A]$ (prob. of event A) and with the following properties (the Axioms of Kolmogorov):*

1. *$P[\Omega] = 1$*
2. *$A_1, \dots \in \mathcal{F}$ with $A_i \cap A_j = \emptyset$ for $i \neq j$ then*

$$P\left[\bigcup_{i=1}^{\infty} A_i\right] = \sum_{i=1}^{\infty} P[A_i]$$

This is called being "σ-additive".

Some properties of probability distributions:

1. $P[A^c] = 1 - P[A]$
2. $A \subseteq B \Rightarrow P[A] \leq P[B]$ (monotonicity)
3. $P[A \cup B] = P[A] + P[B] - P[A \cap B] \leq P[A] + P[B]$ (subadditivity)
4. If $A_1 \subseteq A_2 \subseteq \cdots$ then

$$\lim_{n\to\infty} P[A_n] = P\left[\bigcup_{i=1}^{\infty} A_i\right]$$

This is known as monotone continuity. Taking complements, one can show that if $B_1 \supseteq B_2 \supseteq \cdots$ then i

$$\lim_{n\to\infty} P[B_n] = P\left[\bigcap_{i=1}^{\infty} A_i\right]$$

To prove the first property, take an ascending collection A_k. Set $A_0 = \emptyset$ and $D_k = A_k \setminus A_{k-1}$. Then

$$A_n = \bigsqcup_{1\leq k\leq n} D_k \Rightarrow P[A_n] = \sum_{k=1}^{n} P[D_k]$$

by σ-additivity. Since

$$\bigcup_{n\geq 1} A_n = \bigsqcup_{k\geq 1} D_k$$

then we have

$$P\left[\bigcup_{n\geq 1} A_n\right] = \sum_{k=1}^{\infty} P[D_k] = \lim_{n\to\infty} \sum_{k=1}^{n} P[D_k] = \lim_{n\to\infty} P[A_n]$$

5. For any countable collection $(A_k)_{k\geq 1}$,

$$P\left[\bigcup_{k\geq 1} A_k\right] \leq \sum_{k=1}^{\infty} P[A_k]$$

This property is known as σ-subadditivity. To prove it, notice that the collection $\bigcup_{k\leq n} A_k$ is ascending in n, so

$$\begin{aligned} P\left[\bigcup_{k\geq 1} A_k\right] &= \lim_{n\to\infty} P\left[\bigcup_{k\leq n} A_k\right] \\ &\leq \liminf_{n\to\infty} \sum_{k=1}^{n} P[A_k] = \sum_{k=1}^{\infty} P[A_k] \end{aligned}$$

Lemma 1.9 (Borel-Cantelli I)**.** *Assume* $(A_k)_{k\geq 1} \subseteq \mathcal{F}$ *with* $\sum_{i\geq 1} P[A_i] < \infty$. *Then*

$$P\left[\bigcap_{n=1}^{\infty} \bigcup_{k\geq n} A_k\right] = 0$$

Remark 1.10. Notice that

$$\begin{aligned} P\left[\bigcap_{n=1}^{\infty} \bigcup_{k\geq n} A_k\right] &= P\left[\{\omega : k_i(\omega) \to \infty \text{ s.t. } \omega \in A_{k_i(\omega)}, i \geq 1\right] \\ &= P[\infty \text{ many of } A_k \text{ occur}] = P[A_n \text{ i.o.}] \end{aligned}$$

Proof. Notice that $\bigcup_{k\geq n} A_k$ is a descending collection in n, so

$$P\left[\bigcap_{n=1}^{\infty} \bigcup_{k\geq n} A_k\right] = \lim_{n\to\infty} P\left[\bigcup_{k\geq n} A_k\right] \leq \liminf_{n\to\infty} \sum_{k\geq n} P[A_k] = 0$$

by assumption. □

Theorem 1.11 (Uniqueness)**.** *Let* P_1, P_2 *be two probability measures on* $(\Omega, \mathcal{F})$, *and suppose* $\mathcal{B} \subseteq \mathcal{F}$ *is* $\cap$*-closed. If* $P_1 \restriction_{\mathcal{B}} \equiv P_2 \restriction_{\mathcal{B}}$, *then* $P_1 \restriction_{\sigma(\mathcal{B})} \equiv P_2 \restriction_{\sigma(\mathcal{B})}$.

Proof. Observe that $\mathcal{D} := \{A \in \mathcal{F} : P_1[A] = P_2[A]\}$ is, indeed, a Dynkin system containing $\mathcal{B}$ (which is $\cap$-closed). Thus, $\mathcal{D} \supseteq \mathcal{D}(\mathcal{B}) = \sigma(\mathcal{B})$, by the π-λ Theorem 1.7. □

Example 1.12. 1. **Discrete models**. Let $\mathcal{F} = \sigma(Z)$ where $Z = (A_i)_{i\in\mathbb{N}}$ is a countable partition of Ω. Then every probability measure is determined by its "weights" on the atoms A_i, since

$$\sigma(Z) = \left\{\bigcup_{i\in J} A_i : J \subseteq \mathbb{N}\right\} \Rightarrow P\left[\bigcup_{i\in J} A_i\right] = \sum_{i\in J} P[A_i]$$

Special case: if Ω is countable and $\mathcal{F}' = \mathcal{P}(\Omega)$ then P is determined by $P[\{\omega\}] = p(\omega)$ the weights on the singletons, since $p(\omega) \geq 0$ and $\sum_{\omega\in\Omega} p(\omega) = 1$.

2. **Dirac measure**. Take $(\Omega, \mathcal{F})$ with $\omega_0 \in \Omega$. Then

$$P[A] = \begin{cases} 1 & \text{if } \omega_0 \in A \\ 0 & \text{if } \omega_0 \notin A \end{cases}$$

is a measure, callewd the "Dirac mass concentrated at ω_0".

3. **Uniform distribution** on $[0, 1] = \Omega$. Take $\mathcal{F}$ to be the Borel σ-field. The Lebsgue measure λ yields a probability distribution on $[0, 1]$ with $\lambda((a, b]) = b - a$ for $b > a$. Note that λ is uniquely determined by the above line since the collection of intervals $\{(a, b] : a \leq b \in [0, 1]\}$ is $\cap$-closed and generates the Borel σ-field.

4. **(Discrete) Stochastic process**. Let S be countable and $\mathfrak{S} = \mathcal{P}(S)$ and $\Omega = S^{\mathbb{N}} = \{(\omega_1, \omega_2, \dots) : \omega_i \in S\}$ and

$$\mathcal{F} = \sigma\left(\{\{X_k \in A\} : k \geq 0, A \in \mathfrak{S}\}\right)$$

A *stochastic process* is any measure P on Ω. It is determined by the values

$$P\left[\{X_0 = s_0, X_1 = s_1, \dots, X_k = s_k\}\right] \qquad , k \geq 0, s_i \in S$$

since they generate $\mathcal{F}$ and are $\cap$-closed. The following are special cases of stochastic processes.

   (a) **Independent experiments** with values in S with distribution μ, and
   $$P\left[\{X_0 = s_0, \dots, X_n = s_n\}\right] = \mu(s_0) \cdot \mu(s_1) \cdots \mu(s_n)$$

   (b) **Markov chain** with initial distribution μ on S and translation kernel $K(S, S')$, and

   $$\begin{aligned} &P\left[\{X_0 = s_0, X_1 = s_1, \dots, X_n = s_n\}\right] \\ &\qquad = \mu(s_0)K(s_0, s_1)K(s_1, s_2) \cdots K(s_{n-1}, s_n) \end{aligned}$$

   The existence of P follows from the Theorem below.

Theorem 1.13 (Carathéodory)**.** *Let $\mathcal{B}$ be an algebra on Ω and P a normalized σ-additive set function on $\mathcal{B}$. Then $\exists!$ extension of P on $\sigma(\mathcal{B}) = \mathcal{F}$.*

Remark 1.14. For a proof, see any standard text on measure theory. The uniqueness follows from the fact that $\mathcal{B}$ is $\cap$-closed.

1.4 Independence

Let $(\Omega, \mathcal{F}, P)$ be given.

Definition 1.15. *A collection $(A_i)_{i\in I}$ of events in $\mathcal{F}$ is called* independent *provided*

$$\forall J \subseteq I, |J| < \infty \Rightarrow P\left[\bigcap_{i\in J} A_i\right] = \prod_{i\in I} P[A_i]$$

Definition 1.16. *A collection of set systems $(\mathcal{B}_i)_{i\in I}$ with $\mathcal{B}_i \subseteq \mathcal{F}$ is called* independent *provided for every choice of $A_i \in \mathcal{B}_i$, the chosen events $(A_i)_{i\in I}$ are independent.*

Theorem 1.17. *Let $(\mathcal{B}_i)_{i\in I}$ be an independent collection of $\cap$-closed set systems in $\mathcal{F}$. Then*

1. *$(\sigma(\mathcal{B}_i))_{i\in I}$ is also independent, and*
2. *if $(J_k)_{k\in K}$ is a partition of I, then $\left(\sigma\left(\bigcup_{i\in J_k} \mathcal{B}_i\right)\right)_{k\in K}$ is also independent.*

Note that (1) is a special case of (2), obtained by setting $K = I$ and $J_k = \{k\}$ for $k \in I$.

Proof. 1. Pick $\{i_1, \ldots, i_n\} =: J \subseteq I$ and $A_j \in \sigma(\mathcal{B}_j)$ for $j \in J$. WWTS

$$P\left[\bigcap_{j\in J} A_j\right] = \prod_j P[A_j] \tag{1}$$

Define

$$\mathcal{D} = \{A \in \sigma(\mathcal{B}_{i_1}) : P[A \cap A_{i_2} \cap \cdots \cap A_{i_n}] = P[A]P[A_{i_2}] \cdots P[A_{i_n}]\}$$

By assumption, $\mathcal{B}_{i_1} \subseteq \mathcal{D}$; also, $\mathcal{D}$ is a Dynkin system because

(a) $\Omega \in \mathcal{D}$

(b) If $A \in \mathcal{D}$ then

$$\begin{aligned} P[A^c \cap A_{i_2} \cap \cdots \cap A_{i_n}] &= P[A_{i_2} \cap \cdots \cap A_{i_n}] - P[A \cap A_{i_2} \cap \cdots \cap A_{i_n}] \\ &= \prod_{k=2}^{n} P[A_{i_k}] - P[A] \prod_{k=2}^{n} P[A_{i_k}] \\ &= (1 - P[A]) \prod_{k=2}^{n} P[A_{i_k}] = P[A^c] \prod_{k=2}^{n} P[A_{i_k}] \end{aligned}$$

so $A^c \in \mathcal{D}$, as well.

(c) Observe that

$$\begin{aligned} P\left[\left(\bigcup_{k\geq 1} A_k\right)\cap A_{i_2}\cap\cdots\cap A_{i_n}\right] &= \sum_{k\geq 1} P\left[A_k\cap A_{i_2}\cap\cdots\cap A_{i_n}\right] \\ &= \sum_{k\geq 1} P[A_k]\cdot\prod_{j=2}^{n} P[A_{i_j}] \\ &= P\left[\bigcup_{k\geq 1}\right] P[A_{i_2}]\cdots P[A_{i_n}] \end{aligned}$$

Now, since $\mathcal{B}_{i_1}$ is $\cap$-closed, $\mathcal{D}\supseteq\sigma(\mathcal{B}_{i_1})$ and so Equation 1 holds for the collection $\sigma(\mathcal{B}_{i_1}),\mathcal{B}_{i_2},\ldots,\mathcal{B}_{i_n}$. Iterating the above arguments, we conclude that $\sigma(\mathcal{B}_{i_1}),\cdots,\sigma(\mathcal{B}_{i_n})$ are also independent, as desired.

2. The set systems

$$\mathcal{C}_k := \left\{\bigcap_{i\in J} A_i : J\subseteq J_k, |J|<\infty, A_i\in\mathcal{B}_i\right\}$$

for $k\in K$ are $\cap$-closed and independent. Thus, for any choices $C_{k_\ell}\in\mathcal{C}_\ell$, we have

$$\begin{aligned} P\left[C_{k_1}\cap\cdots\cap C_{k_n}\right] &= P\left[\left(\bigcap_{i\in J_1} A_i\right)\cap\cdots\cap\left(\bigcap_{i\in J_n} A_i\right)\right] \\ &= \left(\prod_{i\in J_1} P[A_i]\right)\cdots\left(\prod_{i\in J_n} P[A_i]\right) \\ &= P[C_{k_1}]\cdots P[C_{k_n}] \end{aligned}$$

since $J_\ell\subseteq J_{k_\ell}$. Now, by part (1), we know that $(\sigma(\mathcal{C}_k))_{k\in K}$ are independent. Finally, note that $\sigma(\mathcal{C}_k)=\sigma\left(\bigcup_{i\in J_k}\mathcal{B}_i\right)$.

□

Lemma 1.18 (Borel-Cantelli II). *Let $(A_i)_{i\in\mathbb{N}}$ be independent with $\sum_i P[A_i]=\infty$. Then*

$$P\left[\bigcap_{n\geq 0}\bigcup_{k\geq n} A_k\right] = 1$$

Proof. First, notice that the equation above is equivalent to

$$P\left[\bigcup_{k\geq n} A_k\right] = 1\ \forall n \iff P\left[\bigcap_{k\geq n} A_k^c\right] = 0\ \forall n$$

But then,

$$
\begin{aligned}
P\left[\bigcap_{k\geq n} A_k^c\right] &= \lim_{m\to\infty} P\left[\bigcap_{n\leq k\leq m} A_k^c\right] \\
&= \lim_{m\to\infty} \prod_{n\leq k\leq m} (1-P[A_k]) \\
&\leq \liminf_{m\to\infty} \exp\left(-\sum_{k=n}^{m} P[A_k]\right) = 0
\end{aligned}
$$

since $\sum P[A_k] = \infty$. □

Example 1.19. An "application" of this lemma is Shakespeare and the monkey.

Theorem 1.20 (0-1 Law of Kolmogorov)**.** *Let $(\mathcal{F}_i)_{i\geq 1}$ be a countable collection of independent σ-fields. Set*

$$
\mathcal{F}^\star := \bigcap_{n\geq 1} \sigma\left(\bigcup_{k\geq n} \mathcal{F}_k\right)
$$

to be the tail field. Then $\mathcal{F}^\star$ is trivial *in the sense that $P[A]=0$ or $P[A]=1$ for every $A\in\mathcal{F}^\star$.*

Proof. Set

$$
\mathcal{F}_\infty = \sigma\left(\bigcup_{k\geq 1} \mathcal{F}_k\right) \supseteq \mathcal{F}^\star
$$

WWTS $\mathcal{F}_\infty$ and $\mathcal{F}^\star$ are independent. Notice that this completes the proof because $\forall A\in\mathcal{F}^\star$, $A\in\mathcal{F}_\infty$ also, so independence implies

$$
P[A] = P[A\cap A] = (P[A])^2 \Rightarrow P[A]=0 \text{ or } P[A]=1
$$

Now, to prove independence, observe that $\mathcal{F}_1,\mathcal{F}_2,\ldots,\mathcal{F}^\star$ are independent since

$$
\mathcal{F}_1,\mathcal{F}_2,\ldots,\mathcal{F}_n,\sigma\left(\bigcup_{k>n}\mathcal{F}_k\right) \text{ are independent } \forall n
$$

and we know $\mathcal{F}^\star \subseteq \sigma\left(\bigcup_{k>n}\mathcal{F}_k\right)$. Next, $\mathcal{F}^\star,\mathcal{F}_1,\mathcal{F}_2,\ldots$ being independent implies that $\mathcal{F}^\star$ and $\sigma\left(\bigcup_{k\geq 1}\mathcal{F}_k\right) = \mathcal{F}_\infty$ are independent, as well, and we're done. Note we have used Theorem 1.17 twice. □

Example 1.21. Independent Bernoulli variables with parameter p. Take $\Omega = \{0,1\}^{\mathbb{N}^+}$ and $X_k(\omega)=\omega_k$ with the σ-fields

$$
\mathcal{F}_k = \sigma\left(\{X_k=1\}\right) \qquad \text{and} \qquad \mathcal{F} = \sigma\left(\bigcup_k \mathcal{F}_k\right)
$$

The probability P is determined by

$$P = P_p\left[X_{k_1} = 1, \dots, X_{k_j} = 1, X_{i_1} = 0, \dots, X_{i_\ell} = 0\right] = p^j(1-p)^\ell$$

Let $c \in [0,1]$ and set

$$A_c := \left\{\omega : \limsup_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} X_k(\omega) = c\right\}$$

Then $A_c \in \mathcal{F}^\star$, so by Kolmogorov's Law 1.20, $P[A_c] = 0$ or $P[A_c] = 1$. Let j be fixed; then

$$A_c = \left\{\limsup_{n\to\infty} \frac{1}{n-j+1}\sum_{k=j}^{n} X_k = c\right\} \in \sigma\left(\bigcup_{k\geq j} \mathcal{F}_k\right) = \mathcal{F}_j^\star$$

and so $A_c \in \mathcal{F}_j^\star\ \forall j$.

Example 1.22. Percolation with parameter p. Take $\Omega = \{0,1\}^{\mathbb{Z}^d}$ and $X_z(\omega) = \omega_z$ for $z \in \mathbb{Z}^d$, with

$$\mathcal{F}_k := \sigma\left(\{X_z = 1\} : \|z\| = k\right)$$

for $k \geq 0$, where $\|z\| = \max_{1\leq i\leq d} |z_i|$. If $X_z = 1$ then z is *open*, and closed otherwise. Set

$$P_p\left[X_{z_1} = \dots = X_{z_k} = 1, X_{y_1} = \cdots X_{y_\ell} = 0\right] = p^k(1-p)^\ell$$

so it uniquely determines a probability. We claim $(\mathcal{F}_k)_{k\geq 0}$ is an *independent* family; in fact, this follows from Theorem 1.17 part (2).

A set C of sites in $\mathbb{Z}^d$ is called *connected* if between any two sites in C $\exists$ a sequence of nearest neighbors $\subseteq C$. An *(open) cluster* is a connected component of open sites. Set $A = \{\exists \infty \text{ cluster}\}$. We claim

$$A \in \mathcal{F}^\star = \bigcap_{n\geq 0} \sigma\left(\bigcup_{k\geq n} \mathcal{F}_k\right)$$

To see why, let n be fixed and set $B(k) = \{z : \|z\| < k\}$. The basic observation is that $\forall n$, $\omega \in A \iff \omega \in \hat{A}_n$ where

$$\hat{A}_n := \{\omega : \exists \infty \text{ cluster in } B^c(n)\}$$

which implies that $A = \hat{A}_n\ \forall n$. But notice that

$$\hat{A}_n \in \sigma\left(\bigcup_{k<n} \mathcal{F}_k\right) \Rightarrow A \in \bigcap_n \sigma\left(\bigcup_{k>n} \mathcal{F}_k\right) = \mathcal{F}^\star$$

By Kolmogorov's Law 1.20, either $P_p[A] = 0$ or $P_p[A] = 1$. In fact, if we set $p_c := \inf\{p \geq 0 : P_p[A] > 0\}$, then

$$p > p_c \Rightarrow P_p[A] > 0 \Rightarrow P_p[A] = 1$$

A further fact is that $0 < p_c < 1$ for $d \geq 2$.

1.5 Measurable Maps and Induced Measures

Let $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ be measurable spaces, and let $T : \Omega \to \Omega'$ be any map. For $A' \in \mathcal{F}'$, we write

$$\{T \in A'\} = \{\omega : T(\omega) \in A'\} = T^{-1}(A')$$

Definition 1.23. *The collection*

$$\sigma(T) = \{\{T \in A'\} : A' \in \mathcal{F}'\}$$

is, indeed, a σ-field on Ω, and it is called the σ-field generated *by T.*

Definition 1.24. *The map T is called* measurable *with respect to $(\mathcal{F}, \mathcal{F}')$ if*

$$\{T \in A'\} \in \mathcal{F} \; \forall A' \in \mathcal{F}' \tag{2}$$

Remark 1.25.

1. It is sufficient to check the condition in Equation (2) for a generator $\mathcal{B}'$ with $\sigma(\mathcal{B}') = \mathcal{F}'$, since the collection

   $$\{A' \subseteq \Omega' : \{T \in A'\} \in \mathcal{F}\}$$

   is, indeed, a σ-field (*** proven on homework) and it contains $\mathcal{B}'$ so it must also contain $\sigma(\mathcal{B}')$.

2. The composition of measurable maps is also measurable. That is, if we're given the measurable spaces $(\Omega, \mathcal{F}), (\Omega', \mathcal{F}'), (\Omega'', \mathcal{F}'')$ and the measurable maps $T : \Omega \to \Omega'$ and $S : \Omega' \to \Omega''$, then $S \circ T : \Omega \to \Omega''$ is measurable, as well.

3. If (Ω, τ) and (Ω', τ') are topological spaces and $T : \Omega \to \Omega'$ is continuous, then T is measurable with respect to the Borel σ-fields $\sigma(\tau)$ and $\sigma(\tau')$.

Remark 1.26. We use $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ where $\mathcal{B}_{\mathbb{R}}$ is the σ-field generated by all open subsets of $\mathbb{R}$, which is equivalent to the σ-field generated by (open) intervals (*** proven on homework).

Also, we use the topological space of the extended reals $\overline{\mathbb{R}} = \{-\infty\} \cup \{\infty\} \cup \mathbb{R}$ with open sets generated by the neighborhood bases

$$\left\{\mathcal{N}\left(r, \frac{1}{n}\right) : r \in \overline{\mathbb{R}}, n \geq 1\right\}$$

where

$$\mathcal{N}\left(\infty, \frac{1}{k}\right) = \left\{x \in \overline{\mathbb{R}} : x > k\right\} = (k, \infty]$$

and

$$\mathcal{N}\left(-\infty, \frac{1}{k}\right) = \left\{x \in \overline{\mathbb{R}} : x < -k\right\} = [-\infty, -k)$$

Open sets are unions of neighborhood basis elements. We can discuss *convergence* of sequences by saying $x_n \to x$ as $n \to \infty$ provided $\forall k \, \exists n$ such that $x_m \in \mathcal{N}(x, \frac{1}{k}) \; \forall m \geq n$. Note: we will use $(\mathbb{R}, \mathcal{B})$ to indicated $(\overline{\mathbb{R}}, \overline{\mathcal{B}})$, sometimes!

Definition 1.27. *Let $(\Omega, \mathcal{F})$ be a measurable space. The map $T : \Omega \to \mathbb{R}$ or $T : \Omega \to \overline{\mathbb{R}}$ is called a* random variable *if it is $(\mathcal{F}, \mathcal{B})$ measurable. General measurable maps are called* abstract valued *random variables.*

Example 1.28. 1. Let $(\Omega, \mathcal{F})$ be a measurable space with $\mathcal{F} = \sigma(Z)$ where Z is a countable partition of Ω with atoms A_i. Let $T : \Omega \to \mathbb{R}$. Then T is a random variable $\iff$ T is constant on every atom.

2. **Tossing a coin**. We begin by tossing a (fair) coin. Let $\Omega = \{0,1\}^{\mathbb{N}^+}$ and $X_n(\omega) = \omega_n$ with

$$\mathcal{F}_n = \sigma(X_n) \quad \text{and} \quad \mathcal{F} = \sigma\left(\bigcup_{k\geq 1} \mathcal{F}_k\right)$$

Set

$$T(\omega) := \sum_{k\geq 1} X_k(\omega) 2^{-k}$$

Note that $T : \Omega \to [0,1]$ where $[0,1]$ is equipped with the Borel-field

$$\mathcal{B} := \sigma\left(\{[0,c) : 0 < c \leq 1\}\right)$$

We claim T is $(\mathcal{F}, \mathcal{B})$ measurable. To see why, we first recall some facts about the dyadic representations of numbers. Let

$$\Omega_0 = \{\omega : X_n(\omega) = 1 \text{ i.o.}\}$$

Then $\forall c \in (0,1]$ $\exists! \overline{c} = (\overline{c}_1, \overline{c}_2, \overline{c}_3, \dots) \in \Omega_0$ such that

$$c = \sum_{k\geq 1} \overline{c}_k 2^{-k} = T(\overline{c})$$

That is, we always choose the dyadic representation that uses infinitely many 1s. Notice that if $d < c$ then $\exists n_0 \geq 1$ such that $\overline{d}_1 = \overline{c}_1, \dots, \overline{d}_n = \overline{c}_n$ but $\overline{d}_{n+1} = 0$ whereas $\overline{c}_{n+1} = 1$. Therefore, for $c \in (0,1]$,

$$\{T < c\} = \bigcup_{n:\overline{c}_n = 1} \left\{ \bigcap_{k<n} \{X_k = \overline{c}_k\} \cap \{X_n = 0\} \right\} \in \sigma\left(\bigcup_n \mathcal{F}_n\right) = \mathcal{F}$$

where n in the first intersection above is the first index such that the digits of $T(\omega)$ and c differ. This implies that $T^{-1}[0,c) \in \mathcal{F}$ for every $c > 0$ and since the sets $[0,c)$ generate $\mathcal{B}$, we may conclude T is $(\mathcal{F}, \mathcal{B})$ measurable.

Lemma 1.29. *Let $(\Omega, \mathcal{F}), (\Omega', \mathcal{F}')$ be measurable spaces. If X is $\sigma(\phi)$ measurable, then $\exists \varphi$ such that $X = \varphi \circ \phi$. This is known as "lifting".*

***** insert picture *****

Definition 1.30. *Let $T : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ be measurable, and let P be a probability measure on $(\Omega, \mathcal{F})$. Then*

$$P'[A'] := P[T^{-1}A'] = P \circ T^{-1}(A')$$

is, indeed, a probability measure and it is called the induced measure *a.k.a. the* image measure *of P under T a.k.a. the* distribution *of T under P.*

Example 1.31. **Bernoulli variable** X with parameter p. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $X : \Omega \to \{0, 1\}$ such that $P[X = 1] = p$ and $P[X = 0] = 1 - p$. Then $P' = P \circ T^{-1}$ is determined by

$$P'[\{0\}] = 1 - p \quad , \quad P'[\{1\}] = p \quad , \quad P'[\emptyset] = 0 \quad , \quad P'[\Omega] = 1$$

Example 1.32. **Tossing a coin**. Let $\Omega = \{0, 1\}^{\mathbb{N}^+}$. Put a measure on $(\Omega, \mathcal{F})$ by setting

$$P\left[X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\right] = 2^{-n} \ \forall n, \forall x_1, \ldots, x_n \in \{0, 1\}$$

Then P is uniquely determined. (We will see later how to construct such a P.) What is the image measure on $(0, 1]$ under T? Observe that

$$\begin{aligned} P'\left[[0, c)\right] &= P \circ T^{-1}[0, c) = P[T < c] \\ &= P\left[\bigcup_{n:\bar{c}_n = 1} \{X_1 = \bar{c}_1, \ldots, X_{n-1} = \bar{c}_{n-1}, X_n = 0\}\right] \\ &= \sum_{n:\bar{c}_n = 1} P\left[\{X_1 = \bar{c}_1, \ldots, X_{n-1} = \bar{c}_{n-1}, X_n = 0\}\right] \\ &= \sum_n \bar{c}_n 2^{-n} = T(\bar{c}) = c \end{aligned}$$

which implies that P' is equivalent to the Lebsgue measure on $(0, 1]$! This is equivalent to saying T is uniformly distributed with respect to P.

Remark 1.33. The existence of 0-1 variables $\Rightarrow$ the existence of the Lebesgue measure, and vice versa.

Example 1.34. **Contraction and Simulation of Probability Distributions** on $\mathbb{R}$ (or $\bar{\mathbb{R}}$). Let λ be a uniform distribution (i.e. probability measure) with respect to the Borel σ-field $\mathcal{F}$ on $[0, 1]$. If μ is a probability measure on $\mathbb{R}$, then $F_\mu(x) = \mu(-\infty, x]$ is called the (cumulative) *distribution function* of μ. Note: μ is uniquely determined by F_μ! Also, F_μ has the following properties

1. $F_\mu : \mathbb{R} \to [0, 1]$ with

$$\lim_{x \to -\infty} F_\mu = 0 \quad , \quad \lim_{x \to \infty} F_\mu = 1$$

2. $F_\mu \nearrow$

3. F_μ is right continuous

These properties follow from basic properties of probability measures.

Suppose we have F that satisfies properties 1,2,3. We now show $F \equiv F_\mu$ for some $\mu \in \mathfrak{M}_1(\mathbb{R})$; specifically, given F we will try to construct μ with $F \equiv F_\mu$. Set

$$G(y) := \inf\{c : F(c) > y\}$$

to be the *unique* right continuous inverse of F. (Proof: ***) It is true that

$$\{G \le c\} = \begin{cases} [0, F(c)) & \text{if } F \text{ is "constant after } c\text{"} \\ [0, F(c)] & \text{otherwise} \end{cases}$$

so G is measurable from $[0,1] \to \bar{R}$. Define $\mu = \lambda \circ G^{-1}$ (the distribution of G with respect to λ). Then

$$\mu([-\infty, c]) = \lambda(\{G \le c\}) = F(c) \Rightarrow F = F_\mu$$

since

$$\mu(\{-\infty\}) = \lim_{n\to\infty} \lambda(G \le n) = \lim_{n\to\infty} F(-n) = 0 = \mu(\{+\infty\})$$

and thus $\mu((-\infty,\infty)) = 1$ which implies $\mu \restriction_{\mathbb{R}}$ *is* a probability measure on $\mathbb{R}$ with $F_\mu = F$.

Lemma 1.35. *Let* $: (\Omega, \mathcal{F}) \to \bar{R}$ *be some measurable map. Let* P *be a probability measure on* $(\Omega, \mathcal{F})$ *such that* P *is 0-1 on* $\mathcal{F}$*. Then* T *is* P*-a.s. constant; i.e.* $\exists a$ *such that* $P[T = a] = 1$.

Proof. (*** homework exercise ***) □

1.6 Random Variables and Expectation

Consider a measurable space $(\Omega, \mathcal{F})$, and $\bar{\mathbb{R}} = [-\infty, \infty]$ with Borel σ-algebra $\mathcal{B} \equiv \sigma([-\infty, c) : c \in \mathbb{R})$.

Definition 1.36. *We say* $X : \Omega \to \bar{\mathbb{R}}$ *is a* random variable *if it is* $(\mathcal{F}, \mathcal{B})$ *measurable.*

Remark 1.37. It suffices to check that

$$\{X < c\} \in \mathcal{F} \; \forall c \in \mathbb{R}$$

since the intervals $[-\infty, c)$ generate $\mathcal{B}$.

If X, Y are RVs, then

$$\{X < Y\} = \bigcup_{r \in \mathbb{R}} \{X < r\} \cap \{Y > r\} \in \mathcal{F}$$

and

$$\{X = Y\} = (\{X = Y\}^c)^c = \{\{X > Y\} \cup \{X < Y\}\}^c \in \mathcal{F}$$

and so forth.

If $X_1, \dots, X_N$ are RVs and $f : \bar{\mathbb{R}}^n \to \bar{\mathbb{R}}$ is measurable, then

$$Y = f(X_1, \dots, X_n)$$

is also a RV. Thus,

$$\sum_{i=1}^{n} X_i \ , \ \prod_{i=1}^{n} X_i \ , \ \max_{1 \le i \le n} X_i \ , \ X_1^+$$

are all RVs, as well.

The class of RVs is closed under "countable operations". That is,

$$X_1 \le X_2 \le \cdots \le X_n \le \cdots \ \Rightarrow \ Y = \lim_{n \to \infty} X_n \text{ is a RV}$$

To see why, notice that

$$\left\{ \lim_{n \to \infty} X_n \le c \right\} = \bigcap_{n \ge 1} \{X_n \le c\} \in \mathcal{F}$$

Also, $Y = \sup_n X_n$ is a RV because

$$\sup_n X_n = \lim_n \nearrow Y_n \quad \text{ where } Y_n = \max_{1 \le i \le n} X_i$$

Similarly, $Y = \liminf_n X_n$ is a RV (and lim sup) because

$$\liminf_{n \to \infty} X_n = \lim_{n \to \infty} \nearrow Y_n \quad \text{ where } Y_n = \inf_{k \ge n} X_k$$

which implies

$$\left\{ \lim_{k \to \infty} X_k \text{ exists} \right\} = \left\{ \liminf_{n \to \infty} X_n = \limsup_{n \to \infty} X_n \right\}$$

Example 1.38.

- If $A \in \mathcal{F}$, then $\mathbf{1}_A$ is a RV.
- If $A_1, \dots, A_n \in \mathcal{F}$ and $c_i \in \mathbb{R}$ (not $\bar{\mathbb{R}}$), then

  $$X := \sum_{i=1}^{n} c_i \mathbf{1}_{A_i} \in \mathcal{F}$$

  is called a *step function.* Note that WOLOG the A_is can be taken to be *disjoint* (since the sum is over a finite index set).

Lemma 1.39. *If $X \ge 0$ is a RV, then $\exists X_n$ a monotone increasing sequence of step functions such that $X_n \nearrow X$ as $n \to \infty$.*

Proof. Define X_n by

$$X_n := \left(\sum_{k=0}^{n^2-1} \frac{k}{n} \mathbf{1}_{\left\{ \frac{k}{n} \le X < \frac{k+1}{n} \right\}} \right)^+ + n \mathbf{1}_{\{X \ge n\}}$$

The idea is that as $n \to \infty$, we generate a finer mesh on the interval $[0, n]$. $\square$

Note: this lemma is *crucial* to be able to define integration!

Theorem 1.40 (Lifting)**.** *Let $T : \Omega \to \Omega'$ be $(\mathcal{F}, \mathcal{F}')$ measurable. If X is $\sigma(T)$ measurable (i.e. "$X \in \sigma(T)$"), then $\exists \varphi$ measurable such that $X = \varphi \circ T$.*

****** insert diagram ******

Proof. This proof technique is (sometimes) known as "measure theoretic induction" or just MTI, for short.

1. Let $X = \mathbf{1}_A$ for
$$A \in \sigma(T) = \left\{T^{-1}(B) : B \in \mathcal{F}'\right\}$$
so $\exists B \in \mathcal{F}'$ such that $A = T^{-1}(B)$. Thus,
$$X = \mathbf{1}_A = \mathbf{1}_B \circ T \Rightarrow \varphi = \mathbf{1}_B$$
since
$$\mathbf{1}_B = \begin{cases} 1 & \text{if } T \in B \\ 0 & \text{if } T \notin B \end{cases}$$

2. Let
$$X = \sum_{i=1}^{n} c_i \mathbf{1}_{A_i} \qquad \text{where } A_i = T^{-1}(B_i)\ , i = 1, \ldots, n$$
where $B_i \in \mathcal{F}'$ are *disjoint*, so the A_i are, as well. Then
$$X = \left(\sum_{i=1}^{n} c_i \mathbf{1}_{A_i}\right) \circ T = \begin{cases} c_i & \text{on } T \in B_i \\ 0 & \text{otherwise} \end{cases}$$
and $T \in B_i \equiv A_i$. Thus, we can say
$$\varphi := \sum_{i=1}^{n} c_i \mathbf{1}_{A_i}$$

3. If $X \geq 0$ then $\exists X_n \nearrow X$ with X_n step functions. But $X_n = \varphi_n \circ T$ by (2), so
$$X = \lim_n \nearrow (\varphi_n \circ T) = \left(\lim_{n\to\infty} \nearrow \varphi_n\right) \circ T$$
and we can set $\varphi := \lim_{n\to\infty} \varphi_n$.

4. If $X = X^+ - X^-$, then $X^+, X^- \geq 0$ are $\sigma(T)$ measurable, so then
$$X = \varphi^+ \circ T - \varphi^- \circ T = \left(\varphi^+ - \varphi^-\right) \circ T$$
so we can set $\varphi := \varphi^+ - \varphi^-$. This completes the proof!

□

Remark 1.41. **Special case**: Suppose X, Y are RVs and X is $\sigma(Y)$ measurable. Then $\exists \varphi : \mathbb{R} \to \mathbb{R}$ (or $\varphi : \bar{\mathbb{R}} \to \bar{\mathbb{R}}$) measurable such that $X = \varphi(Y)$! (i.e. X depends *deterministically* on Y)

1.6.1 Integral (expected value)

Definition 1.42. *Let X be a RV on $(\Omega, \mathcal{F}, P)$. We write*

$$E[X] := \int_\Omega X\, dP \equiv \int_\Omega X(\omega) P(d\omega)$$

(whenever the integral exists) and

$$E[X; A] := \int_A X\, dP = \int_\Omega X \cdot \mathbf{1}_A\, dP$$

Sketch of the construction:

1. If $X = \mathbf{1}_A$ then $E[X] := P[A]$.
2. If $X = \sum_{i=1}^n c_i \mathbf{1}_{A_i}$ with A_is disjoint then $E[X] := \sum_i c_i P[A_i]$.
3. If $X \geq 0$ then find $X_n \nearrow X$ step functions and set

   $$E[X] := \lim_{n\to\infty} \nearrow E[X_n]$$

   (note: this is $\leq \infty$!)
4. If $X = X^+ - X^-$ then $E[X] := E[X^+] - E[X^-]$. Note: the RHS exists except when $E[X^+] = E[X^-] = \infty$.

Definition 1.43. *If $E[X] \in [-\infty, \infty]$ exists, we say X is* semi-integrable. *If $E[X]$ is finite, we say X is* integrable. *We let $\mathfrak{S}(\Omega, \mathcal{F}, P)$ denote the class of semi-integrable functions, and $\mathcal{L}^1(\Omega, \mathcal{F}, P)$ denote the class of integrable functions.*

In measure theory, one verifies the following properties of the integral (i.e. of $E[\cdot]$):

1. **Linearity**: $E[X + cY] = E[X] + cE[Y]$
2. **Monotonicity**: $X \leq Y$ a.s. $\Rightarrow$ $E[X] \leq E[Y]$
3. **Monotone convergence**: If $X_0 \in \mathcal{L}^1$ and $X_0 \leq X_1 \leq \cdots$ a.s., then

   $$E\left[\lim_{n\to\infty} \nearrow X_n\right] = \lim_{n\to\infty} \nearrow E[X_n]$$

   This is a Theorem due to Beppo-Levi.

These 3 properties are the basic ones; all others follow from these!

Remark 1.44. Let $P[A] = 0$ and $X \in \bar{\mathbb{R}}$ measurable. Then $E[\mathbf{1}_A \cdot X] = 0$. To see why, assume $X \geq 0$ and observe that

$$X \cdot \mathbf{1}_A = \lim_{n\to\infty} \nearrow (X \wedge n) \cdot \mathbf{1}_A$$

and so

$$E[X \cdot \mathbf{1}_A] \le E[n \cdot \mathbf{1}_A] = n \cdot 0 = 0$$

Then, by Beppo-Levi, $E[X \cdot \mathbf{1}_A] = \lim 0 = 0$. For general X, notice that

$$E[X^+ \cdot \mathbf{1}_A] - E[X^- \cdot \mathbf{1}_A] = 0 - 0 = 0$$

Also, note that in monotone convergence, the assumption that $X_0 \in \mathcal{L}^1$ is *necessary.* As a counterexample, consider $\Omega](0,1]$ and $X_0 = f(x) = -\frac{1}{x}$. Set

$$X_n = X_0 \cdot \mathbf{1}_{(0,\frac{1}{n}]}$$

Notice that $X_n \nearrow 0$, but

$$-\infty = E[X_n] \not\to E\left[\lim_{n\to\infty} X_n\right] = 0$$

Theorem 1.45 (Fatou's Lemma)**.** *Suppose* $(X_n)_{n\ge 1} \ge Y$ *a.s. and* $Y \in \mathcal{L}^1$. *Then*

$$-\infty < E\left[\liminf_{n\to\infty} X_n\right] \le \liminf_{n\to\infty} E[X_n] \le +\infty$$

Proof. Observe that

$$X_n \ge \inf_{k\ge n} X_n \ge Y \text{ a.s.}$$

so we can apply monotone convergence (since $Y \in \mathcal{L}^1$) to write

$$E\left[\lim_{n\to\infty} \nearrow \left(\inf_{k\ge n} X_n\right)\right] = \lim_{n\to\infty} \nearrow E\left[\inf_{k\ge n} X_k\right] \le \liminf_{n\to\infty} E[X_n]$$

□

By taking minus signs in the proof above, we can show that $(X_n) \le Y \in \mathcal{L}^1$ a.s. implies

$$E\left[\limsup_{n\to\infty} X_n\right] \ge \limsup_{n\to\infty} E[X_n]$$

A (silly but useful) mnemonic to remember the direction of the inequality in the statement of Fatou's Lemma above is ILLLI ("the Integral of the Limit is Less than the Limit of the Integrals").

Theorem 1.46 (Dominated Convergence)**.** *Assume* $X_n \to X$ *a.s. and* $\exists Y \in \mathcal{L}^1$ *such that* $|X_n| \le Y$ *a.s.* $\forall n$. *Then*

1. $E\left[\lim_{n\to\infty} X_n\right] = \lim_{n\to\infty} E[X_n]$, *and*

2. $X_n \to X$ *in* $\mathcal{L}^1$*; i.e.*

$$E\left[|X - X_n|\right] =: \|X - X_n\|_1 \xrightarrow[n\to\infty]{} 0$$

Proof. First, notice that

$$|X_n| \to |X| \text{ a.s. } \Rightarrow |X| \leq Y \text{ a.s. } \Rightarrow E[|X|] \leq E[Y] < \infty \Rightarrow X \in \mathcal{L}^1$$

Now, to prove (1), we apply Fatou's Lemma 1.45 twice to write

$$\begin{aligned} E[X] = E\left[\liminf_{n\to\infty} X_n\right] &\leq \liminf_{n\to\infty} E[X_n] \\ &=\leq \limsup_{n\to\infty} E[X_n] \leq E\left[\limsup_{n\to\infty} X_n\right] = E[X] \end{aligned}$$

so everything is equal in the line above.

To prove (2), define $D_n := X - X_n$, so that

$$|D_n| \leq |X| + |X_n| \leq 2Y \in \mathcal{L}^1$$

so that $|D_n| \leq 2Y$ and $|D_n| \to 0$ a.s. Thus, we can apply the conclusion of part (1) to write

$$0 = E\left[\lim_{n\to\infty} |D_n|\right] = \lim_{n\to\infty} E[D_n] = \lim_{n\to\infty} E[|X - X_n|]$$

□

Theorem 1.47 (Chebyshev-Markov Inequality)**.** *Let $\varphi : \bar{\mathbb{R}} \to \bar{\mathbb{R}}$ with $\varphi \geq 0$ and let A be a Borel set. Define $c_A := \inf_A \varphi$. Then for any RV X,*

$$c_A P[X \in A] \leq E\left[\varphi(X); X \in A\right] \leq E\left[\varphi(X)\right]$$

Proof. Note that $c_A \mathbf{1}_A \leq \varphi \mathbf{1}_A$ and so

$$c_A \mathbf{1}_{\{X\in A\}} = c_A \mathbf{1}_A \circ X \leq \varphi \mathbf{1}_A \circ X = \varphi(X) \mathbf{1}_{\{X\in A\}}$$

Taking $E[\cdot]$ of both sides yields

$$c_A P[X \in A] \leq E\left[\varphi(X); X \in A\right] \leq E\left[\varphi(X)\right]$$

since $\varphi \geq 0$. □

As an application of this inequality, we show that for $X \geq 0$,

$$E[X] < \infty \Rightarrow X < \infty \text{ a.s.}$$

and

$$E[X] = 0 \Rightarrow X = 0 \text{ a.s.}$$

Observe that, for the first case,

$$\begin{aligned} P[X = \infty] = P\left[\bigcap_{n=1}^{\infty} \{X \geq n\}\right] &= \lim_{n\to\infty} P[X \geq n] \\ &\leq \liminf_{n\to\infty} \frac{1}{n} E\left[\varphi(X)\right] = \liminf_{n\to\infty} \frac{1}{n} E[X] = 0 \end{aligned}$$

where we have applied Chebyshev-Markov with $A = [n, \infty]$ and $\varphi = \mathbf{1}_{[0,\infty]} \cdot \mathrm{id}$. For the second case, we use a similar technique to write

$$P[X > 0] = P\left[\bigcup_{n=1}^{\infty}\left\{X \geq \frac{1}{n}\right\}\right] = \lim_{n\to\infty} \nearrow P\left[X \geq \frac{1}{n}\right]$$
$$\leq \liminf_{n\to\infty} nE[X] = \lim 0 = 0$$

Theorem 1.48. *Suppose $X \in \mathcal{L}^1$ and u is convex. Then*

$$E[u(X)] \geq u\left(E[X]\right)$$

Furthermore, if u is strictly convex *then the inequality above is* strict *and X is* not *a.s. constant.*

Note: this theorem only holds for *probability* measures!

Proof. **** insert diagram ***** If u is convex, then $\forall x \in \mathbb{R}$ there is "support line" $\ell(x) = ax + b$ such that $u(y) \geq ay + b$ for every y and $\ell(x) = u(x)$ (note: ℓ is not unique). Pick $x_0 = E[X] < \infty$. Then

$$E[u(X)] \geq E[\ell(X)] = \ell(E[X]) = u(E[X])$$

where the first equality holds because P is a probability measure, so

$$E[aX + b] = aE[X] + E[b] = aE[x] + b$$

The proof of (2) is left as an exercise (***). □

As an application of Jensen's Inequality, consider the space

$$\mathcal{L}^p = \{X : E\left[|X|^p\right] < \infty\}$$

Then for $1 \leq p \leq q < \infty$, we have $\|X\|_p \leq \|X\|_q$; thus, in particular, $\mathcal{L}^q \subseteq \mathcal{L}^p$. First, we show $\mathcal{L}^q \subseteq \mathcal{L}^p$ directly:

$$E\left[|X|^p\right] \leq E\left[|X|^p \vee 1\right] \leq E\left[|X|^q \vee 1\right] \leq E\left[X|^q\right] + 1 < \infty$$

Next, define $\varphi(x) := |X|^{q/p}$ (which is, indeed, convex). Jensen's Ineqaulity tells us

$$E\left[(|X|^p)^{q/p}\right] \geq E\left[|X|^p\right]^{q/p}$$

where $|X|^p \in \mathcal{L}^1$ since $X \in \mathcal{L}^p$.s This implies

$$\|X\|_q = E\left[|X|^q\right]^{1/q} \geq E\left[|X|^p\right]^{1/p} = \|X\|_p$$

Theorem 1.49 (Transformation Formula)**.** *Let $T : (\Omega, \mathcal{F}, P) \to (\Omega', \mathcal{F}', P')$ be measurable and take P' to be the induced measure; i.e. $P' = P \circ T^{-1}$. Let $X' \geq 0$ be a RV on Ω'. Then*

$$E_{P'}[X'] = E_p[X' \circ T]$$

Note that if $X' \in \mathcal{L}^1(P')$ then the same is true, of course.

Proof. We use Measure Theoretic Induction:

1. If $X' = \mathbf{1}_A$ then

$$E_{P'}\left[\mathbf{1}_{A'}\right] = P'(A') = P \circ T^{-1}(A) = P[T \in A] = E_P\left[\mathbf{1}_A \circ T\right]$$

2. By linearity of $E[\cdot]$, step functions also work.

3. If $X' = \lim \nearrow X'_n$ for X'_n step functions, then

$$X' \circ T = \lim_{n\to\infty} \nearrow (X'_n \circ T)$$

which implies

$$E_P\left[X' \circ T\right] = \lim_{n\to\infty} E_P\left[X'_n \circ T\right] = \lim_{n\to\infty} E_{P'}\left[X'_n\right] = E_{P'}\left[X'\right]$$

and this completes the proof! □

Corollary 1.50. *Let X be a RV with $P[X \in \mathbb{R}] = 1$. Then the distribution $\mu = P \circ X^{-1}$ is concentrated on $\mathbb{R}$ and for each measurable function $\varphi \geq 0$, we have*

$$E\left[\varphi(X)\right] = \int_{\mathbb{R}} \varphi(x)\, \mu(dx)$$

Additionally, we note that if $\varphi : \mathbb{R} \to \mathbb{R}$ and either $E[\varphi^+(X)]$ or $E[\varphi^-(X)]$ is $< \infty$, then the same equality above holds.

We note a couple of special cases:

1. We have

$$E[X^+] = \int_0^\infty x\, \mu(dx) \quad \text{and} \quad E[X^-] = \int_{-\infty}^0 |x|\, \mu(dx)$$

and so

$$E[X] = \int_{-\infty}^\infty x\, \mu(dx)$$

whenever $E[X^+]$ or $E[X^-]$ is $< \infty$. Also, we consider the so-called "k-th moment" defined by

$$E\left[|X|^k\right] = \int_{\mathbb{R}} |x|^k\, \mu(dx)$$

and the *variance* of X (assuming $X \in \mathcal{L}^2$) defined by

$$\begin{aligned}\mathrm{Var}(X) &= E\left[(X - E[X])^2\right] = E[X^2] - 2E[X]^2 + E[X]^2 = E[X^2] - E[X]^2 \\ &= \int x^2\, \mu(dx) - \left(\int x\, \mu(dx)\right)^2\end{aligned}$$

2. Let $(X, Y) : (\Omega, T, P) \to \mathbb{R}^2$ and $\mu(dx\, dy)$ be the induced measure on $\mathbb{R}^2$, a.k.a. the product distribution of X, Y. Then we consider the *covariance* of X and Y defined by

$$\begin{aligned}\operatorname{Cov}(X,Y) &= E\left[(X - E[X])(Y - E[Y])\right] = E[XY] - E[X]E[Y] \\ &= \int_{\mathbb{R}^2} xy\, \mu_{xy}(dx\, dy) - \int_{\mathbb{R}} x\, \mu_x(dx) \cdot \int_{\mathbb{R}} y\, \mu_y(dy)\end{aligned}$$

Note: $\operatorname{Cov}(X, X) = \operatorname{Var}(X)$.

1.6.2 Convergence of RVs

Assume $(\Omega, \mathcal{F}, P)$ is a probability space and $(X_n)_{n\geq 1}, X$ are all $\mathbb{R}$-valued RVs.

Definition 1.51. *We say*

1. $X_n \to X$ *a.s. provided*

$$P[X_n \not\to X] = 0$$

2. $X_n \to X$ *in probability (a.k.a. in measure) provided*

$$\lim_{n\to\infty} P\left[|X_n - X| > \varepsilon\right] = 0 \quad \forall \varepsilon > 0$$

3. $X_n \to X$ *in* $\mathcal{L}^1$ *provided*

$$\lim_{n\to\infty} E\left[|X_n - X|\right] = 0$$

The following theorem characterizes these 3 types of convergence.

Theorem 1.52. *1. Almost sure convergence $\Rightarrow$ convergence in probability.*

2. $\mathcal{L}^1$ convergence $\Rightarrow$ convergence in probability.
In general, there are no other implications!

Proof. 1. Suppose $X_n \to X$ a.s. Then

$$\{X_n \not\to X\} = \bigcup_{\ell} \bigcap_{n} \bigcup_{m\geq n} \left\{|X_m - X| \geq \frac{1}{\ell}\right\}$$

is a measure zero set, and so

$$P\left[\bigcap_{n} \bigcup_{m\geq n} \left\{|X_m - X| \geq \frac{1}{\ell}\right\}\right] = 0 \quad \forall \ell$$

Thus,

$$0 = \lim_{n\to\infty} \searrow P\left[\bigcup_{m\geq n} \left\{|X_m - X| \geq \frac{1}{\ell}\right\}\right] \quad \forall \ell$$

and since

$$\bigcup_{m\geq n}\left\{|X_m - X| \geq \frac{1}{\ell}\right\} \supseteq \left\{|X_n - X| \geq \frac{1}{\ell}\right\} \quad \forall \ell$$

we can conclude that

$$\lim_{n\to\infty} P\left[|X_n - X| \geq \frac{1}{\ell}\right] = 0 \quad \forall \ell$$

which is precisely convergence in probability.

2. Assume convergence in $\mathcal{L}^1$. Then we can apply Chebyshev's Inequality 1.47 to conclude

$$P\left[|X_n - X| \geq \varepsilon\right] \leq \frac{1}{\varepsilon} E\left[|X_n - X|\right] \xrightarrow[n\to\infty]{} 0 \quad \forall \varepsilon > 0$$

□

Now, if we add extra assumptions, then the previous theorem becomes more complicated and admits some implications between modes of convergence, as summarized in the following diagram

***** insert diagram (unit 5+ page 3) ******

These implications will be stated and proven in the following series of lemmas and theorems.

Lemma 1.53. *Suppose* $\sum_{n\geq 1} E[|X_n - X|] < \infty$*. Then* $X_n \to X$ *a.s.*

The sum condition above is known as "fast $\mathcal{L}^1$ convergence".

Proof. Define

$$S_n := \sum_{k=1}^{n} |X_k - X| \quad \text{ and } S := \lim_{n\to\infty} \nearrow S_n$$

By monotone integrability, we have

$$E[S] = \lim_{n\to\infty} \nearrow E[S_n] = \lim_{n\to\infty} \sum_{k=1}^{n} E[|X_k - X|] < \infty$$

by assumption, so S is finite a.s. Thus, for a.e. ω,

$$S(\omega) = \sum_{k=1}^{\infty} |X_k(\omega) - X(\omega)| < \infty \;\Rightarrow\; |X_k(\omega) - X(\omega)| \to 0$$

which means $X_k \to X$ a.s. □

Theorem 1.54. $X_n \to X$ *in probability* $\iff$ *for each subsequence* X_{n_k} *there is a further subsequence* $X_{n_{k_\ell}}$ *which converges to* X P*-a.s.*

Proof. (⇐) See measure theory (***)
(⇒) Suppose

$$P\left[|X_n - X| > \varepsilon\right] \xrightarrow[n\to\infty]{} 0 \quad \forall \varepsilon > 0$$

Choose a subsequence $K_1 < k_2 < \cdots < k_n < \cdots$ such that

$$P\left[|X_{k_n} - X| > \frac{1}{n}\right] < 2^{-n}$$

By Borel-Cantelli I 1.9, only finitely many of these events occur simultaneously. That is, for a.e. ω and $\forall n$ sufficiently large,

$$|X_{k_n}(\omega) - X(\omega)| \le \frac{1}{n}$$

But then, this implies $X_{k_n}(\omega) \to X(\omega)$. (Note that it is sufficient to work with the original sequence as opposed to a subsequence of a subsequence.) □

Before the next theorem and proof, we need to introduce the notion of *uniform integrability.*

Definition 1.55. *A collection* $\mathcal{H} \subseteq \mathcal{L}^1(\Omega, \mathcal{F}, P)$ *of functions is called* uniformly integrable *(written u.i., or sometimes called equi-integrable in measure theory) provided*

$$\lim_{c\to\infty} \searrow \sup_{X\in\mathcal{H}} E[|X|; |X| > c] = 0$$

Remark 1.56. If $X \in \mathcal{L}^1$ then $\{X\}$ is u.i. If $X \in \mathcal{L}^1$ and $\mathcal{H}$ u.i. then $\{X\} \cup \mathcal{H}$ is u.i. If $\mathcal{H}$ is $\mathcal{L}^1$-dominated, i.e.

$$\sup_{\mathcal{H}} |X_n| \le Y \in \mathcal{L}^1$$

then $\mathcal{H}$ is u.i. If $(X_n)_{n\ge1}$ is u.i. then $\liminf X_n, \limsup X_n \in \mathcal{L}^1$.

Theorem 1.57. *TFAE:*

1. $\mathcal{H}$ *is u.i.*
2. $\mathcal{H}$ *is* $\mathcal{L}^1$*-bounded (* $\iff \sup_{\mathcal{H}} E[|X|] < \infty$*) and* $\forall \varepsilon > 0\ \exists \delta > 0$ *such that*

$$\sup_{\mathcal{H}} E[|X|; A] < \varepsilon \quad \forall A \textit{ with } P(A) < \delta$$

3. $\exists g : \mathbb{R}^+ \to \mathbb{R}^+$ *Borel measurable with* $\frac{g(x)}{x} \to \infty$ *as* $x \to \infty$ *such that*

$$\sup_{\mathcal{H}} E[g(|X|)] < \infty$$

Example 1.58. An example of such a $g(x)$ in (3) above is $g(x) = |x|^p$ for $p > 1$. If $\mathcal{H}$ is $\mathcal{L}^p$ bounded then $\mathcal{H}$ is u.i. (but this is *not* true for $p = 1$). Also, $g(x) = x \log x$, etc.

Proof. See textbook. □

Theorem 1.59. *Let X_n, X be RVs. Then*

$$X_n \to X \text{ in } \mathcal{L}^1 \iff \begin{cases} X_n \to X \text{ in probability, and} \\ (X_n)_{n\geq 1} \text{ is u.i.} \end{cases}$$

Remark 1.60. The first condition implies $X_{n_k} \to X$ a.s., and by the previous theorem this implies $\{(X_n), X\}$ is u.i., which in turn implies that $\{|X_n - X|\}$ is u.i. Also, the theorem statement is about $|X_n - X|$ and therefore, WOLOG $X \equiv 0$ and $X_n \geq 0$.

Proof. ($\Rightarrow$) Observe that

$$E[|X_n - 0|] = E[X_n] = \underbrace{E\left[X_n; X_n \leq \varepsilon\right]}_{\leq \varepsilon \text{ always}} + \underbrace{E\left[X_n; X_n > \varepsilon\right]}_{:=(*)}$$

We claim $(*) < \varepsilon$ for $n \geq N$. For a given $\varepsilon > 0$, $\exists \delta = \delta(\varepsilon)$ such that if $P[A] < \delta$, we have

$$\sup_n E[X_n; A] < \varepsilon$$

by uniform integrability. Also, for $\delta = \delta(\varepsilon)$, we can choose $N = N(\delta, \varepsilon) = N(\varepsilon)$ such that

$$\sup_{n \geq N} P[X_n > \varepsilon] < \delta$$

by convergence in probability. This proves the claim.

($\Leftarrow$) Assume $X_n \to 0$ in $\mathcal{L}^1$. Then $X_n \to X$ in probability by Theorem 1.52. To prove the second condition, take $\varepsilon > 0$ and write

$$\sup_n E[X_n; X_n \geq c] \leq \sup_{n \leq N} E[X_n; X_n \geq c] + \sup_{n > N} E[X_n; X_n \geq c] =: E_1 + E_2$$

Choose $N = N(\varepsilon)$ such that

$$E_2 \leq \sup_{n > N} E[|X_n - 0|] \leq \varepsilon \quad \forall c$$

since the supremum in E_2 is over a quantity guaranteed to be $\leq E[X_n]$. Then, for the given N, choose c large enough such that $E_1 \leq \varepsilon$ (noting that the collection $\{X_1, \ldots, X_n\}$ is u.i.). These two estimates hold simultaneously for our choice of $N(\varepsilon)$ and $c = c(N) = c(\varepsilon)$. Thus $\exists c = c(\varepsilon)$ such that

$$\sup_n E[X_n; X_n \geq c] < 2\varepsilon$$

which implies $(X_n)_{n \in \mathbb{N}}$ is u.i. □

This concludes the analysis of modes of convergence.

1.7 Product Spaces

Let $(S_i, \mathfrak{S}_i)$ for $i = 1, 2$ be measurable spaces and set $S := S_1 \times S_2$. Let $X_i : S \to S_i$ be the coordinate maps (i.e. projections).

Definition 1.61. *A* stochastic kernel $K(x_1, dx_2)$ *from* S_1 *to* S_2 *is a map*

$$\begin{aligned} K : & S_1 \times \mathfrak{S}_2 \to [0,1] \\ & (x_1, A_2) \mapsto K(x_1, A_2) \end{aligned}$$

such that $K(x, \cdot)$ *is a probability measure on* $\mathfrak{S}_2$ *for all* $x \in S_1$ *and* $K(\cdot, A_2)$ *is* $\mathfrak{S}_1$*-measurable for all* $A_2 \in \mathfrak{S}_2$

Example 1.62.

1. Set $K(x, \cdot) = \mu(\cdot)\ \forall x \in S_1$. Then there is no dependency on x; i.e. "X_2 is independent of X_1" and so $X_2 \sim \mu$.
2. Set $K(x, \cdot) = \delta_{T(x)}(\cdot)$ for $T : S_1 \to S_2$ measurable. That is, "$X_2 = T \circ X_1$" i.e. X_2 depends deterministically on X_1.
3. **Countable Markov chain**: Let $S_1 = S_2 =: S$ be countable and set $\mathfrak{S} = \mathcal{P}(S)$. Let $K_{x,y}$ be a matrix with $K_{x,y} \geq 0$ and $\sum_y K_{x,y} = 1$ (i.e. a stochastic matrix). Set

   $$K(x, A) := \sum_{y \in A} K_{x,y}$$

   This is known as the *transition kernel.*
4. Set $S_1 = S_2 = \mathbb{R}$ and $K(x, \cdot) = \mathcal{N}(0, \beta x^2)$. **Question**: Does $\exists \beta > 0$ such that the Markov Chain converges to 0? What do we mean by "converges" in this case?

Let P_1 be a probability measure on $(S_1, \mathfrak{S}_1)$ and let K be a stochastic kernel from S_1 to S_2. We construct a probability measure $P (= P_1 \cdot K)$ on $\Omega := S_1 \times S_2$ such that

$$P[X_1 \in A_1] = P_1(A_1) \quad \text{for } A_1 \in \mathfrak{S}_1$$

and

$$\text{“}P\left[X_2 \in A_2 | X_1 = x_1\right] = K(x_1, A_2)\text{”} \quad \text{for } A_2 \in \mathfrak{S}_2$$

Definition 1.63. *The product* σ*-algebra is given by*

$$\mathcal{F} := \sigma\left(A_1 \times A_2 : A_i \in \mathfrak{S}_i\right)$$

The sets $A_1 \times A_2$ *are "rectangles" in the product space.*

Definition 1.64. *For* $A \in \mathcal{F}$ *and* $x_1 \in S_1$*, the set*

$$A_{x_1} := \{x_2 : (x_1, x_2) \in A\} \subseteq S_2$$

is called the x_1*-section of* A*.*

We will see that $A \in \mathcal{F} \Rightarrow A_{x_1} \in \mathfrak{S}_2$. Note that

$$\mathbf{1}_A(x_1, x_2) = \mathbf{1}_{A_{x_1}}(x_2)$$

Theorem 1.65. *1. The set function P defined by*

$$P[A] := \int_{S_1} P_1(dx_1) K(x_1, A_{x_1}) = \int_{S_1} \int_{S_2} K(x_1, dx_2) \mathbf{1}_A(x_1, x_2)\, P_1(dx_1)$$

is a probability measure on $(\Omega, \mathcal{F})$.

2. If $f \in \mathcal{F}$ and f is semi-integrable w.r.t. P, then

$$\int_\Omega f\, dP = E[f] = \int_{S_1} P_1(dx_1) \int_{S_2} K(x_1, dx_2) f(x_1, x_2)$$

Proof. We prove (1); claim (2) follows from (1) by MTI. To check that P is, indeed, a probability measure, we verify

1. $P[\Omega] = 1$ is true
2. If $A = \biguplus_i A_i$ then, applying Monotone Integrability twice, we have

$$\begin{aligned} P\left[\biguplus_i A_i\right] &= \int_{S_1} P_1(dx_1) \int_{S_2} K(x_1, dx_2) \left(\sum_i \mathbf{1}_{A_i}(x_1, x_2)\right) \\ &= \int_{S_1} P_1(dx_1) \sum_i \int_{S_2} K(x_1, dx_2) \mathbf{1}_{A_i}(x_1, x_2) \\ &= \sum_i \int_{S_1} P_1(dx_1) \int_{S_2} K(x_1, dx_2) \mathbf{1}_{A_i}(x_1, x_2) \\ &= \sum_i P(A_i) \end{aligned}$$

   which is what we want.

This proves the theorem. □

Lemma 1.66. *For all $x_1 \in S_1$ and $f \in \mathcal{F}^+$ (meaning $f \geq 0$ and f is $\mathcal{F}$-measurable), we have $f_{x_1}(\cdot) := f(x_1, \cdot)$ is $\in \mathfrak{S}_2$. Furthermore, $f \in \mathcal{F}^+$ implies that the function φ defined by*

$$x_1 \mapsto \int K(x_1, dx_2) f(x_1, x_2) \in \bar{\mathbb{R}}^+$$

is well-defined and φ is $\in \mathfrak{S}_1^+$.

Proof. (***) homework □

Note that the first statement implies $A_{x_1} \in \mathfrak{S}_2$, and the second statement implies

$$\int \varphi(x_1)P_1(dx_1) = \int_{S_1} P_1(dx_1) \int_{S_2} K(x_1, dx_2) f(x_1, x_2)$$

is well-defined. These conclusions are used in the proof of the theorem above.

A classical case of the theorem above is Fubini's Theorem. Let $K(x, \cdot) := P_2(\cdot)$ so there is no x_1-dependence. Let $P = P_1 \cdot P_2$ and

$$P[A] = \int_{S_1} P_1(dx_1) \int_{S_2} P_2(dx_2) \mathbf{1}_A(x_1, x_2)$$

Let's define $\tilde{P}$ on $\Omega = S_1 \times S_2$ with σ-algebra $\mathcal{F}$ as follows:

$$\tilde{P}[A] = \int_{S_2} P_2(dx_2) \int_{S_1} P_1(dx_1) \mathbf{1}_A(x_1, x_2)$$

This corresponds to a constant kernel $\tilde{K}$ from S_2 to S_1 given by $\tilde{K}(x_2, dx_1) = P_1(dx_1)$. Then, by MTI, for $f \in \mathcal{F}$ with $f \geq 0$, we have

$$\int_\Omega f \, d\tilde{P} = \int_{S_2} P_2(dx_2) \int_{S_1} P_1(dx_1) f(x_1, x_2)$$

Note, however, that $\tilde{P} = P$ since they agree on rectangles,

$$P[A_1 \times A_2] = P_1(A_1) \cdot P_2(A_2) = \tilde{P}(A_1 \times A_2)$$

and rectangles are $\cap$-closed and generate $\mathcal{F}$. Therefore, the equality holds $\forall f \in \mathcal{F}^+$, so

$$\int_{S_1} P_1(dx_1) \int_{S_2} P_2(dx_2) f(x_1, x_2) = \int_{S_2} P_2(dx_2) \int_{S_1} P_1(dx_1) f(x_1, x_2)$$

This equality is Fubini's Theorem.

Remark 1.67. Fubini is valid for σ-finite measures only! Also, the integrand must be semi-integrable w.r.t P.

Example 1.68. Consider the following application of Fubini's Theorem. Let $X \geq 0$ be a RV. Then

$$E[X] = \int_0^\infty P[X > s] \, \lambda(ds)$$

Proof. Observe that

$$\begin{aligned} \int_0^\infty P[X > s] \, ds &= \int_0^\infty \left(\int_\Omega P(d\omega) \mathbf{1}_{(S,\infty]}(X(\omega)) \right) ds \\ &= \int_\Omega P(d\omega) \cdot \int_0^\infty \mathbf{1}_{(-\infty, X(\omega))}(s) \, ds \\ &= \int_\Omega X(\omega) P(d\omega) = E[X] \end{aligned}$$

since $P[X > s] = E\left[\mathbf{1}_{(s,\infty]}(X)\right]$. □

Remark 1.69. Notice that when $\mu(\Omega) < \infty$,

$$\int f\, d\mu = \int_0^\infty \mu(f > c)\, dc$$

assuming $f \geq 0$. Also,

$$\int |f|\, d\mu \geq \sup_c c \cdot \mu(|f| > c)$$

1.7.1 Infinite product spaces

This short section presents the powerful Ionescu-Tulcea Theorem. Consider a (countable) sequence of measurable spaces $(S_i, \mathfrak{S}_i)_{i\geq 0}$ and define

$$(S^n, \mathfrak{S}^n) := \left(\prod_{i=0}^n S_i \ ,\ \sigma\left(\{A_1 \times A_2 \times \cdots \times A_n : A_i \in \mathfrak{S}_i\}\right) \right)$$

Let μ_0 be a normed measure on S_0, and for $n \geq 1$ let K_n be a stochastic kernel from S^{n-1} to S_n; i.e. $K(x_0 x_1 \ldots x_{n-1}, dx_n)$ with $K_n(x_0 \ldots x_n, S_n) = 1$. Set $\mu^0 := \mu_0$ and iteratively define

$$\mu^n := \mu^{n-1} \cdot K_n$$

to be a measure on $\mathfrak{S}^n$. That is, for $f \in (\mathfrak{S}^n)^+$,

$$\begin{aligned}
\int_{S^n} f\, d\mu^n &= \int_{S^{n-1}} \mu^{n-1}(dy) \int_{S_n} K_n(y, dx_n) f(y, x_n) \\
&= \int_{S^{n-1}} \mu^{n-1}\left(d(x_0 \ldots x_{n-1})\right) \int_{S_n} K_n(x_0 \ldots x_{n-1}, dx_n) f(x_0 \ldots x_n) \\
&= \int_{S_0} \mu^0(dx_0) \cdot \int_{S_1} K_1(x_0, dx_1) \cdot \int_{S_2} K_2(x_0 x_1, dx_2) \cdots \\
&\qquad \cdots \int_{S_n} K_n(x_0 \ldots x_{n-1}, dx_n) f(x_0 \ldots x_n)
\end{aligned}$$

Set

$$X = \prod_{i=1}^\infty S_i = \{x = (x_0, x_1, \ldots) : x_i \in S_i\}$$

and define the canonical projections $\pi_i : X \to S_i$ by $\pi_i(x) = x_i$. Also, set

$$\mathcal{A}_n := \sigma\left(\pi_0, \pi_1, \ldots, \pi_n\right) = \{A^n \times S_{n+1} \times S_{n+2} \times \cdots : A^n \in \mathfrak{S}^n\}$$

and

$$\mathcal{A} = \sigma\left(\pi_0, \ldots, \pi_n, \ldots\right) = \sigma\left(\bigcup_{n\geq 0} \mathcal{A}_n \right)$$

Question: Does $\exists\mu$ a probability measure on $(X, \mathcal{A})$ such that

$$\mu \circ \pi^{-1}_{\{0,\dots,n\}} = \mu^n \quad \forall n \quad ?$$

That is, we want to guarantee that μ satisfies

$$\mu\left(A^n \times S_{n+1} \times S_{n+2} \times \cdots\right) := \mu^n(A_n) \tag{3}$$

Answer: Yes, and this measure μ is *unique*! This is the conclusion of the Ionescu-Tulcea Theorem, with the conditions being the discussion in this section leading up to this sentence.

Proof. First, observe that $\bigcup_n \mathcal{A}_n$ is an algebra and μ is *consistently* defined on $\bigcup_n \mathcal{A}_n$ by Equation (3); that is, for $A \in \mathcal{A}_n \cap \mathcal{A}_{n-1}$, we can write

$$A = A^n \times S_{n+1} \times \cdots = \underbrace{(A^{n-1} \times S_n)}_{=A^n} \times S_{n+1} \times \cdots$$

and have

$$\begin{aligned}
\mu^n(A^n) &= \mu^n\left(A^{n-1} \times S_n\right) \\
&= \int_{S^{n-1}} \mu^{n-1}(d(x_0 \dots x_{n-1})) \\
&\qquad \cdot \int_{S_n} K(x_0 \dots x_{n-1}, dx_n) \mathbf{1}_{A^{n-1}}(x_0 \dots x_{n-1}) \mathbf{1}_{S_n}(x_n) \\
&= \mu^{n-1}(A^{n-1}) \cdot 1
\end{aligned}$$

Also, observe that μ is additive on $\bigcup_n \mathcal{A}_n$ (which implies monotonicity). This is easy and follows from the additivity of μ_n.

We now have to show that μ is σ-additive; given $A_n \in \bigcup_n \mathcal{A}_n$ with $A_n \searrow \emptyset$, we need $\lim_n \mu(A_n) = 0$. WOLOG we can take $A_n \in \mathcal{A}_n$. To see why this is okay, let

$$A_1, A_2, \dots \in \bigcup_n \mathcal{A}_n \;\Rightarrow\; \forall k, A_k \in \mathcal{A}_{m_k}$$

Set $n_1 = m_1$, $n_2 = m_2 \vee (n_1 + 1), \dots, n_k = m_k \vee (n_{k-1} + 1), \dots$ and so on. Then $n_1 < n_2 < n_3 < \cdots$ and $A_k \in \mathcal{A}_{n_k}$ since the collection $\mathcal{A}_n$ is increasing in n. Now, define $B_{n_k} = A_k$ for $k \geq 1$ and fill the "gaps" in the sequence as follows:

$$\begin{gathered}
B_1, \dots, B_{n_1 - 1} = \Omega \\
B_{n_1}, \dots, B_{n_2 - 1} = A_1 \\
\vdots \\
B_{n_k}, \dots, B_{n_{k+1} - 1} = A_k \\
\vdots
\end{gathered}$$

Then $B_k \in \mathcal{A}_k$ and $B_k \searrow \emptyset$ and $\lim \mu(A_k) = \lim \mu(B_k)$. Now, using this assumption, we can write

$$A_n = A^n \times S_{n-1} \times S_{n+2} \times \cdots \quad \text{and} \quad A_{n+1} = A^{n+1} \times S_{n+2} \times \cdots$$

and so $A^{n+1} \subseteq A^n \times S_n$. Now, assume by way of contradiction that $\inf_n \mu(A_n) > 0$. Then

$$\begin{aligned} \mu(A_n) &= \mu\left(A^n \times S_s n + 1 \times \cdots\right) = \mu^n(A^n) \\ &= \int_{S_0} \mu_0(dx_0) \cdot \int_{S_1} K_1(x_0, dx_1) \cdot \\ &\quad \cdots \int_{S_n} K_n(x_0 \ldots x_{n-1}, dx_n) \mathbf{1}_{A^n}(x_0 \ldots x_n) \\ &=: int_{S_0} \mu_0(dx_0) f_{0,n}(x_0) \end{aligned}$$

and notice that $f_{0,n}(x_0) \searrow$ in n: by assumption,

$$\inf_{n \geq 1} \mu(A_n) = \inf_{n \geq 1} \int_{S_0} \mu_0(dx_0) f_{0,n}(x_0) > 0$$

and by monotone integration,

$$\exists \bar{x}_0 \text{ such that } \inf_n f_{0,n}(\bar{x}_0) > 0$$

Thus,

$$\begin{aligned} f_{0,n}(x_0) &= \int_{S_1} K_1(x_0, dx_1) \cdots \int_{S_n} K_n(x_0 \ldots x_{n-1}, dx_n) \mathbf{1}_{A^n}(x_0 \ldots x_n) \\ &\leq \int_{S_1} K_1(x_0, dx_1) \cdot \\ &= f_{0,n-1}(x_0) \end{aligned}$$

since

$$\mathbf{1}_{A^n}(x_0 \ldots x_n) \leq \mathbf{1}_{A^{n-1} \times S_n}(x_0 \ldots x_n) = \mathbf{1}_{A^{n-1}}(x_0 \ldots x_n)$$

This shows that $f_{0,n}(x_0) \searrow$ in n. Similarly, $\forall k \geq$ and $\forall x_0 \ldots x_k$, with $n > k$ we have

$$\begin{aligned} f_{k,n}(x_0 \ldots x_k) &:= \int_{S_{k+1}} K_{k+1}(x_0 \ldots x_k, dx_{k+1}) \cdot \\ &\quad \cdots \int_{S_n} K_n(x_0 \ldots x_{n-1}, dx_n) \mathbf{1}_{A^n} \\ &\leq f_{k,n-1}(x_0 \ldots x_k) \end{aligned}$$

since $\mathbf{1}_{A^n} \leq \mathbf{1}_{A^{n-1} \times S_n}$. This shows $f_{k,n}(x_0 \ldots x_k) \searrow$ in n, as well.

Now, it follows from $\inf_n f_{n,0}(\bar{x}_0) > 0$ that $\exists \bar{x}_1 \in S_1$ such that

$$\inf_{n\geq 2} \int_{S_2} K_2(\bar{x}_0\bar{x}_1, dx_2) \cdot \int_{S_3} K_3(\bar{x}_0\bar{x}_1 x_2, dx_3) \cdot \\ \cdots \int_{S_n} K_n(\bar{x}_0\bar{x}_1 x_2 \ldots x_{n-1}, dx_n) \mathbf{1}_{A^n}(\bar{x}_0\bar{x}_1 \ldots x_n) > 0$$

That is, $\inf_n f_{1,n}(\bar{x}_0, \bar{x}_1) > 0$. Iterating this process shows us that $\forall k\ \exists \bar{x}_k \in S_k$ such that

$$\inf_{n\geq k+1} \int_{S_{k+1}} K_{k+1}(\bar{x}_0 \ldots \bar{x}_k, dx_{k+1}) \cdots \int_{S_n} K_n(\bar{x}_0 \ldots \bar{x}_k x_{k+1} \ldots x_{n-1}, dx_n) \\ \cdot \mathbf{1}_{A^n}(\bar{x}_0 \ldots \bar{x}_k x_{k+1} \ldots x_n) =: \inf_{n\geq k+1} f_{k,n}(\bar{x}_0 \ldots \bar{x}_k) > 0$$

In particular, for $u = k+1$,

$$\int_{K_{k+1}} (\bar{x}_0 \ldots \bar{x}_k, dx_{k+1}) \, \mathbf{1}_{A^{k+1}} \, (\bar{x}_0 \ldots \bar{x}_k, x_{k+1}) > 0$$

since $\mathbf{1}.(\cdot) \neq 0$ and since

$$A^{k+1} \subseteq A^k \times S_{k+1} \ \Rightarrow \ (\bar{x}_0, \bar{x}_1, \ldots) \in \bigcap_k A_k \neq \emptyset$$

This completes the proof. □

2 Laws of Large Numbers

First applications to a classical limit theorem of probability.

Theorem 2.1 (Weak Law of Large Numbers)**.** *Let $X_1, X_2, \ldots$ be a sequence of uncorrelated i.i.d. RVs with finite variance σ^2 and mean μ. Set*

$$\bar{X}_n = \frac{1}{n} S_n = \frac{1}{n} \sum_{k=1}^{n} S_k$$

Then

1. *$\bar{X}_n \to \mu$ in $\mathcal{L}^2$, i.e. $E[|\bar{X}_n - \mu|^2] \to 0$, where we think of μ as a constant RV.*

2. *$\bar{X}_n \to \mu$ in probability, i.e.*

$$\lim_{n\to\infty} P\left[|\bar{X}_n - \mu| \geq \varepsilon\right] = 0 \quad \forall \varepsilon > 0$$

Proof. To prove (1), observe that

$$E[\bar{X}_n] = \frac{1}{n} \sum_{k=1}^{n} E[X_k] = \mu$$

and so

$$\begin{aligned} E\left[|\bar{X}_n - \mu|^2\right] = \operatorname{Var}\left[\bar{X}_n\right] &= \frac{1}{n^2} \operatorname{Var}\left[\sum_{k=1}^{n} X_K\right] \\ &= \frac{1}{n^2} \sum_{k=1}^{n} \operatorname{Var}\left[X_k\right] = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \xrightarrow[n\to\infty]{} 0 \end{aligned}$$

To prove (2), we apply (1) to say that $\bar{X}_n \to \mu$ in $\mathcal{L}^2$ and the appeal to Lemma 2.2 below to conclude that $\bar{X}_n \to \mu$ in probability, as well. □

Lemma 2.2. *$\mathcal{L}^p$ convergence $\Rightarrow$ convergence in probability; i.e. if $X_k \to X$ in $\mathcal{L}^p$ for $p > 0$, then for every $\varepsilon > 0$, $\lim_n P[|X_n - X| \geq \varepsilon] = 0$.*

Proof. WOLOG $X = 0$ (just set $X_n' = X_n - X$, say). Then $\mathcal{L}^p$ convergence says $E[|X_n - X|^p] \to)$ as $n \to \infty$. Let $\varepsilon > 0$. Then we apply the Chebyshev-Markov Inequality 1.47 with $A = \{X : |X| \geq \varepsilon\}$ and $\varphi = |X|^p$ and $c_A = \varepsilon^p$ to write

$$P\left[|X_n| \geq \varepsilon\right] \leq \varepsilon^{-p} E\left[|X_n|^p\right] \xrightarrow[n\to\infty]{} 0$$

□

Lemma 2.3. *Let $X \geq 0$ be a RV on $(\Omega, \mathcal{F}, P)$. Let $F : [0, \infty) \to \mathbb{R}$ be absolutely continuous, i.e. $F(x) = \int_0^x f(t)\, dt$ for some $\mathcal{L}^1 \ni f \geq 0$ measurable. Then*

$$E[F(X)] = \int_0^\infty P[X > t] f(t)\, dt = \int_0^\infty P[X \geq t] f(t)\, dt$$

Proof. Homework exercise (***) □

Lemma 2.4. *Let $X \geq 0$ a.s. Then*

$$\sum_{k\geq 1} P[X \geq k] \leq E[X] \leq \sum_{n \geq 0} P[X > n]$$

Proof. Define $\varphi(t)$ and $\psi(t)$ to be the upper and lower step functions, respectively; that is,

$$\varphi(t) = n \text{ for } t \in (n-1, n] \quad \text{ and } \psi(t) = n - 1 \text{ for } t \in (n-1, n]$$

so that $\psi(t) \leq t \leq \varphi(t)$ for any t. We now work with the LHS and write

$$\begin{aligned} LHS &= \sum_{k\geq 1} \int_{(k-1,k]} P[X \geq \varphi(t)] \\ &\leq \int_{(0,\infty]} P[X \geq t]\,dt = E[X] \\ &= \int_{(0,\infty]} P[X > t]\,dt \leq \int_{(0,\infty)} P[X > \psi(t)]\,dt \\ &= \sum_{k\geq 0} \int_{(k,k+1]} P[X > \psi(t)]\,dt = \sum_{k\geq 0} P[X > k] \end{aligned}$$

□

Remark 2.5. If $X \geq 0$ and $X \in \mathbb{N}$ then LHS=RHS= $E[X]$.

Theorem 2.6 (Strong Law of Large Numbers, Etemardi)**.** *Assume* $X_1, X_2, \ldots$ *are* pair-wise *independent, identically distributed RVs with* $E[X_i] =: \mu < \infty$ *for all* i. *Let* $S_n := \sum_{k=1}^n X_k$. *Then* $\frac{S_n}{n} \to \mu$, P*-a.s.*

Proof. We follow 5 steps.

1. **WOLOG** $X_1 \geq 0$. Write $X_i = X_i^+ - X_i^-$. Then the $(X_i^+)_i$ are pair-wise independent (***), identically distributed with $E[X_i^+] < \infty$ and $X_1^+ \geq 0$. The same holds for the X_i^-, as well. Moreover,

$$\frac{1}{n}\sum_{i=1}^n X_i^{\pm} \xrightarrow[n\to\infty]{\text{a.s.}} E\left[X^{\pm}\right] \Rightarrow \frac{1}{n}\sum_{i=1}^n (X_i^+ - X_i^-) \to E[X] \text{ a.s.}$$

2. **Truncation**. Let $Y_k := X_k \cdot \mathbf{1}_{[X_k \leq k]} \geq 0$. Then the $(Y_i)_i$ are still independent (***). Let $T_n = \sum_{k=1}^n Y_k$. It will be easy to show that $\frac{T_n}{n} \to \mu$ a.s., since

$$\sum_{k\geq 1} P[X_k > k] = \sum_{k\geq 1} P[X_1 > k] \leq E[X_1] < \infty$$

by Lemma 2.4, and so

$$\sum_{k\geq 1} P[X_k > k] = \int_0^\infty P[X_1 > \varphi(t)]\,dt \leq \int_0^\infty P[X_1 > t]\,dt = \mu < \infty$$

where φ is the upper step function we used in the proof of Lemma 2.4. Applying Borel-Cantelli 1.9, we can conclude that for a.e. ω, $X_k(\omega) = Y_k(\omega)$ for all $k \geq k_0(\omega)$. But then,

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n X_k(\omega) = \lim_{n\to\infty} \Bigg\{ &\frac{1}{n}\sum_{k=1}^{k_0(\omega)} (X_k(\omega) - Y_k(\omega)) \\ &+ \frac{1}{n}\sum_{k=1}^n Y_k(\omega) \Bigg\} \end{aligned}$$

and since the first term is a finite sum, we can conclude that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k(\omega) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} Y_k(\omega)$$

for a.e. ω.

3. **Variance estimates**. This part is quite technical. We claim

$$\sum_{k=1}^{\infty} \frac{\operatorname{Var}[Y_k]}{k^2} \le 4E[X_1] < \infty$$

where, really, any constant will do instead of 4. We apply Lemma 2.3 to write

$$\operatorname{Var}[Y_k] \le E[Y_k^2] = \int_0^\infty 2tP[Y_k > t]\,dt \le \int_0^k 2tP[X_k > t]\,dt$$

and thus

$$\begin{aligned}\sum_{k=1}^{\infty} \frac{\operatorname{Var}[Y_k]}{k^2} &\le \sum_{k\ge 1} \frac{1}{k^2} \int_0^\infty \mathbf{1}_{[0,k)}(t) 2tP[X_1 > t]\,dt \\ &= \int_0^\infty 2t \left(\sum_{k\ge 1} \frac{1}{k^2} \mathbf{1}_{(t,\infty)}(k) \right) P[X_1 > t]\,dt \\ &\le 4 \int_0^\infty P[X_1 > t]\,dt = 4E[X_1]\end{aligned}$$

since

$$2t \left(\sum_{k\ge 1} \frac{1}{k^2} \mathbf{1}_{(t,\infty)}(k) \right) = 2t \sum_{k>t} \frac{1}{k^2} \sim 2t \cdot \frac{1}{t} \le 4$$

or some other constant, it doesn't really matter ...

4. **Convergence along a subsequence**. We **claim** $\frac{T_{k_n}}{k_n} \to \mu$ a.s. as $n \to \infty$. By Chebyshev-Markov 1.47, for an arbitrary subsequence,

$$\begin{aligned}\sum_{n=1}^{\infty} P\left[\frac{1}{k_n} \left| T_{k_n} - E\left[T_{k_n}\right] \right| > \varepsilon \right] &\le \frac{1}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{1}{k_n^2} \operatorname{Var}\left[T_{k_n}\right] \\ &= \frac{1}{\varepsilon^2} \sum_{n\ge 1} \frac{1}{k_n^2} \sum_{m=1}^{k_n} \operatorname{Var}\left[Y_m\right] \\ &= \frac{1}{\varepsilon^2} \sum_{m\ge 1} \operatorname{Var}\left[Y_m\right] \underbrace{\sum_{n:k_n\ge m} \frac{1}{k_n^2}}_{:=\Gamma}\end{aligned}$$

To verify the last equality, we choose $\alpha > 1$ and set $k_n := \lfloor \alpha^n \rfloor$. Then observe that

$$\begin{aligned}\Gamma &\le \sum_{n:\lfloor \alpha^n \rfloor \ge m} \frac{1}{\lfloor \alpha^n \rfloor^2} \le 4 \sum_{n:\alpha^n \ge m} \frac{1}{\alpha^{2n}} \\ &\le 4\left(\frac{1}{\alpha^{2n_0}} + \frac{1}{\alpha^{2(n_0+1)}} + \cdots\right) \\ &\le 4\frac{1}{m^2}\left(1 + \frac{1}{\alpha^2} + \frac{1}{\alpha^4} + \cdots\right) = \frac{4}{m^2} \cdot \frac{1}{1-\alpha^{-2}}\end{aligned}$$

Now, we have for $\alpha > 1$ fixed and $k_n = \lfloor \alpha^n \rfloor$,

$$\begin{aligned}\sum_{n\ge 1} P\left[\frac{1}{k_n}\left|T_{k_n} - E\left[T_{k_n}\right]\right| > \varepsilon\right] &\le \frac{4}{\varepsilon^2} \cdot \frac{1}{1-\alpha^{-2}} \sum_{m\ge 1} \frac{\mathrm{Var}\left[Y_m\right]}{m^2} \\ &\le \frac{16}{\varepsilon^2} \cdot \frac{1}{1-\alpha^{-2}} \mu < \infty\end{aligned}$$

Then, by Borel-Cantelli 1.9, the set

$$\begin{aligned}A_\varepsilon &= \left\{\frac{1}{k_n}\left|T_{k_n} - E\left[T_{k_n}\right]\right| > \varepsilon \text{ only finitely many times}\right\} \\ &\equiv \left\{\frac{1}{k_n}\left|T_{k_n} - E\left[T_{k_n}\right]\right| \le \varepsilon \text{ for all suffic. large } n\right\}\end{aligned}$$

satisfies $P[A_\varepsilon] = 1$. Let $A := \bigcap_{j\ge 1} A_{1/j}$. Then $P[A] = 1$ and

$$\frac{1}{k_n}\left|T_{k_n} - E\left[T_{k_n}\right]\right| \xrightarrow[n\to\infty]{} 0 \text{ on } A$$

But $E[Y_k] \nearrow E[X_1] = \mu$ as $k \to \infty$, so by monotone convergence, $\frac{E[T_{k_n}]}{k_n} \to \mu$. Thus, on A,

$$\mathrm{dist}\left(\frac{T_{k_n}}{k_n}, \frac{E[T_{k_n}]}{k_n}\right) \xrightarrow[n\to\infty]{} 0$$

since each sequence $\to \mu$.

5. **Filling the gap between the subsequence and the full sequence**. For $k_n \le m \le k_{n+1}$, we have

$$\frac{k_n}{k_{n+1}} \cdot \frac{T_{k_n}}{k_n} = \frac{T_{k_n}}{k_{n+1}} \le \frac{T_{k_n}}{m} \le \frac{T_m}{m} \le \frac{T_{k_{n+1}}}{m} \le \frac{T_{k_{n+1}}}{k_n} = \frac{T_{k_{n+1}}}{k_{n+1}} \cdot \frac{k_{n+1}}{k_n}$$

Notice that

$$\frac{k_{n+1}}{k_n} = \frac{\lfloor \alpha^{n+1} \rfloor}{\lfloor \alpha^n \rfloor} \xrightarrow[n\to\infty]{} \alpha$$

and so the line above reads, in the limit,

$$\frac{1}{\alpha}\mu \le \liminf_{m\to\infty} \frac{T_m}{m} \le \limsup_{m\to\infty} \frac{T_m}{m} \le \alpha\mu \text{ a.s.}$$

Since $\alpha > 1$ is arbitrary, $\lim \frac{T_m}{m} = \mu$ a.s. (let $\alpha = 1 + \frac{1}{n}$, for instance).

□

Theorem 2.7 (Strong LLN for semintegrable functions)**.** *Let $(X_i)_i$ be i.i.d. with $E[X_1^+] = +\infty$ and $E[X_i^-] < \infty$. Then*

$$\frac{S_n}{n} \xrightarrow[n\to\infty]{} E[X_1] = +\infty \quad a.s.$$

Proof. **Truncation**: Let $M \in \mathbb{N}$ large be fixed, and let

$$X_i^M := M \wedge X_i \quad \text{and} \quad S_n^M = \sum_{k=1}^n X_i^M$$

Then $(X_i^M)_i$ are i.i.d. with finite mean μ^M. As $M \to \infty$, $\mu^M \nearrow \infty$ by monotone integration. Define the sets

$$A_M := \left\{ \liminf_{n\to\infty} \frac{S^M}{n} \geq \mu^M - i \right\}$$

and note $P[A_M] = 1$, so

$$A := \bigcap_{M\geq 1} A_m \;\Rightarrow\; P[A] = 1$$

and on A,

$$\liminf_n \frac{S_n}{n} \geq \liminf_n \frac{S_n^M}{n} \geq \mu^M - 1 \; \forall M \;\Rightarrow\; \liminf_n \frac{S_n}{n} \geq \infty$$

since $\mu^M \to \infty$. □

What if the X_i are not semi-integrable?

Theorem 2.8. *Let $(X_i)_i$ be i.i.d. with $E[|X_i|] = +\infty$. Then*

$$\limsup_{n\to\infty} \left| \frac{S_n}{n} \right| = +\infty \; a.s.$$

In particular,

$$P\left[\limsup_{n\to\infty} \left| \frac{S_n}{n} \right| < \infty \right] > 0 \;\Rightarrow\; X_1 \in \mathcal{L}^1$$

never mind converging a.s. to some finite RV! Anyway, $\limsup \left|\frac{S_n}{n}\right|$ *is constant by Kolmogorov's 0-1 law 1.20.*

This (in some way) shows that the $\mathcal{L}^1$ condition is necessary for the Strong LLN 2.6.

Proof. Notice

$$|X_n| = |S_n - S_{n-1}| \le |S_n| + |S_{n-1}| \;\Rightarrow\; \limsup_{n\to\infty} \left|\frac{X_n}{n}\right| \le 2\limsup_{n\to\infty}\left|\frac{S_n}{n}\right|$$

This tells us it suffices to show $\limsup \left|\frac{X_n}{n}\right| = \infty$. Fix $C \ge 1$. Then

$$\sum_{n=0}^{\infty} P\left[\frac{|X_n|}{n} \ge C\right] = \sum_{n=0}^{\infty} P\left[\frac{|X_1|}{C} \ge n\right] \ge E\left[\frac{|X_1|}{C}\right] = \infty$$

by Lemma 2.4. Since the X_i are independent, then Borel-Cantelli II 1.18 says that infinitely many of the events $\left\{\frac{|X_n|}{n} \ge C\right\}$ occur a.s. Thus,

$$\limsup_{n\to\infty} \frac{|X_n|}{n} \ge C \text{ a.s.}$$

Since C is arbitrary, we can take intersections and conclude that

$$\limsup_{n\to\infty} \frac{|X_n|}{n} = +\infty \text{ a.s.}$$

□

2.1 Examples and Applications of LLN

1. **Renewals**: Let $(X_i)_i$ be i.i.d. with $0 < X_i < \infty$ and set $T_n = \sum_{k=1}^{n} X_k$. We interpret the X_j as waiting times and T_n as the time of the nth occurrence. Set

   $$N_t = \sup\{n : T_n \le t\} = \text{ \# of occurrences up to time } t$$

   Theorem 2.9. *If* $E[X_1] = \mu \le \infty$ *then* $\frac{N_t}{t} \to \frac{1}{\mu}$ *a.s.*

   Proof. By Strong LLN 2.6, $\frac{T_n}{n} \to \mu$ a.s. By the definition of N_t, $T_{N_t} \le t < T_{N_t+1}$, and dividing by N_t gives

   $$\underbrace{\frac{T_{N_t}}{N_t}}_{\to\mu} \le \frac{t}{N_t} < \underbrace{\frac{T_{N_t+1}}{N_t+1}}_{\to\mu} \cdot \underbrace{\frac{N_{t+1}}{N_t}}_{\to 1}$$

   and so $\frac{t}{N_t} \to \mu$. Note that we have used the fact that $N_t \to \infty$ as $t\infty$ a.s. and so $\frac{T_n}{n} \to \mu$ a.s. implies $\frac{T_{N_t}}{N_t} \to \mu$. □

2. **Glivenko-Cantelli Theorem**: Let $(X_i)_i$ be i.i.d. with *arbitrary* distribution F. Consider the *empirical distribution functions*

   $$F_n(x) = F_n(x,\omega) := \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{(-\infty,x]}(X_k(\omega))$$

Note that $F_n(x, \omega)$ is the observed *frequency* of values $\leq x$.
Claim: for all x, $F_n(x) \xrightarrow[n\to\infty]{} F(x)$ a.s.; i.e. $F_n(x)$ is a "good estimator" of F. To prove this claim, let

$$Y_n(\omega) = \mathbf{1}_{\{X_n \leq x\}}$$

so that (Y_n) are i.i.d. and

$$E[Y_n] = P[X_n \leq x] = F(x)$$

and so by the Strong LLN 2.6

$$\frac{1}{n}\sum_{k=1}^{n} Y_n = F_n(x, \omega) \to E[Y_1] = F(x)$$

Theorem 2.10.

$$\sup_{x\in\mathbb{R}} |F_n(x, \omega) \to F(x)| \xrightarrow[n\to\infty]{} 0 \quad a.s.$$

Proof. Let

$$F(x^-) := \lim_{y \nearrow x} F(y) = P[X_1 < x]$$

By setting $Z_n(\omega) = \mathbf{1}_{\{X_n < x\}}$, then $F_n(x^-) \to F(x^-)$ a.s. for all x. Fix $1 \leq k \in \mathbb{N}$. For $1 \leq j \leq k-1$, let

$$x_j := \inf\left\{y : F(y) \geq \frac{j}{k}\right\}$$

so then $F(x_j^-) - F(x_{j-1}) \leq \frac{1}{k}$. Also, for a.e. ω, $\exists N = N(k, \omega)$ such that $\forall n \geq N$ and $\forall 0 \leq j \leq k$, we have

$$|F_n(x_j) - F(x_j)| < \frac{1}{k} > \left|F_n(x_j^-) - F(x_j^-)\right|$$

Applying these three inequalities and monotone convergence, we can write

$$F_n(x) \leq F_n(x_j^-) \leq F(x_j^-) + \frac{1}{k} \leq F(x_{j-1}) + \frac{2}{k} \leq F(x) + \frac{2}{k}$$

and

$$F_n(x) \geq F)n(x_{j-1}) \geq F(x_{j-1}) - \frac{1}{k} \geq F(x_j^-) - \frac{2}{k} \geq F(x) - \frac{2}{k}$$

which implies that for a.e. ω and $\forall n \geq N(\omega, k)$, we have

$$\sup_x |F_n(x, \omega) - F(x)| \leq \frac{2}{k}$$

which proves the claim since k is arbitrary. □

3. **Monte-Carlo Integration**: How can we compute (i.e. approximate) an integral of the form

$$I := \int \cdots \int_{[0,1]^n} \varphi\left(x_1, x_2, \ldots, x_n\right) \, dx_1 dx_2 \ldots dx_n$$

for a potentially irregular, complicated φ? The main idea (due to Fermi) is to use the Transformation Formula 1.49 and the Strong LLN 2.6!

Assume $(X_i)_{i=1\ldots n}$ are i.i.d. with X_1 uniform in $[0,1]$. Thus, $\vec{X}$ has the distribution λ_n and so

$$E\left[\varphi(\vec{X})\right] = \int_{[0,1]^n} \varphi(x) \, \lambda_n(dx) = I$$

by the Transformation Formula. Accordingly, the integral in question boils down to finding $E\left[\varphi(\vec{X}(\omega))\right]$. By the Strong LLN,

$$\frac{1}{m} \sum_{k=1}^{m} \varphi\left(\vec{X}_k(\omega)\right) \xrightarrow[m \to \infty]{} E\left[\varphi(\vec{X})\right] \quad \text{a.s.}$$

so, we can generate i.i.d. random vectors (with uniform distribution on $[0,1]^n$) $\vec{X}_1, \ldots, \vec{X}_k, \ldots$ and use sums of the form $\frac{1}{m} \sum_{k=1}^{m} \varphi\left(\vec{X}_k(\omega)\right)$ to approximate the integral in question.

3 Weak Convergence of Probability Measures

Let (Ω, ρ) be a metric space and $\mathcal{F} = \sigma(\tau)$ where τ is open sets.

Definition 3.1. *Let $(\mu_n)_{n \geq 1}, \mu$ be probability (or finite) measures on $(\Omega, \mathcal{F})$. Then $\mu_n \xrightarrow{w} \mu$ (read: "the μ_n converge weakly to μ") if and only if for all $\varphi \in \mathcal{C}_b$ (bounded continuous RVs)*

$$\int \varphi d\mu_n \xrightarrow[n \to \infty]{} \int \varphi d\mu$$

Note that

$$\mathbf{1}_\Omega \in \mathcal{C}_b(\Omega) \Rightarrow \mu_n \to \mu \Rightarrow \mu_n(\Omega) \to \mu(\Omega) \in \mathbb{R} \Rightarrow \sup_n \mu_n(\Omega) = M < \infty.$$

Definition 3.2. *If $X_n, X : \Omega \to \mathbb{R}$ on some probability space $(\Omega, \mathcal{F}, P)$, then $X_n \xrightarrow{w} X$ if and only if $\mu_{X_n} \xrightarrow{w} \mu_X$*

Example 3.3. Why is this not stronger? i.e. $\mu_n(A) \to \mu(A)$ for example? Let $X \sim F$ and let $X_n := X + \frac{1}{n}$. Then $X_n \searrow X$ a.s., so that $X_n \xrightarrow{w} X$. BUT,

$$F_n(x) = P\left[X + \frac{1}{n} \leq x\right] = F\left(x - \frac{1}{n}\right).$$

Hence $\lim_n F_n(x) = F(x^-)$, so that $F_n(x) \to F(x) \iff x$ is a continuity point of F. Hence, we shouldn't expect that $\mu_n(A) \to \mu(A) \quad \forall A \in \mathcal{F}$.

Note that if $\mu \sim F$ and $\mathcal{C}_F = \{x \mid F \text{ is cts. at } x\}$, then $\mathcal{C}_F$ is dense ($\mathcal{C}_F^c$ is countable). Hence, $\mu(F)$ is uniquely determined by its values at $\mathcal{C}_F$ (since it is right continuous).

Example 3.4. Let (X_i) be i.i.d $\mathcal{N}(0,1)$. Let $S_n = \sum_{i=1}^n X_i$. Then $S_n \sim \mathcal{N}(0,n)$. Let

$$\mu_n := \frac{1}{n} S_n \sim \mathcal{N}\left(0, \frac{1}{n}\right)$$

Then "obviously" $\mu_n \xrightarrow[n\to\infty]{} \delta_0 =: \mu$ in some sense. BUT,

$$0 = \mu_n(\{0\}) \not\to \mu(\{0\}) = 1.$$

Theorem 3.5 (Portmanteau). *Let $(\mu_n)_n, \mu \in \mathfrak{M}_{finite}$ be such that $\lim \mu_n(\Omega) = \mu(\Omega)$. Then TFAE:*

1. $\mu_n \xrightarrow{w} \mu$
2. $\forall \varphi \in \mathcal{C}_{b,L}(\Omega)$ *(bounded Lipschitz cts),* $\int \varphi \, d\mu_n \to \int \varphi \, d\mu$
3. $\forall G$ *open,* $\liminf_n \mu_n(G) \geq \mu(G)$
4. $\forall D$ *closed:* $\limsup_n \mu_n(D) \leq \mu(D)$
5. $\forall A \in \mathcal{F}$ *such that* $\mu(\partial A) = 0$, $\lim_n \mu_n(A) = \mu(A)$
6. $\forall \varphi \in \mathcal{F}_b$ *such that* $\mu(\mathcal{D}_\varphi) = 0$, $\lim_n \int \varphi \, d\mu_n = \int \varphi \, d\mu$ *where $\mathcal{D}_\varphi$ is any set containing all of the discontinuities of φ.*

Proof.

- $(1 \Rightarrow 2)$ Trivial.
- $(2 \Rightarrow 3)$ Define $\operatorname{dist}(y, D) := \inf\{\rho(x,y) : x \in D\}$. Then for $r \geq 0$, let

  $$f_k(r) := (1 - kr)^+$$

  and

  $$\varphi_k(x) := f_k(\operatorname{dist}(x, D))$$

  for some closed set D. Observe that φ_k is clearly Lipschitz with $\varphi_k \geq \mathbf{1}_D$ and, in fact, $\varphi_k \searrow \mathbf{1}_D$ as $k \to \infty$. Thus,

  $$\limsup_{n\to\infty} \mu_n(0) \leq \liminf_{k\to\infty} \underbrace{\limsup_{n\to\infty} \int \varphi_k \, d\mu_n}_{=\int \varphi_k \, d\mu} = \int \varphi \, d\mu = \mu(D)$$

- $(3 \Rightarrow 4)$ Let $G = D^c$ open. Then

  $$\liminf_{n\to\infty} \mu_n(G) = \liminf_{n\to\infty} (\mu_n(\Omega) - \mu_n(D)) = \overbrace{\lim_{n\to\infty} \mu_n(\Omega)}^{=\mu(\Omega)} - \limsup_{n\to\infty} \mu_n(D)$$
  $$\geq \mu(\Omega) - \mu(D) = \mu(D^c) = \mu(G)$$

- (4 ⇒ 3) Analogous to (3 ⇒ 4).
- (3 ⇒ 5) or (4 ⇒ 5) Observe that

$$\mu(A) = \mu(A^\circ) \leq \liminf_{n\to\infty} \mu_n(A^\circ) \leq \liminf_{m\to\infty} \mu_n(A)$$
$$\leq \limsup_{n\to\infty} \mu_n(A) \leq \limsup_{n\to\infty} \mu_n(\bar{A}) \leq \mu(\bar{A}) = \mu(A)$$

  and so everything in the line above is equal.
- (5 ⇒ 6) We apply MTI, but with a careful approximation. First, note that the distribution of $\varphi \in \mathcal{F}_b$ (bounded measurable functions) on $\mathbb{R}$ can have (at most) countably many atoms, i.e. the set

$$\mathcal{A} = \{a \in \mathbb{R} : \mu(\varphi = a) > 0\}$$

  is (at most) countable. This means $\mathcal{A}^c$ is dense and, therefore, $\forall n \geq 1$ we can find points

$$\alpha_1 < \alpha_2 < \cdots < \alpha_\ell \text{ such that } |\alpha_i - \alpha_{i+1}| < \frac{1}{k}$$

  and $\alpha_1 \leq -k \leq k < \alpha_\ell$ where $k = \sup|\varphi| < \infty$. Set

$$\varphi_k = \sum_{i=1}^{\ell} \alpha_i \mathbf{1}_{\{\alpha_{i-1} < \varphi \leq \alpha_i\}} =: \sum_{i=1}^{\ell} \alpha_i \mathbf{1}_{A_i}$$

  Notice that

$$\partial A_i \subseteq \{\varphi \in \{\alpha_{i-1}, \alpha_i\}\} \cup D_\varphi$$

  where D_φ is any set $\in \mathcal{F}$ containing all the discontinuity points of φ. By the assumptions on D_φ and the choice of α_i, we have

$$\mu(\partial A_i) = 0 \;\forall i \;\Rightarrow\; \lim_{n\to\infty} \mu_n(A_i) = \mu(A_i)$$

  Finally, since $\varphi_k \searrow \varphi$ we can apply dominated convergence, and so

$$\left|\int \varphi\, d\mu_n - \int \varphi\, d\mu\right| \leq \left|\int \varphi\, d\mu_n - \int \varphi_k\, d\mu_n\right| + \left|\int \varphi_k\, d\mu_n - \int \varphi_k\, d\mu\right|$$
$$+ \left|\int \varphi_k\, d\mu - \int \varphi\, d\mu\right|$$
$$\leq \frac{1}{k} \sup_n \mu_n(\Omega) + (\to 0 \text{ by } (5)\ ) + \frac{1}{k} M$$

  Specifically, given any $\varepsilon > 0$, choose k such that $\frac{M}{k} < \frac{\varepsilon}{2}$ and, given k, choose N big enough such that the middle term is $\leq \frac{\varepsilon}{2}$. Then, $\forall \varepsilon > 0, \exists n$ st $\forall n \geq N$, $|\int \varphi\, d\mu_n - \int \varphi\, d\mu| < \varepsilon$.

□

Special case: $\Omega = \mathbb{R}$ or $= \mathbb{R}^d$. Let $\mu_n, \mu \in \mathfrak{M}_1(\mathbb{R})$ such that $F_n(x) \to F(x)$ as $n \to \infty$.

Theorem 3.6. *For all x, if F is continuous at x then $\exists (X_n)_n, X$ RVs on $((0,1), \mathcal{B}, \lambda)$ such that $X_n \to X$ a.s. and $X_n \sim \mu_n, X \sim \mu$.*

Remark 3.7. This is a special case of the following: Let (Ω, ρ) be a separable, complete metric space and let $(\mu_n)_n, \mu \in \mathfrak{M}_1(\mathbb{R})$. If $\mu_n \xrightarrow{w} \mu$ then $\exists (X_n), X$ on some probability space with $X_n \sim \mu_n$ such that $X_n \to X$ a.s.

Proof. If F is a distribution function (i.e. increasing, right-continous, $F(-\infty) = 0$ and $F(\infty) = 1$), then $\forall x \in (0,1)$ we set

$$a_x := \sup \{y : F(y) < x\} =: F^{-1}(x)$$
$$b_x := \int \{y : F(y) > x\}$$

to be the left-continuous inverse and right-continuous inverse, respectively. Notice that $\forall x$, $a_x \le b_x$ with strict inequality $\iff$ F is locally constant in (a_x, b_x). Also, $x < x' \Rightarrow b_x \le a_{x'}$, and $\exists$ (at most) countably many points x such that $a_x < b_x$.

Given $(F_n), F$ we set $Y_n(x) := F_n^{-1}(x)$ and $Y(x) = F^{-1}(x)$ on $\Omega = (0,1)$ with $\mathcal{B}$ and dx. We claim: $Y_n \sim dF_n$ and $Y \sim dF$. To see why, we check

$$\{Y \le x\} = \{y : Y(y) \le x\} = \{y : a_y \le x\} = \{y : y \le F(x)\}$$

and so $\lambda(Y \le x) = \lambda((0, F(x)]) = F(x)$.

Assume now that $F_n(y) \to F(y)$ whenever F is continuous at y. We will show that if x is such that $a_x = b_x$ then $F_n^{-1}(x) \to F^{-1}(x)$ which proves the theorem (since there are at most countable many such points). We make the following two claims

$$\liminf_{n\to\infty} F_n^{-1}(x) \ge F^{-1}(x)$$
$$\limsup_{n\to\infty} F_n^{-1}(x) \le F^{-1}(x)$$

To prove the first inequality, let $y < F^{-1}(x)$ and assume F is continuous at y. Notice $F(y) < x$ necessarily and so $\forall n$ large enough, $F_n(y) < x$, which implies

$$\sup \{z : F_n(z) < x\} = F_n^{-1}(x) \ge y$$

Then,since the continuity points of F are dense we have

$$\liminf_{n\to\infty} F_n^{-1}(x) \ge \sup \left\{y : y < F^{-1}(x), F \text{ cts at } y\right\} = F^{-1}(x)$$

Proving the other claim is similar (see Durrett p. 84). ☐

Theorem 3.8. *On $\mathbb{R}$ (or $\mathbb{R}^d$),*

$$\mu_n \xrightarrow{w} \mu \iff F_n(x) \xrightarrow[n\to\infty]{} F(x) \quad \forall x : F \text{ is cts at } x$$

Proof. ($\Rightarrow$) If F is continuous at x, then

$$\begin{aligned} &\mu\left(\partial(-\infty, x]\right) = \mu(\{x\}) = 0 \\ &\Rightarrow \lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} \mu_n\left((-\infty, x]\right) = \mu\left((-\infty, x]\right) = F(x) \end{aligned}$$

The same argument works in $\mathbb{R}^d$: if F is continuous at $\vec{x}$, then

$$\mu\left(\partial\left((-\infty, x_1] \times \cdots \times (-\infty, x_d]\right)\right) = 0$$

and so on.

($\Leftarrow$) If $F_n(x) \to F(x)$ whenever F is continuous at x then we can apply (2) from Theorem 3.5 to say $\exists X_n \sim dF_n = \mu_n$ such that $X_n \to X \sim dF = \mu$ a.s. For $\varphi \in \mathbb{C}_n(\mathbb{R})$, we have, by dominated convergence,

$$\int \varphi\, d\mu_n = E\left[\varphi(X_n)\right] \xrightarrow[n\to\infty]{} E\left[\varphi(X)\right] = \int \varphi\, d\mu$$

Note: a direct proof for $\mathbb{R}^d$ is given in Durrett p. 165. □

Remark 3.9. If $X_n \xrightarrow{w} X$ and $g \in C_b^+$ then $\liminf E[g(X_n)] \geq E[g(X)]$, etc. ($\exists Y_n \to Y$ a.s. with $Y_n \sim X_n$ and $Y \sim X$, apply Fatou)

This is related to other notions of convergence, as the next theorem demonstrates.

Theorem 3.10. *$X_n \to X$ in probability $\Rightarrow X_n \xrightarrow{w} X$.*

Notice that $X_n \to X$ a.s. $\Rightarrow X_n \xrightarrow{w} X$ immediately, by the definition.

Proof. Define the sets

$$A_{\varepsilon,n} := \{|X_n - X| \leq \varepsilon\} \;\Rightarrow\; \lim_{n\to\infty} P\left[A_{\varepsilon,n}^c\right] = 0\ \forall\varepsilon$$

WWTS $F_n(x) \to F(x)$ if x is a continuity point of F. First, we see that

$$\begin{aligned} F_n(x) = P\left[X_n \leq x\right] &\leq P\left[\{X_n \leq x\} \cap A_{\varepsilon,n}\right] + P\left[A_{\varepsilon,n}^c\right] \\ &\leq P\left[A_{\varepsilon,n}^c\right] + P\left[X \leq x + \varepsilon\right] \end{aligned}$$

and so

$$\forall\varepsilon : \limsup_{n\to\infty} F_n(x) \leq 0 + F(x+\varepsilon) \;\Rightarrow\; \limsup_{n\to\infty} F_n(x) \leq F(x^+) = F(x)$$

Similarly,

$$F_n(x) = P\left[X_n \leq x\right] \geq P\left[\{X_n \leq x\} \cap A_{\varepsilon,n}\right] \geq P\left[X \leq x - \varepsilon\right] - P\left[A_{\varepsilon,n}^c\right]$$

and so

$$\forall\varepsilon : \liminf_{n\to\infty} F_n(x) \geq F(x-\varepsilon) \;\Rightarrow\; \liminf_{n\to\infty} F_n(x) \geq F(x^-) = F(x)$$

Combining these, we have $\liminf F_n(x) \geq F(x) \geq \limsup F_n(x)$ so it must be that $\lim F_n(x) = F(x)$. □

Theorem 3.11. *In* $\mathbb{R}^d$,

$$\mu_n \xrightarrow{w} \mu \iff \int \varphi \, d\mu_n \to \int \varphi \, d\mu \ \forall \varphi \in C_c(\mathbb{R})$$

where $C_c(\mathbb{R})$ *denotes functions that are continuous with compact support.*

Proof. (⇒) Exercise (***).
(⇐) Let G be open and set $G_k = G \cap (-k,k)^d$ which is open and bounded. Set

$$\varphi_k(x) = 1 \wedge k \underbrace{\operatorname{dist}(x, G_k^c)}_{\in C(\mathbb{R}^d)} \in C(\mathbb{R}^d)$$

Note $\varphi(k) = 0$ on G_k^c and > 0 on G_k and $\varphi_k \nearrow \mathbf{1}_G$ as $k \to \infty$, so $\varphi_k \in C_c(\mathbb{R}^d)$. Let $\mathcal{L}^1 \ni \varphi \geq \varphi_k \geq 0$ with $\varphi_k \in C_c$ and $\varphi_k \to \varphi$ as $k \to \infty$. Then

$$\liminf_{n\to\infty} \int \varphi \, d\mu_n \geq \limsup_{k\to\infty} \underbrace{\liminf_{n\to\infty} \int \varphi_k \, d\mu_n}_{= \int \varphi_k \, d\mu \ \forall k} = \int \varphi \, d\mu$$

by assumption and by dominated convergence. So for G open set $\varphi = \mathbf{1}_G$ and define φ_k as before. Then

$$\liminf_{n\to\infty} \mu_n(G) \geq \mu(G)$$

and by the Portmanteau Theorem 3.5, this shows $\mu_n \xrightarrow{w} \mu$. □

3.1 Fourier Transforms of Probability Measures

Let $\mu \in \mathfrak{M}_f(\mathbb{R})$. For instance, $X \sim \mu$ for some RV on $(\Omega, \mathcal{F}, P)$. For $t \in \mathbb{R}$, define

$$\begin{aligned} \hat{\mu}(t) :&= \int_{\mathbb{R}} \exp(itx) \, \mu(dx) = E\left[\exp(itX(\omega))\right] \\ &= \int_{\mathbb{R}} \cos(tx) \, \mu(dx) + i \int_{\mathbb{R}} \sin(tx) \, \mu(dx) \end{aligned} \tag{4}$$

This is called the *Fourier Transform* of μ (or of X).

Lemma 3.12. *For* $\mu \in \mathfrak{M}_f(\mathbb{R})$, *the function* $\hat{\mu}(t)$ *exists; in fact,* $|\hat{\mu}(t)| \leq \mu(\mathbb{R})$. *Furthermore,*

1. $\hat{\mu}$ *is uniformly continuous*
2. $\sup_t |\hat{\mu}(t)| = \hat{\mu}(0) = \mu(\mathbb{R})$ *and* $\hat{\mu}(-t) = \overline{\hat{\mu}(t)}$.
3. $\hat{\mu}$ *is a positive-definite function, i.e.*

$$\sum_{i,j} \hat{\mu}(t_i - t_j) z_i \overline{z}_j \geq 0 \quad \forall \vec{t} \in \mathbb{R}^n, \vec{z} \in \mathbb{C}^n$$

Proof. To show existence of $\hat{\mu}$, notice that

$$|E[\exp(itX)]| \leq E[|\exp(itX)|] \leq 1$$

for finite measures. For uniform continuity, we use the identity

$$|\exp(ia) - \exp(ib)| = |1 - \exp(i(b-a))| \, |\exp(ia)| \leq || \wedge 2$$

which implies

$$\sup_{|s-t|\leq\delta} |\hat{\mu}(t) - \hat{\mu}(s)| \leq \sup_{|s-t|\leq\delta} E[\exp(itX) - \exp(isX)] \leq E[\delta X \wedge 2] \xrightarrow[\delta\to 0]{} 0$$

Proving the sup and complex conjugate conditions are trivial and left as exercises (***). To prove positive-definiteness, we use the fact that $|z|^2 = z\bar{z}$ and observe that

$$\begin{aligned} E\left[|\sum_{j=1}^{n} \exp(it_jX)\cdot z_j|^2 \right] &= \sum_{j,k} E[\exp(it_jX)\cdot z_j \exp(-it_kX)\cdot \bar{z}_k] \\ &= \sum_{j,k} E[\exp(i(t_j - t_k)X z_j \bar{z}_k] \\ &= \sum_{j,k} \hat{\mu}(t_i - t_j) z_i \bar{z}_k \geq 0 \end{aligned}$$

since $E[\cdot^2] \geq 0$. □

Definition 3.13. *For $f \in \mathcal{L}^1(dx)$ we define*

$$\hat{f}(t) = \int \exp(itx) f(x)\, dx = \hat{\nu}(t)$$

where $d\nu = f\, dx$ is a finite signed measure.

Notice that $|\hat{f}(t)| \leq \int |\exp(itx)||f|\, dx = \|f\|_1 < \infty$.

Definition 3.14. *Let $f, g \in \mathcal{L}^1$. Then*

$$f \star g(x) := \int_{\mathbb{R}} f(x-y) g(y)\, dy$$

is called the convolution *of f and g, whenever the integral exists and is finite $\forall x$.*

Remark 3.15. Notice $f \star g = g \star f$ since

$$\begin{aligned} \int \underbrace{f(x-y)}_{:=\tilde{f}_x(y)} g(y)\, dy &= \int \tilde{f}_x(y) g(y)\, dy = \int \tilde{f}_x(-y) g(-y)\, dy \\ &= \int \tilde{f}_x(x-y) g(x-y)\, dy = \int f(y) g(x-y)\, dy \end{aligned}$$

where shifting by x in the second line does not alter the integral.

In the future, we will use the Gaussian function

$$\varphi_\varepsilon(x) := \frac{1}{\varepsilon\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\varepsilon^2}\right) \tag{5}$$

which is $\sim \mathcal{N}(0, \varepsilon^2)$.

Theorem 3.16 (Fejer)**.** *Let $f \in C_c(\mathbb{R})$ and φ_ε as in (5). Then*

$$\lim_{\varepsilon \to 0} \sup_{x \in \mathbb{R}} |(f \star \varphi_\varepsilon)(x) - f(x)| = 0$$

Proof. Let $Z \sim \mathcal{N}(0,1)$ so $\varepsilon Z \sim \mathcal{N}(0, \varepsilon^2)$ and then

$$E\left[f(x - \varepsilon Z)\right] = \int_{\mathbb{R}} f(x-y)\varphi_\varepsilon(y)\, dy = f \star \varphi_\varepsilon(x)$$

Note that $|f| \le M < \infty$ and for a.e. ω,

$$f\left(x - \varepsilon Z(\omega)\right) \xrightarrow[\varepsilon \to 0]{} f(x)$$

Moreover, since f is uniformly continuous, in particular,

$$\sup_x |f(x - \varepsilon Z(\omega) - f(x)| =: \overbrace{W(\omega, \varepsilon)}^{\le 2M} \xrightarrow[\varepsilon \to 0]{} 0 \text{ for a.e.} \omega$$

and therefore

$$\sup_x |\overbrace{f \star \varphi_\varepsilon(x)}^{E[f(x-\varepsilon Z]} - \overbrace{f(x)}^{E[f(x-0Z)]} | \le \sup_x E\big[|\overbrace{f(x - \varepsilon Z(\omega) - f(x)}^{\le W_\varepsilon(\omega)\ \forall x}|\big] \le E\left[W_\varepsilon(\omega)\right] \to 0$$

by dominated convergence. ☐

Theorem 3.17 (Planchard)**.** *1. Let $f \in \mathcal{L}^1(dx)$ and $\varepsilon > 0$. Then*

$$\int_{\mathbb{R}} (f \star \varphi_\varepsilon)(x)\mu(dx) = \frac{1}{2\pi} \int_{\mathbb{R}} \exp\left(-\frac{\varepsilon^2 t^2}{2}\right) \hat{f}(t)\overline{\hat{\mu}(t)}\, dt$$

2. Let $f \in C_c(\mathbb{R})$ and $\hat{f} \in \mathcal{L}^1(dx)$. Then

$$\int_{\mathbb{R}} f\, d\mu = \frac{1}{2\pi} \int_{\mathbb{R}} \hat{f}(t)\overline{\hat{\mu}(t)}\, dt$$

Proof. 1. Notice that

$$\frac{1}{2\pi}\int \exp\left(-\frac{\varepsilon^2t^2}{2}\right)\hat{f}(t)\overline{\hat{\mu}(t)}\,dt$$

$$= \frac{1}{2\pi}\int \exp\left(-\frac{\varepsilon^2t^2}{2}\right)\cdot\left(\int \exp(itx)f(x)\,dx\right)\cdot\left(\exp(-ity)\,\mu(dy)\right)\,dt$$

$$= \frac{1}{2\pi}\int \big(f(x)\underbrace{\int \exp(it(x-y))\exp\left(-\frac{\varepsilon^2t^2}{2}\right)dt}_{=\frac{\sqrt{2\pi}}{\varepsilon}\exp\left(-\frac{(x-y)^2}{2\varepsilon^2}\right)=2\pi\varphi_\varepsilon(y-x)}\,dx\big)\,\mu(dy)$$

$$= \int_{\mathbb{R}}(f\star\varphi_\varepsilon)(y)\,\mu(dy)$$

2. For all $\varepsilon > 0$, we have

$$\int f_\varepsilon(x)\,\mu(dx) = \frac{1}{2\pi}\int \exp\left(-\frac{\varepsilon^2t^2}{2}\right)\hat{f}(t)\overline{\hat{\mu}}(t)\,dt$$

Since $f_\varepsilon(x) \to f(x)$ uniformly, the LHS $\to \int f(x)\,\mu(dx)$ as $\varepsilon \to 0$. Likewise, on the RHS, the $\exp(\cdot)$ term $\to 1$, so by dominated convergence, the whole RHS $\to \frac{1}{2\pi}\int 1\cdot\hat{f}(t)\overline{\hat{\mu}(t)}\,dt$.

□

Theorem 3.18 (Uniqueness)**.** *If $\mu,\nu \in \mathfrak{M}_1(\mathbb{R})$ such that $\hat{\mu} = \hat{\nu}$ (λ-a.e.) then $\mu = \nu$.*

Proof. Let $f \in C_c(\mathbb{R})$ and set $f_\varepsilon = f\star\varphi_\varepsilon$. Apply Planchard's Theorem 3.17 part (1) to write

$$\int f_\varepsilon\,d\mu = \int f_\varepsilon\,d\nu \quad \forall\varepsilon > 0$$

Since $f_\varepsilon \to f$ uniformly as $\varepsilon \to 0$, then letting $\varepsilon \to 0$ shows $\int f\,d\mu = \int f\,d\nu$. Since this holds for arbitrary such f, it must be that $\mu = \nu$. Specifically, for any $-\infty < a < b < \infty$, notice that $f_\varepsilon \searrow \mathbf{1}_{[a,b]}$ and so $\int f_\varepsilon\,d\mu \searrow \mu[a,b]$ and similarly for $\nu[a,b]$, as well. □

Theorem 3.19 (Pleny & Glivenko)**.** *Let $\mu,(\mu_n)_n \in \mathfrak{M}_1(\mathbb{R})$. Suppose $\hat{\mu}_n(t) \to \hat{\mu}(t)$ for λ-a.e. t. Then $\mu_n \xrightarrow{w} \mu$.*

Remark 3.20. Notice that the converse is trivial since $\exp(itx) \in \mathbb{C}_b(\mathbb{R})$, so we can apply the Transformation Formula 1.49 and say $E[\varphi(X_n)] = \int \varphi\,d\mu_n \to \int \varphi\,d\mu = E[\varphi(X)]$. Also, the same theorem holds on $\mathbb{R}^d$ (see Theorem 9.4 in Durrett).

Proof. Let $f \in C_c(\mathbb{R})$. WWTS $\int f\,d\mu_n \to \int f\,d\mu$. Set

$$\delta(\varepsilon) := \|f_\varepsilon - f\|_\infty = \|f\star\varphi_\varepsilon - f\|_\infty$$

so that $\delta(\varepsilon) \to 0$ as $\varepsilon \to -0$. We first apply the triangle inequality (thrice) and then Planchard's Theorem 3.17 part (1) to write, $\forall \varepsilon > 0$,

$$\begin{aligned}
&\left|\int f \, d\mu_n - \int f \, d\mu\right| \\
&\quad \leq \int \overbrace{|f_\varepsilon - f| \, d\mu_n}^{\leq \delta(\varepsilon)\ \forall n} + \int \overbrace{|f_\varepsilon - f| \, d\mu}^{\leq \delta(\varepsilon)} + \left|\int f_\varepsilon \, d\mu_n - \int f_\varepsilon \, d\mu\right| \\
&\quad = \int \cdot + \int \cdot + \frac{1}{2\pi}\left|\underbrace{\int \exp\left(-\frac{\varepsilon^2 t^2}{2}\right) \underbrace{\hat{f}(t)}_{|\cdot| \leq \|f\|_1 < +\infty}}_{\text{a finite measure on } \mathbb{R}\ \forall \varepsilon > 0} \ (\underbrace{\overline{\hat{\mu}_n(t)} - \overline{\hat{\mu}(t)}}_{\to 0 \text{ a.e. as } n \to \infty}) \, dt\right|
\end{aligned}$$

This means we can apply Dominated Convergence to say

$$\limsup_{n\to\infty} \left|\int f \, d\mu_n - \int f \, d\mu\right| \leq 2\delta(\varepsilon) + 0 \xrightarrow[\varepsilon \to 0]{} 0$$

□

Remark 3.21. Let $\mu_n, \mu \in \mathfrak{M}_1(\mathbb{R}^d)$. Then

$$\mu_n \to \mu \iff \hat{\mu}_n(t) \to \hat{\mu}(t) \quad \forall t \in \mathbb{R}$$

Corollary 3.22 (Cramer-Wold Device). *Let $(\vec{X}_n)_{n\geq 1}, \vec{X}$ be RVs with values in $\mathbb{R}^d$. If $(\vec{t} \cdot \vec{X}_n) \xrightarrow{w} (\vec{t} \cdot \vec{X})$ for every $t \in \mathbb{R}^d$, then $\vec{X}_n \xrightarrow{w} \vec{X}$.*

Proof. Since $\exp(ix) \in C_b(\mathbb{R})$, we know

$$E\left[\exp(i(\vec{t} \cdot \vec{X}_n))\right] \xrightarrow[n\to\infty]{} E\left[\exp(i(\vec{t} \cdot \vec{X}))\right]$$

for every $\vec{t} \in \mathbb{R}^d$. But notice that this is just the pointwise convergence of the Fourier Transforms of $\vec{X}_n \to \vec{X}$! □

4 Central Limit Theorems and Poisson Distributions

Theorem 4.1 (CLT in $\mathbb{R}$). *Suppose $(X_n)_n$ are i.i.d. with finite second moments. Let $\mu = E[X_1]$ and $\sigma^2 = \mathrm{Var}[X_1]$. Then*

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{w} \mathcal{N}(0,1)$$

Proof. WOLOG $\mu = 0$ and $\sigma = 1$, so WWTS $\frac{S_n}{\sqrt{n}} \to \mathcal{N}(0,1)$. That is, WWTS

$$\lim_{n\to\infty} E\left[\exp\left(it\frac{S_n}{\sqrt{n}}\right)\right] = \exp\left(-\frac{t^2}{2}\right) \quad \forall t$$

We look at the Taylor expansion for $x \in \mathbb{R}$ to write

$$\exp(ix) = 1 + ix - \frac{x^2}{2} + R(x) \quad , \quad |R(x)| \le \frac{1}{6}|x^3| \le |x|^3$$

for $x \in \mathbb{R}$ where $R(x)$ is the remainder term. For *large* x, we will use the estimate

$$|R(x)| \le |\exp(ix)| + 1 + |x| + \frac{x^2}{2} \le 2 + |x| + \frac{|x|^2}{2} \le x^2$$

for $x \ge 4$, and so

$$|R(x)| \le |x|^3 \wedge 4|x|^2 \quad \forall x \in \mathbb{R}$$

since $|x|^3 \le 4|x|^2$ for $|x| \le 4$. Now, we write

$$\begin{aligned}
E\left[\exp\left(it\frac{S_n}{\sqrt{n}}\right)\right] &= \prod_{k=1}^{n} E\left[\exp\left(it\frac{X_k}{\sqrt{n}}\right)\right] \\
&= \left(1 + itE\left[\frac{X_1}{\sqrt{n}}\right] - \frac{t^2}{2n}E\left[X_1^2\right] + E\left[R\left(\frac{tX_1}{\sqrt{n}}\right)\right]\right)^n \\
&= \left(1 - \frac{1}{n}\left(\frac{t^2}{2} - nE\left[R\left(\frac{tX_1}{\sqrt{n}}\right)\right]\right)\right)^n \\
&=: \left(1 - \frac{1}{n}\left(\frac{t^2}{2} - \varepsilon_n\right)\right)^n
\end{aligned}$$

and observe that

$$\begin{aligned}
|\varepsilon_n| &\le nE\left[\left(\frac{tX_1}{\sqrt{n}}\right)^3 \wedge 4\left(\frac{tX_1}{\sqrt{n}}\right)^2\right] = E\left[\frac{t^3X_1^3}{\sqrt{n}} \wedge 4t^2X_1^2\right] \\
&= t^2E\left[\frac{t}{\sqrt{n}}X_1^3 \wedge 4X_1^2\right] \xrightarrow[n\to\infty]{} 0
\end{aligned}$$

where convergence in the last line follows by dominated convergence, since $|\cdot| \le 4X_1^2 \in \mathcal{L}^1$ (note: this shows why only finite *second* moment needed!) Thus,

$$\lim_{n\to\infty} E\left[\exp\left(it\frac{S_n}{\sqrt{n}}\right)\right] = \lim_{n\to\infty}\left(1 - \frac{1}{n}\left(\frac{t^2}{2} - \varepsilon_n\right)\right)^n = \exp\left(-\frac{t^2}{2}\right)$$

which is implied by the following claim:

$$\lim_{n\to\infty}\left(1 - \frac{c_n}{n}\right)^n = \exp(-c) \quad \text{if } \mathbb{C} \ni c_n \to c$$

To prove this, we use the complex logarithmic function and write

$$\text{RHS} = \lim_{n\to\infty} \exp\left(n \cdot \log\left(1 - \frac{c_n}{n}\right)\right) = -\frac{c_n}{n} + o\left(\frac{1}{n}\right)$$

using the Taylor power series for log. Then

$$\begin{aligned}\text{RHS} &= \lim_{n\to\infty} \exp\left(n \cdot \left(-\frac{c_n}{n} + o\left(\frac{1}{n}\right)\right)\right) \\ &= \lim_{n\to\infty} \exp(-c_n) \cdot \exp\left(n \cdot o\left(\frac{1}{n}\right)\right) = \exp(-c)\end{aligned}$$

□

For completeness, we also present an alternative (direct) proof that

$$\lim_{z_n\to z} \left(1 - \frac{z_n}{n}\right)^n = \mathrm{e}^z$$

Set $a_n = \left(1 + \frac{z_n}{n}\right)$ and $b_n = \exp(z_n/n)$, and choose $|\gamma| > |z|$. For large n, $|z_n| < \gamma$, so $\frac{|z_n|}{n} \leq 1$, which implies

$$\left|\left(1 - \frac{z_n}{n}\right)^n - \mathrm{e}^{z_n}\right| \leq (\exp(z_n/n))^{n-1} \cdot n \cdot \left|\frac{z_n}{n}\right|^2 \leq \mathrm{e}^\gamma \cdot \frac{\gamma^2}{n} \xrightarrow[n\to\infty]{} 0$$

where the first inequality follows from Lemma AB on page 3 in Unit 11 ******** reference ****** Therefore,

$$\left|\left(1 - \frac{z_n}{n}\right)^n - \mathrm{e}^z\right| \leq \left|\left(1 - \frac{z_n}{n}\right)^n - \mathrm{e}^{z_n}\right| + |\mathrm{e}^{z_n} - \mathrm{e}^z| \to 0$$

Theorem 4.2 (Lindeberg-Feller)**.** *For any n, let $X_{n,1}, X_{n,2}, \ldots, X_{n,k_n}$ be RVs on the probability space $(\Omega_n, \mathcal{F}_n, P_n)$. Assume that*

1. *For each n, $(X_{n,k})_{k=1,\ldots,k_n}$ are independent and have 0 mean, with respect to P_n*
2. *$\sum_{k=1}^{k_n} \operatorname{Var}(X_{n,k}) \to \sigma_2 \in (0,\infty)$ as $n \to \infty$*
3. *For every $\varepsilon > 0$,*

$$\lim_{n\to\infty} \sum_{k=1}^{k_n} E_n\left[X_{n,k}^2; |X_{n,k}| > \varepsilon\right] = 0$$

Then

$$S_n := \sum_{k=1}^{k_n} X_{n,k} \xrightarrow{w} \mathcal{N}(0,\sigma^2)$$

i.e. $\mu_n := P_n \circ S_n^{-1} \xrightarrow{w} \mathcal{N}(0,\sigma^2)$.

Proof. See Durrett. It is not much more complicated than the proof of the classical CLT 4.1. □

Observe that Lindeberg-Feller 4.2 includes the classical CLT 4.1. Set $k_n = n$ and $X_{n,k} = \frac{X_k}{\sqrt{n}}$ where X_k are i.i.d. with $E[X_k] = 0$ and $E[X_k^2] = \sigma^2$. Then assumption (1) is satisfied for all n by definition, assumption (2) holds because

$$\sum_{k=1}^{n} E[X_{n,k}^2] = \sum_{k=1}^{n} \left(\frac{1}{\sqrt{n}}\right)^2 E[X_k^2] = \sigma^2$$

and assumption (3) holds because

$$\begin{aligned} E\left[X_{n,k}^2 \cdot \mathbf{1}_{\{|X_{n,k}|>\varepsilon\}}\right] &= \sum_{k=1}^{n} E\left[\frac{1}{n} X_k^2 \cdot \mathbf{1}_{\{|X_k|>\varepsilon\sqrt{n}\}}\right] \\ &= E\left[X_1^2 \cdot \mathbf{1}_{\{X_1^2>\varepsilon^2 n\}}\right] \xrightarrow[n\to\infty]{} 0 \end{aligned}$$

by dominated convergence, since $X_1^2 \cdot \mathbf{1}_{\{X_1^2>\varepsilon^2 n\}} \to 0$ as $n \to \infty$ a.s. dy dominated convergence, since $X_1^2 \in \mathcal{L}^1$.

Example 4.3 (Cycles in a random permutation). Let $\Omega_n = \{\omega : \{1, \dots, n\} \to \{1, \dots, n\} \text{ bijective}\}$ be the space of permutations of $\{1, \dots, n\}$, where $k \mapsto \omega_k$. We write $\Pi = \Pi(\omega) = (\Pi_1(\omega), \dots, \Pi_n(\omega)) = (\omega_1, \dots, \omega_n)$. Let P_n be the uniform measure on Ω_n, i.e. $P_n[\Pi = \sigma] = \frac{1}{n!}$ for any fixed permutation σ.

Notation: as an example, consider the permutation

$$(1, 2, 3, 4, 5, 6, 7, 8) \mapsto (2, 5, 8, 4, 1, 7, 3, 6) = (\sigma_1, \sigma_2, \dots, \sigma_8) = \sigma$$

We write $\sigma = (125)(3867)(4) = C_1C_2C_3 = c_1c_2 \dots c_8$ in its *cycle decomposition* form, where the first term is the cycle containing 1, the second term is the cycle containing the lowest number not in the first cycle, etc. **Question**: What is the "typical" number of cycles in the decomposition of a random permutation?

An algorithmic way to "generate" uniformly distributed random permutations is as follows: we generate the cycle decomposition directly beginning with the cycle containing 1, i.e. $C_1 = (1, c_2, ?)$. Let $U_1(\omega)$ be uniformly distributed on $\{1, \dots, n\}$. If $U_1 = 1$ then $C_1 = (1)$ (a fixed point) and set $c_2 = 2$ and continue. If $U_1 \neq 1$ then $c_2 = U_1(\omega)$ and continue. Assuming we already have the first k entries in the form $(c_1c_2 \dots)(\cdots) \dots (c_m \dots c_k)$, what is the next one? Let $U_k(\omega)$ be independent of $U_1, \dots, U_{k-1}$ and uniformly distributed on $\{1, n\} \setminus \{c_1, \dots, c_{m-1}, c_{m+1}, \dots, c_k\}$, so the total number of choices is $n - k + 1$. If $U_k = c_m$ then close the current cycle and begin the next with $c_{k+1} = \min\{\{1 \dots n\} \setminus \{c_1 \dots c_k\}\}$. If $U_k \neq c_m$ set $c_{k+1} = U_k(\omega)$ and proceed. Note:

$$P\left[U_k = c_m \mid \text{ given } U_1, \dots, U_{k-1}\right] = \frac{1}{n - k + 1}$$

is the probability that c_k is the end of the cycle. We introduce the variables $X_{n,k}$, for $k = 1, \dots, n$, that take the value 1 if c_k is the last element of a cycle, and 0 otherwise. For instance, with the length-8 permutation above, $X_{8,3} = X_{8,7} = X_{8,8} = 1$ and all others are 0. Note: $P[X_{n,k} = 1] = \frac{1}{n-k+1}$. More precisely, given the sequence $c_1 \dots c_{k-1} c_k$ and the cycles (i.e. the appropriate brackets), we have

$$P[X_{n,k} = 1 \mid X_{n,1} = x_1, \dots, X_{n,k-1} = x_{k-1}] = \frac{1}{n-k+1}$$

for each $x_1 \dots x_{k-1} \in \{0,1\}$. That is, the $(X_{n,k})$ are independent!

Now, the # of cycles N can be determined by setting

$$\sum_{k=1}^{n} X_{n,k} =: S_n$$

and finding

$$E[S_n] = \sum_{k=1}^{n} \frac{1}{n-k+1} = \sum_{k=1}^{n} \frac{1}{k} = \log n + O(1)$$

and

$$\begin{aligned}
\mathrm{Var}[S_n] = \sum_{k=1}^{n} \mathrm{Var}[X_{n,k}] &= \sum_{k=1}^{n} \left(\frac{1}{n-k+1} - \frac{1}{(n-k+1)^2} \right) \\
&= \sum_{k=1}^{n} \left(\frac{1}{k} - \frac{1}{k^2} \right) \sim \log n + O(1)
\end{aligned}$$

Define

$$Y_{n,k} = \left(X_{n,m} - \frac{1}{n-m+1} \right) / \sqrt{\log n}$$

Then $E[Y_{n,k}] = 0$ and

$$\mathrm{Var}[S_n] = \frac{1}{\log n} \sum_{k=1}^{n} \left(\frac{1}{k} - \frac{1}{k^2} \right) = \frac{1}{\log n} (\log n + O(1)) \to 1$$

To prove assumption (3) from Lindeberg-Feller 4.2 holds, observe that

$$\begin{aligned}
&\sum_{k=1}^{n} E\left[Y_{n,k}^2 ; |Y_{n,k}| > \varepsilon \right] \\
&\quad = \frac{1}{\log n} \sum_{k=1}^{n} E\Big[\left(X_{n,k} - \frac{1}{n-k+1} \right)^2 ; \underbrace{\underbrace{\Big| X_{n,k} - \frac{1}{n-k+1} \Big|}_{<1} > \varepsilon \sqrt{\log n}}_{= \emptyset \text{ if } \log n > \varepsilon^{-2}} \Big] \\
&\qquad \xrightarrow[n \to \infty]{} 0
\end{aligned}$$

since eventually $n > \exp(\varepsilon^{-2})$. Therefore, by Lindeberg-Feller 4.2,

$$\frac{1}{\sqrt{\log n}} \left(S_n - \sum_{k=1}^{n} \frac{1}{k} \right) \xrightarrow{w} \mathcal{N}(0,1)$$

We split this as a sum, to say

$$\frac{S_n - \log n}{\sqrt{\log n}} + \underbrace{\frac{\log n - \sum_{k=1}^n 1/k}{\sqrt{\log n}}}_{\to 0} \xrightarrow{w} \mathcal{N}(0,1)$$

4.1 Poisson Convergence

The "Law of small numbers" could be more aptly titled as the "law of rare events".

Theorem 4.4. *Let A be an array of 0-1 RVs, i.e. $A = (X_{n,m})$ for $n \geq 1$ and $1 \leq m \leq k_n$. Let $P[X_{n,m} = 1] = p_{n,m}$ and set $S_n = X_{n,1} + \cdots + X_{n,k_n}$. Suppose*

1. *$X_{n,1}, \dots X_{n,k_n}$ are independent $\forall n$*
2. $$E[S_n] = \sum_{k=1}^{k_n} p_{n,k} \xrightarrow[n\to\infty]{} \lambda \in [0, \infty)$$

   and

   $$\max_{1 \leq k \leq k_n} p_{n,k} \xrightarrow[n\to\infty]{} 0$$

Then $S_n \xrightarrow{w} Poi(\lambda)$.

Example 4.5. Roll two dice $n = 36$ times. Let $X_{n,k} = 1$ if we get two 6s at time k and 0 otherwise. Then S_n is the count of the number of double 6s, and $E[S_n] = 36 \cdot \frac{1}{6^2} = 1$. This is a rare event with $\lambda = 1$ so $S_n \approx \text{Poi}(1)$. For particular values of k, we can calculate

k :	0	1	2	3
exact :	0.3678	.3678	.1834	.0613
Poi(1) :	0.3627	.3730	.1865	.0604

so we see that the approximation is good even though $n = 36$ is rather small.

Proof. Let

$$\varphi_{n,m}(t) = E\left[\exp\left(itX_{n,m}\right)\right] = (1 - p_{n,m}) + p_{n,m}\exp(it)$$

Then

$$\begin{aligned} E\left[\exp(itS_n)\right] &= E\left[\exp\left(it\sum_{m=1}^n X_{n,m}\right)\right] = \prod_{m=1}^n E\left[\exp(itX_{n,m})\right] \\ &= \prod_{m=1}^n \left(1 + p_{n,m}(\exp(it) - 1)\right) \end{aligned}$$

WWTS, $\forall t \in \mathbb{R}$,

$$E\left[\exp(itS_n)\right] \xrightarrow[n\to\infty]{} \exp\left(\lambda(\exp(it)-1)\right)$$

Note:

$$\begin{aligned}
&\left|\exp\left(\lambda(\exp(it)-1)\right) - E\left[\exp(itS_n)\right]\right| \\
&\qquad \le \left|\exp\left(\lambda(\exp(it)-1)\right) - \exp\left(\sum_{m=1}^{k_n} p_{n,m}(\exp(it)-1)\right)\right| \\
&\qquad\qquad + \left|\exp\left(\sum_{m=1}^{k_n} p_{n,m}(\exp(it)-1)\right) - E\left[\exp(itS_n)\right]\right| =: I_1 + I_2
\end{aligned}$$

Notice $I_1 \xrightarrow[n\to\infty]{} 0$ since $\sum p_{n,m} \to \lambda$. Write

$$I_2 = \left|\prod_{m=1}^{k_n} \exp\left(p_{n,m}(\exp(it)-1)\right) - \prod_{m=1}^{k_n}\left(1 + p_{n,m}(\exp(it)-1)\right)\right| =: |a_m - b_m|$$

and note

$$|a_m| = \exp\left(p_{n,m} - \Re(\exp(it)-1)\right) \le \exp(1\cdot 0) = 1$$

since $|\exp(z)| = \exp(\Re(z))$ and $\Re(\exp(it)-1 \le 0$. Also, $|b_m| \le 1$ since $1 + p_m(\exp(it)-1)$ satisfies ***** picture *****

Applying Lemma 4.6 below tells us

$$\begin{aligned}
I_2 &\le \sum_{m=1}^{k_n} \left|\exp\left(p_{n,m}(\exp(it)-1) - (1 + p_{n,m}(\exp(it)-1)\right)\right)\right| \\
&\le \sum_{m=1}^{k_n} p_{n,m}^2 \left|\exp(it)-1\right|^2
\end{aligned}$$

where the second inequality follows from Lemma 4.7 with $z = p_{n,m}(\exp(it)-1)$, and the fact that $|z| \le 1$ when $\max_m p_{n,m} \le \frac{1}{2}$. Continuing, we have

$$I_2 \le 4 \cdot \underbrace{\max_{1\le m\le k_n} p_{n,m}}_{\to 0} \cdot \underbrace{\sum_{m=1}^{k_n} p_{n,m}}_{\to\lambda} \xrightarrow[n\to\infty]{} 0$$

and this completes the proof. □

The following two lemmas are used in the proof above.

Lemma 4.6. *Let $a_1, \ldots, a_n, b_1, \ldots, b_n \in \mathbb{C}$ such that $|a_i|, |b_i| \le \theta$. Then*

$$\left|\prod_{i=1}^{n} a_i - \prod_{i=1}^{n} b_i\right| \le \theta^{n-1} \sum_{i=1}^{n} |a_i - b_i|$$

Proof. We use induction. The $n = 1$ case is trivial. Now, assume this holds for $k = n - 1$. Then,

$$\begin{aligned}
\left|\prod_{i=1}^{n} a_i - \prod_{i=1}^{n} b_i\right| &\leq \left|a_n \prod_{i=1}^{n-1} a_i - a_n \prod i = 1^{n-1} b_i\right| \\
&\quad + \left|a_n \prod_{i=1}^{n-1} b_i - b_n \prod_{i=1}^{n-1} b_i\right| \\
&\leq |a_n| \cdot \left|\prod_{i=1}^{n-1} a_i - \prod_{i=1}^{n-1} b_i\right| + \left|\prod_{i=1}^{n-1} b_i\right| \cdot |a_n - b_n| \\
&= \theta^{n-1}\left(\left(\sum_{i=1}^{n-1} |a_i - b_i|\right) + |a_n - b_n|\right)
\end{aligned}$$

where the last line follows by the inductive assumption. □

Lemma 4.7. *If $b \in \mathbb{C}$ and $|b| \leq 1$, then*

$$|\exp(b) - (1 + b)| \leq |b|^2$$

Proof. For $|b| \leq 1$, we can write

$$\begin{aligned}
e^b - (1 + b) &= \frac{b^2}{2!} + \frac{b^3}{3!} + \cdots \\
&\leq \frac{|b|^2}{2}\left(1 + \frac{1}{3} + \frac{1}{3 \cdot 4} + \cdots\right) \\
&\leq \frac{|b|^2}{2}\left(1 + \frac{1}{2} + \frac{1}{4} + \cdots\right) = |b|^2
\end{aligned}$$

□

Theorem 4.8. *Let $(X_{n,k})$ for $1 \leq k \leq K_n$ and $N \geq 1$ be $\mathbb{N}$-valued with $P[X_{n,k}] = p_{n,k}$ and $P[X_{n,k} \geq 2] = \varepsilon_{n,k}$. If*

1. *$X_{n,1}, \ldots, X_{n,K_n}$ are independent $\forall n$*
2. *$\sum_{k=1}^{K_n} p_{n,k} \xrightarrow[n\to\infty]{} \lambda \in [0, \infty)$ and $\max_k p_{n,k} \xrightarrow[n\to\infty]{} 0$*
3. *$\sum_{k=1}^{K_n} \varepsilon_{n,k} \to 0$ i.e. the expected number of values $\geq 2 \to 0$*

then

$$S_n = \sum_{k=1}^{K_n} X_{n,k} \xrightarrow[n\to\infty]{} \text{Poi}(\lambda)$$

Proof. Let

$$X'_{n,k} = \mathbf{1}_{\{X_{n,k}=1\}} = X_n \cdot \mathbf{1}_{\{X_n \le 1\}}$$

and

$$S'_n := X'_{n,1} + \cdots + X'_{n,K_n}$$

By the previous Theorem 4.4, ($p'_{n,k} = p_{n,k}$) we have $S'_n \xrightarrow{w} \text{Poi}(\lambda)$. Assumption (3) then implies that

$$P\left[S_n \neq S'_n\right] \le \sum_{k=1}^{K_n} P\left[X_{n,k} \neq X'_{n,k}\right] = \sum_{k=1}^{K_n} \varepsilon_{n,k} \to 0$$

since $\{X_{n,k} \neq X'_{n,k}\} = \{X_{n,k} \ge 2\}$. Note: $Y_n := S_n - S'_n \ge 0$. We now **claim** $Y_n \to 0$ in probability. To prove this claim, observe that

$$P[Y_n > \varepsilon] = P[S_n > \neq S'_n] = P[S_n \neq S'_n] \xrightarrow[n\to\infty]{} 0$$

Since $S'_n \xrightarrow{w} \text{Poi}(\lambda)$ and $Y_n \xrightarrow{w} 0$, then $S_n := S'_n + Y_n \xrightarrow{w} \text{Poi}(\lambda)$ (as proven, in general, on homework ***). □

Theorem 4.9 (Characterization of the Poisson process)**. *Intepretation*:** *Assume that we have random arrival times (occurrences) $\tau(\omega)$ in $\mathbb{R}^+$ (or $\mathbb{R}$) and let*

$$N_{s,t}(\omega) = \#\left|\tau(\omega) \cap (s,t]\right|$$

(For instance, this can represent the replacement times of light bulbs, arrival times at a bank line, arrival times of α-particles at a Geiger-Muller counter, etc.) Assume

1. *The # of points in disjoint intervals is independent*
2. *The $N_{s,t}$ distribution depends only on $t - s$*
3. $P\left[N_{s,t} - 1\right] = \lambda t + o(t)$ *as* $t \searrow 0$
4. $P\left[N_{s,t} \ge 2\right] = o(t)$ *as* $t \searrow 0$

Then $N_{0,t} - N_t \sim \text{Poi}(\lambda t)$.

Proof. Let

$$X_{n,k} = N_{(k-1)\frac{t}{n}, \frac{kt}{n}} \quad \text{for } k = 1, \ldots, n$$

Then

$$p_{n,k} = P\left[X_n = 1\right] = \lambda \cdot \frac{t}{n} + o\left(\frac{t}{n}\right)$$

and so

$$\sum_{k=1}^{n} \lambda \cdot \frac{t}{n} = \lambda t + \underbrace{n \cdot o(t/n)}_{\to 0 \text{ as } n\to\infty}$$

Also
$$\sum_{k=1}^{n} \varepsilon_{n,k} = \sum_{k=1}^{n} o\left(\frac{t}{n}\right) = n \cdot o\left(\frac{t}{n}\right) = \frac{o(t/n)}{(t/n)} \cdot t \xrightarrow[n\to\infty]{} 0$$
Then by the previous theorem,
$$N_t = \sum_{k=1}^{n} X_{n,k} \xrightarrow[n\to\infty]{} \mathrm{Poi}(\lambda t)$$

□

Remark 4.10. Such processes do exist. One way to construct such a process is to look at N_t as a renewal process with (i.i.d.) lifetime distribution $\exp(\lambda)$. That is, let
$$N_{0,t} = N_t := \inf\{k : T_1 + \cdots T_k \le t\}$$
with each $T_i \sim \exp(\lambda)$, and let $N_{s,t} := N_t - N_s$. In this case,
$$\tau(\omega) = \{T_1(\omega), T_1 + T_2(\omega), \dots\}$$
is the set of "replacement times" and
$$N_A(\omega) := |\tau(\omega) \cap A| = \text{ \# of points in } A$$

Theorem 4.11 (Law of Small Numbers)**.** *Assume we have a triangular array of 0-1 RVs $X_{n,1}, \dots, X_{n,k_n}$ for $n \ge 1$ where, for all n, $X_{n,1}, \dots, X_{n,n}$ are* independent *and such that $p_{n,m} := P[X_{n,m} = 1]$ satisfies*
$$\sum_{m=1}^{k_n} p_{n,m} \xrightarrow[n\to\infty]{} \lambda \in (0,\infty)$$
and
$$\max_{1\le m\le k_n} p_{n,m} \xrightarrow[n\to\infty]{} 0$$
Set
$$S_{k_n} := \sum_{m=1}^{k_n} X_{n,m}$$
so that $E[S_n] = \sum_m p_{n,m}$. Then, $S_{k_n} \xrightarrow{w} \mathrm{Poi}(\lambda)$ as $n \to \infty$.

Before we prove the theorem, recall that if $z \sim \mathrm{Poi}(\lambda)$ then $z \in \mathbb{N}$ and
$$P[z = k] = \exp(-\lambda)\frac{\lambda^k}{k!} =: \pi_\lambda(k)$$
Also,
$$\hat{\pi}_\lambda(t) = \int_{\mathbb{R}} \exp(itx)\pi_\lambda(dx) = \sum_{k\ge 0} \exp(itk)\exp(-\lambda)\frac{\lambda^k}{k!} = \exp(\lambda(\exp(it) - 1))$$

Proof. Set $\mu = \text{Poi}(\lambda)$ for $\lambda > 0$, so then $\hat{\mu}(t) = \exp(-\lambda)\exp(\lambda(\exp(it) - 1))$. (This also works for $\lambda = 0$.) Then,

$$\begin{aligned} E\left[\exp(itS_n)\right] &= \prod_{k=1}^{k_n} E\left[\exp(itX_{n,k}\right] = \prod_{k=1}^{k_n} \left(1 + p_{n,k}(\exp(it) - 1)\right) \\ &= \prod_{k=1}^{k_n} \exp\left(\log\left(1 + p_{n,k}(\exp(it) - 1)\right)\right) \\ &= \exp \sum_{k=1}^{k_n} p_{n,k}(\exp(it) - 1) + R_n \end{aligned}$$

Thus,

$$\lim_{n\to\infty} E\left[\exp(itS_n)\right] = \exp\left(\lambda(\exp(it) - 1)\right))$$

and so by Theorem 3.19, we have $S_n \xrightarrow{w} \text{Poi}(\lambda)$. □

5 Conditional Expectations

Let $X \geq 0$ or $X \in \mathcal{L}^1$ on $(\Omega, \mathcal{F}, P)$. The expectation $E[X]$ can be interpreted as an a priori prognosis for the value of X. Say we have a subfield $\mathcal{F}_0 \subseteq \mathcal{F}$ such that for every $A \in \mathcal{F}_0$, we know whether $\omega \in A$ or not. (For example, if $\mathcal{F}_0 = \sigma(Y)$, then we know for each c whether $\omega \in \{Y \leq c\}$ or not, so we know exactly $Y(\omega)$!) How does this partial information modify our a priori prognosis? If $X \in \mathcal{F}_0$ then our prognosis is exact $\Rightarrow X(\omega) = \hat{X}(\omega)$. For X, Y, we observe $S = X + Y$ and attempt $\hat{X}(\omega) = S(\omega) - E[Y]$ heuristically, but this is actually wrong. Notationally, we write $\hat{X}(\omega) = E[X|\mathcal{F}_0]$ for the *conditional* expectation. What exactly should $\hat{X}(\omega)$ be?

A *partial observation* is a collection of events (= observable events). When is an event A "observable"? Iff we can tell whether A occurred or not, i.e. whether $\omega \in A$ or $\omega \notin A$ (note: we don't know ω, only whether it is $\in A$).

Example 5.1. $\Omega = \{$people attending a film at the theatre$\}$ and $A = \{$more than 20 people$\}$ etc. Let $\mathcal{O} := \{A | A \text{is observable}\}$. Note $\mathcal{O}$ is closed under arbitrary unions and intersections, so it is a σ-algebra. This implies that partial information is associated with a σ-algebra.

Example 5.2. Information is often obtained by observing a RV Y (or several RVs) . Then $\mathcal{O} = \sigma(Y)$ since knowing the value of $Y(\omega)$ (but not ω) allows us to decide whether $\{Y \in B\}$ occurred or not for every Borel set $B \in \mathfrak{S}$. Moreover,

$$\sigma(Y) = \{\{Y \in B : B \in \mathfrak{S}\}$$

Prediction after an observation.

Example 5.3. Observe two events A, B, with

$$\mathcal{O} = \sigma(A, B) = \{A, B, \emptyset, \Omega, \underbrace{A \cap B}_{:=A_1}, \underbrace{A \setminus B}_{:=A_2}, \underbrace{B \setminus A}_{:=A_3}, \underbrace{(A \cup B)^c}_{:=A_4}\}$$

Notice $\mathcal{O}$ is atomic with atoms A_1, A_2, A_3, A_4. Let X be a RV (on $\mathbb{R}$ or $\bar{R}$). How should we refine our prediction (expectation) for X considering the observation $\mathcal{O}$? i.e. $E[X|\mathcal{O}] =?$ Note

$$E[X|\mathcal{O}] = E[X|A_i] \text{ if } A_i \text{ occurred}$$

and this motivates the definition

$$E[X|\mathcal{O}](\omega) = \sum_{i=1}^{4} \mathbf{1}_{A_i}(\omega) E[X|A_i]$$

This specific formula works for atomic σ-algebra, but it represents the general idea (i.e. a weighted average).

In general, we define

$$E[X|\mathcal{F}_0](\omega) = \sum_{\substack{i \geq 1 \\ P[A_i] \neq 0}} \mathbf{1}_{A_i}(\omega) E[X|A_i]$$

where $\mathcal{F}_0 = \sigma(Z)$. Note $\mathbf{1}_{A_i}$ is a RV and $E[X|A_i]$ is a constant, in the sum.

Example 5.4. For an atomic σ-algebra $\mathcal{F}_0 = \sigma(Z)$ (where Z is countable), the conditional expectations can be explicitly computed (see above)! We know which atom happens, meaning $\omega \in A_i$ for a certain i, and this happens with > 0 probability if $P(A_i) > 0$. Then

$$P[\cdot|A_i] = \frac{P[\cdot \cap A_i]}{P[A_i]}$$

is the conditional measure and

$$E[X|A_i] := \int X \, dP[\cdot|A_i] = \frac{1}{P[A_i]} E[X \mathbf{1}_{A_i}]$$

so we define

$$E[X|\mathcal{F}_0](\omega) = \sum_{\substack{i \geq 0 \\ P[A_i] \neq 0}} \mathbf{1}_{A_i}(\omega) E[X|A_i]$$

Theorem 5.5. *Let $X \geq 0$ or $X \in \mathcal{L}^1$ and let $\mathcal{F}_0 = \sigma(Z)$. Then $E[X|\mathcal{F}_0]$ has the following properties:*

1. *$E[X|\mathcal{F}_0]$ is $\in \mathcal{F}_0$.*
2. *$\forall Y_0 \geq 0$ with $Y_0 \in \mathcal{F}_0$,*

$$E[Y_0 \cdot X] = E[Y_0 \cdot E[X|\mathcal{F}_0]]$$

In particular,

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

using $Y_0 = \mathbf{1}_\Omega$.

Proof. (1) is trivial (constant on atoms). For (2), first use $Y_0 = \mathbf{1}_{A_i}$, and so

$$E\left[\mathbf{1}_{A_i} \cdot E[X|\mathcal{F}_0]\right] = E\left[\mathbf{1}_{A_i}(\omega) \sum_j \mathbf{1}_{A_j}(\omega) \cdot E[X|A_j]\right]$$
$$= E[X|A_i] \cdot E[\mathbf{1}_{A_i}] = E[\mathbf{1}_{A_i} X]$$

For $Y_0 = \sum_i c_i \mathbf{1}_{A_i}$, this follows from linearity and monotone integration for general $X \geq 0$. If $X \in \mathcal{L}^1$, separate X^+ and X^-. □

Proposition 5.6. *Let $X_1, \ldots, X_n$ be independent with $p = P[X_i = 1] = 1 - P[X_i = 0]$ and fix $\mathcal{F}_0 = \sigma(S_n)$ where $S_n = \sum_{i=1}^n X_i$. Then*

$$E[X_1|S_n](\omega) = \frac{1}{n} S_n(\omega)$$

Proof. Notice that

$$E[X_1|\sigma(S_n)](\omega) = \sum_{k=0}^{n} \mathbf{1}_{\{S_n = k\}} P[X_1 = 1|S_n = k]$$

and

$$P[X_1 = 1|S_n = k] = \frac{p \cdot \binom{n-1}{k-1} p^{k-1}(1-p)^{(n-1)-(k-1)}}{\binom{n}{k} p^k (1-p)^{n-k}} = \frac{k}{n}$$

Thus,

$$E[X_1|S_n](\omega) = \sum_{k=0}^{n} \mathbf{1}_{\{S_n = k\}} \frac{k}{n} = \frac{S_n}{n} \sum_{k=0}^{n} \mathbf{1}_{\{S_n = k\}} = \frac{S_n}{n}$$

□

Example 5.7. **Random sums**. Let $X_1, X_2, \ldots$ be RVs with $E[X_i] = m \in \mathbb{R}$ for all i. Let $T : \Omega \to \mathbb{N}$ be independent from $\sigma(X_1, X_2, \ldots)$. Let

$$S_{T(\omega)} := X_1(\omega) + \cdots + X_{T(\omega)}(\omega)$$

be a random sum. **Question**: does it follow that

$$E[S_T] = E[X_1] \cdot E[T]$$

Yes, and this is known as Wald's Identity. **Idea**: $E[S_T] = E[E[S_T|\sigma(T)]]$! Notice that

$$E[S_T|T](\omega) = \sum_{k \geq 0} \mathbf{1}_{\{T=k\}}(\omega) E[S_T|T = k] = \sum_{k \geq 0} \mathbf{1}_{\{T=k\}}(\omega) \cdot k \cdot m = m \cdot T(\omega)$$

since

$$E[S_T \mathbf{1}_{\{T=k\}}] = E[S_k \mathbf{1}_{\{T=k\}}] = E[S_k] P[T = k] = kE[X_1]$$

Thus,

$$E[E[S_T|T](\omega)] = mE\left[\sum_{k \geq 0} \mathbf{1}_{\{T=k\}}(\omega) T(\omega)\right] = mE[T]$$

General conditional expectation:

Definition 5.8. *Let $X \in \mathcal{F}^+$ or $X \in \mathcal{L}^1$ on $(\Omega, \mathcal{F}, P)$ with $\mathcal{F}_0 \subseteq \mathcal{F}$. Any RV X_0 with*

- $X_0 \in \mathcal{F}_0$ *and*
- $\forall A_0 \in \mathcal{F}_0,\ E[X\mathbf{1}_{A_0}] = E[X_0\mathbf{1}_{A_0}]$

is called a (version of) the conditional expectation of X given $\mathcal{F}_0$.

Theorem 5.9. *X_0 exists and is unique up to zero-measure sets.*

Proof. **Uniqueness**. Let X_0, X_0' be RVs with the properties above. Set $A_0 := \{X_0 > X_0'\} \in \mathcal{F}_0$. Then

$$E[\mathbf{1}_{A_0}X_0] = E[\mathbf{1}_{A_0}X] = E[\mathbf{1}_{A_0}X_0']$$

so $E[(X_0 - X_0')\mathbf{1}_{A_0}] = 0$ and thus $P[A_0] = 0$. Similarly, $P[X_0 < X_0'] = 0$, so $P[X_0 = X_0'] = 1$.

Existence. If $X \in \mathcal{F}^+$ already, define $Q[A_0] = E[X\mathbf{1}_{A_0}]$. This defines a ($\sigma$-finite) measure on $\mathcal{F}_0$ which is absolutely continuous with respect to $P \restriction_{\mathcal{F}_0} =: P_0$. By the Radon-Nikodym Theorem, $\exists X_0 \in \mathcal{F}_0^+$ such that

$$Q[A_0] = \int \mathbf{1}_{A_0}X_0\, dP_0 = E[X_0\mathbf{1}_{A_0}] = E[X\mathbf{1}_{A_0}]$$

for every $A_0 \in \mathcal{F}_0$. But then $Q[A_0] = E[\mathbf{1}_{A_0}X] = E[X_0\mathbf{1}_{A_0}]$, so X_0 is the conditional expectation.

For general $X \in \mathcal{L}^1$, write $X = X^+ - X^-$ with $X^+, X^- \in \mathcal{L}^1 \cap \mathcal{F}^+$. By the previous part, $(X^+)_0, (X^-)_0$ exist and are on $\mathcal{L}^1$. Set $X_0 := (X^+)_0 - (X^-)_0$ and check that the second condition

$$\begin{aligned} E[\mathbf{1}_{A_0}X] &= E[\mathbf{1}_{A_0}X^+] - E[\mathbf{1}_{A_0}X^-] = E[\mathbf{1}_{A_0}(X^+)_0] - E[\mathbf{1}_{A_0}(X^-)_0] \\ &= E[\mathbf{1}_{A_0}((X^+)_0 - (X^-)_0)] = E[\mathbf{1}_{A_0}X_0] \end{aligned}$$

is satisfied.

Finally, if $X \geq 0$ but $\notin \mathcal{L}^1$, then set $X_n := X \wedge n$ so that $X_n \nearrow X$. Set

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

which exists a.s. since it is $\nearrow$ (***). Then, for $A_0 \in \mathcal{F}_0$,

$$\begin{aligned} E[\mathbf{1}_{A_0}X] &= \lim \nearrow E[\mathbf{1}_{A_0}X] = \lim \nearrow E[\mathbf{1}_{A_0}E[X_0|\mathcal{F}_0]] \\ &= E[\mathbf{1}_{A_0}\underbrace{\lim \nearrow E[X_n|\mathcal{F}_0]}_{=E[X|\mathcal{F}_0]}] \end{aligned}$$

and so we have the conditional expectation of X. $\square$

5.1 Properties and computational tools

1. $E[\cdot|\mathcal{F}_0]$ is *monotone*; i.e. $X \geq 0 \;\Rightarrow\; E[X|\mathcal{F}_0] \geq 0$.
2. $E[E[X|\mathcal{F}_0]] = E[X]$ (use $\mathbf{1}_{A_0} = \Omega$)
3. Let $Y_0 \in \mathcal{F}_0$ and $X_0 = E[X|\mathcal{F}_0]$. Then

$$X, Y_0 \geq 0 \text{ a.s. } \Rightarrow E[XY_0] = E[X_0 Y_0]$$

and

$$XY_0 \in \mathcal{L}^1 \Rightarrow X_0 Y_0 \in \mathcal{L}^1 \text{ and } E[XY_0] = E[X_0 Y_0]$$

To prove the first claim, write $Y_0 = \lim \nearrow Y_{0,n}$ where $Y_{0,n}$ are simple functions $\in \mathcal{F}_0$. Then,

$$\begin{aligned} E[XY_0] = E\left[X(\lim \nearrow Y_{0,n})\right] &= \lim_{n\to\infty} \nearrow E[XY_{0,n}] \\ &= \lim_{n\to\infty} E[X_0 Y_{0,n}] = E[X_0 \lim Y_{0,n}] = E[X_0 Y_0] \end{aligned}$$

by monotone integration and linearity. To prove the second claim, assume WOLOG $Y_0 \geq 0$ and let $X = X^+ - X^-$ with $0 \leq X^-, X^+ \in \mathcal{L}^1$. Then,

$$\begin{aligned} E[XY_0] &= E[X^+Y_0] - E[X^-Y_0] = E[(X^+)_0 Y_0] - E[(X^-)_0 Y_0] \\ &= E[Y_0((X^+)_0 - (X^-)_0)] = E[Y_0 X_0] \end{aligned}$$

using the first claim.

4. Let $Y_0 \in \mathcal{F}_0$ and assume $X, Y_0 \in \mathcal{F}^+$ and $X \in \mathcal{L}^1$ and $XY_0 \in \mathcal{L}^1$. Then

$$E[XY_0|\mathcal{F}_0] = Y_0 E[X|\mathcal{F}_0] \text{ a.s.}$$

To prove this, we have to check that the RHS is a version of the conditional expectation of $XY_0|\mathcal{F}_0$. First, RHS$\in \mathcal{F}_0$. Second, we can apply (3) to say

$$E[\mathbf{1}_{A_0} \cdot (Y_0 E[X|\mathcal{F}_0])] = E[\overbrace{\mathbf{1}_{A_0} Y_0}^{\in \mathcal{F}_0} E[X|\mathcal{F}_0]] = E[\mathbf{1}_{A_0} Y_0 X]$$

There are two special cases of this property.

   (a) If $\mathcal{F}_0 = \mathcal{F}$ (total information) and $X \in \mathcal{F}$ then $E[X|\mathcal{F}_0] = X \cdot E[1|\mathcal{F}_0] = X$.

   (b) If $\mathcal{F}_0$ is trivial (i.e. 0-1) or $\mathcal{F}_0$ is independent of X then $X_0 = E[X]$ a.s.

5. If $X, Y \in \mathcal{F}^+$ or $X, Y, XY_0, YX_0 \in \mathcal{L}^1$, then

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

To prove this, set $Y_0 = E[Y|\mathcal{F}_0] \in \mathcal{F}_0$ and apply (4):

$$E[XY_0] = E[X_0 Y_0] = E[X_0 Y]$$

where the second equality is by symmetry.

6. **Projectivity**: Let $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}$ and $X \in \mathcal{F}^+$ or $X \in \mathcal{L}^1$. Then

$$E[\underbrace{E[X|\mathcal{F}_1]}_{:=X_1}|\mathcal{F}_0] = \underbrace{E[X|\mathcal{F}_0]}_{:=X_0}$$

To prove this, we show that X_0 is the conditional expectation of $X_1|\mathcal{F}_0$. For $A_0 \in \mathcal{F}_0$, we apply the second property in the definition of conditional expectation twice to write

$$E[\mathbf{1}_{A_0} X_1] = E[\mathbf{1}_{A_0} X] = E[\mathbf{1}_{A_0} X_0]$$

which implies X_0 is a conditional expectation of $X_1|\mathcal{F}_0$.

Further properties of conditional expectation.

Theorem 5.10. *Let $Y, X \in \mathcal{F}^+$ or $\mathcal{L}^1$ on $(\Omega, \mathcal{F}, P)$ with $\mathcal{F}_0 \subseteq \mathcal{F}$. Then*

1. *Linearity: $E[X + Y|\mathcal{F}_0] = E[X|\mathcal{F}_0] + E[Y|\mathcal{F}_0]$ a.s. and $E[cX|\mathcal{F}_0] = cE[X|\mathcal{F}_0]$.*

2. *Monotonicity: if $X \geq Y$ and X or $Y \in \mathcal{F}+$ or $\mathcal{L}^1$ then $E[X|\mathcal{F}_0] \geq E[Y|\mathcal{F}_0]$.*

3. *"Monotone continuity" (Beppo-Levi) If $\mathcal{L}^1 \ni Y \leq X_1 \leq X_2 \leq \cdots$ a.s. then*

$$E\left[\lim_{n\to\infty} \nearrow X_n|\mathcal{F}_0\right] = \lim_{n\to\infty} E[X_n|\mathcal{F}_0] \text{ a.s.}$$

****

5.2 Conditional Expectation and Product Measures

Let $(\Omega, \mathcal{F}, P)$ be a probabiltiy space and let $X_i : \Omega \to (S_i, \mathfrak{S}_i)$ for $i = 0, 1$. Assume that the joint distribution of (Z_0, Z_1) (on $S_0 \times S_1$ with the product σ-algebra $\mathfrak{S}_0 \times \mathfrak{S}_1$) is of the form $P_0 \otimes K(\cdot,)$ for some stochastic kernel K, where P_0 is the distribution of Z_0.

Question: Given $f \in (\mathfrak{S}_0 \times \mathfrak{S}_1)^+$, what is

$$E\left[f(Z_0, Z_1)|Z_0\right](\omega) =?$$

Example 5.11. Let T_0, T_1 be independent $\exp(\alpha)$ distributed RVs and define $f(X, Y) = \min\{X, Y\}$. Then

$$E[\min\{T_0, T_1\}|T_0] =? = \varphi(T_0)$$

for some measurable function φ.

Theorem 5.12.

$$E\left[f(Z_0, Z_1)|Z_0\right](\omega) = \int_{S_1} f(Z_0(\omega), s) K(Z_0(\omega), ds) \text{ a.s.}$$

Proof. Omitted for now. (***) □

Corollary 5.13. *If Z_0, Z_1 are independent ($\iff K(Z_0, \cdot) = P_1(\cdot)$ where P_1 is the distribution of Z_1 on S_1), then*

$$E\left[f(Z_0, Z_1)|Z_0\right](\omega) = \int_{S_1} f(Z_0(\omega), s)P_1(ds) = E\left[f(Z_0(\omega), Z_1)\right]$$

Sometimes the following extension is useful.

Corollary 5.14. *Let $Z_0 \in \mathcal{F}_0 \subseteq \mathcal{F}$ and assume Z_1 is independent from $\mathcal{F}_0$ (i.e. $\sigma(Z_1), \mathcal{F}_0$ are independent). Note: this is* stronger *than assuming Z_0, Z_1 are independent. Then*

$$E\left[f(Z_0, Z_1)|\mathcal{F}_0\right](\omega) = E\left[f(Z_0(\omega), Z_1)\right]$$

is still valid.

Proof. Apply Corollary 5.13 to $\tilde{Z}_0 := \mathrm{id} : (\Omega, F) \to (\Omega, \mathcal{F}_0)$. Then $\forall g : (\Omega, \mathcal{F}) \times S_1 \to \mathbb{R}^+$, Corollary 5.13 tells us

$$E\left[g(\tilde{Z}_0, Z_1)|\mathcal{F}_0\right](\omega) = E\left[g(\omega, Z_1)\right]$$

noting that $\mathcal{F}_0 = \sigma(\tilde{Z}_0)$. Set g to be the particular function given by $g(\omega, s) = f(Z_0(\omega), s)$. Then

$$E[g(\tilde{Z}_0, Z_1)|\mathcal{F}_0](\omega) = E[f(Z_0(\tilde{Z}_0), Z_1)|\mathcal{F}_0]$$

and note that the LHS is

$$E[g(\omega, Z_1)] = E[f(Z_0(\omega), Z_1)]$$

and the RHS is

$$E[f(Z_0, Z_1)|\mathcal{F}_0](\omega)$$

□

Example 5.15. Let T_0, T_1 be independent $\exp(\alpha)$ distributed with $\alpha > 0$. Then

$$\begin{aligned}
E[\min\{T_0, T_1\}|T_0](\omega) &= E[\min\{T_0(\omega), T_1\}] \\
&= \int \min\{T_0(\omega), T_1)\, P_1(ds) \\
&= \int_0^\infty (T_0(\omega) \wedge s) \cdot \alpha \exp(-\alpha s)\, ds \\
&= \int_0^{T_0(\omega)} s\alpha \exp(-\alpha s)\, ds + T_0(\omega) \int_{T_0(\omega)}^\infty \alpha \exp(-\alpha s)\, ds \\
&= -T_0(\omega) \exp(-\alpha T_0(\omega)) + 0 - \frac{1}{\alpha} \exp(-\alpha T_0) \\
&\quad + \frac{1}{\alpha} + (T_0 \exp(-\alpha T_0(\omega))) \\
&= \frac{1}{\alpha}\left(1 - \exp(-\alpha T_0)\right)
\end{aligned}$$

5.2.1 Conditional Densities

Let X_i for $i = 0, 1$ be RVs on $(S_i, \mathfrak{S}_i)$ with joint distribution on $S_0 \times S_1$ given by

$$(X_0, X_1) \sim_P \varphi(x_0, x_1)\mu_0(dx_0)\mu_1(dx_1)$$

where μ_i are σ-finite measures; i.e. (X_0, X_1) has a joint density $\varphi \geq 0$ with respect to the product measure $\mu_0 \otimes \mu_1$ (which is also σ-finite). In this case, we can write the joint distribution μ of (X_0, X_1) as $P_0 \otimes K$, where

$$P_0(dx) = \varphi_0(x_0)\mu_0(dx_0)$$

is a measure on S_0 and where

$$\varphi_0(x_0) := \int_{S_1} \varphi(x_0, x_1)\mu_1(dx_1)$$

is the density of x_0 with respect to μ_0, and

$$K(x_0, dx_1) = \begin{cases} \varphi_{x_1|x_0}(x_0, x_1) & \text{if } \varphi_0(x_0) > 0 \\ \text{any (fixed) prob. dist.} & \text{if } \varphi_0(x_0) = 0 \end{cases}$$

where we recall that

$$\varphi_{x_1|x_0}(x_0, x_1) := \frac{\varphi(x_0, x_1)}{\int_{S_1} \varphi(x_0, x_1)\mu_1(dx_1)} = \frac{\varphi(x, y)}{\varphi_0(x)}$$

Check K on rectangles! (***)

Now, we return to the question of what $E[f(X_0, X_1)|X_0](\omega)$ should be. We can write

$$\begin{aligned} E[f(X_0, X_1)|X_0](\omega) &= \int_{S_1} f(X_0(\omega), x_1) \cdot K(X_0(\omega), dx_1) \\ &= \int_{S_1} f(X_0(\omega), x_1)\varphi_{x_1|x_0}(X_0(\omega), x_1)\mu_1(dx_1) \end{aligned}$$

Remark 5.16. If $f(X_0, X_1) = f(X_1)$ then notice that

$$E[f(X_1)|X_0](\omega) = \int_{S_1} f(x_1)\varphi_{X_1|X_0}(X_0(\omega), x_1)\mu_1(dx_1)$$

and compare this to

$$E[f(X_1)] = \int_{S_1} f(x_1) \cdot \varphi(x_1)\mu_1(dx_1)$$

The difference is in the φ_1 versus $\varphi_{X_1|X_0}$ term, where the first one is the marginal distribution of X_1 and the second one is the conditional distribution of X_1 given $X_0(\omega)$.

Remark 5.17. If X_0, X_1 are independent, then $\varphi_{X_1|X_0}(X_0, X_1) = \varphi_1(X_1)$ and then

$$E[f(X_1)|X_0](\omega) = \int_{S_1} f(x_1)\varphi_1(x_1)\mu_1(dx_1) = E[f(X_1)]$$

6 Martingales

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\mathcal{A}_0 \subseteq \mathcal{A}_1 \subseteq \cdots \subseteq \mathcal{F}$ be a sequence of σ-algebras; we call such a sequence a *filtration* and refer to $(\Omega, \mathcal{F}, (\mathcal{A}_k), P)$ as a *filtered probability space*. Let $(X_n)_{n\geq 0}$ be a stochastic process.

Definition 6.1. *We say X is* adapted *on $(\mathcal{A}_k)$ if $X_k \in \mathcal{A}_k$ for all $k \geq 0$. We say $(Y_k)_{k\geq 1}$ is* previsible *(with respect to $(\mathcal{A}_k)$) if $Y_k \in \mathcal{A}_{k-1}$ for all $k \geq 1$. We say (Z_k) is* innovative *if $Z_k \in \mathcal{L}^1$ and satisfies the martingale property (see below).*

Definition 6.2. *We say X is a* martingale *with respect to $\mathcal{A}$ if*

1. *X is adapted and $X_k \in \mathcal{L}^1$ for all k, and*
2. *X satisfies the martingale property*

$$E[\Delta_{n+1}X|\mathcal{A}_n] := E[(X_{n+1} - X_n)|\mathcal{A}_n] = 0 \text{ a.s.} \tag{6}$$

which is equivalent to

$$E[X_{n+1}|\mathcal{A}_n] = X_n \text{ a.s.}$$

Remark 6.3. For all $n, k \geq 0$,

$$\begin{aligned} E[X_{n+k} - X_n|\mathcal{A}_n] &= E\left[\sum_{\ell=n+1}^{n+k} \Delta_\ell X|\mathcal{A}_n\right] \\ &= \sum_{\ell=n+1}^{n+k} E[\Delta_\ell X|\mathcal{A}_n] = 0 \text{ a.s.} \\ &= E[\underbrace{E[\Delta_\ell X|\mathcal{A}_{\ell-1}]}_{=0 \text{ a.s.}}|\mathcal{A}_n] \end{aligned}$$

where the last line follows by projectivity. In particular, for $n = 0$ fixed, then for every $k \geq 0$,

$$E[X_k|\mathcal{A}_0] = X_0 \Rightarrow E[X_k] = E[X_0]$$

Example 6.4. Let $Y_1, Y_2, \ldots$ be independent $\mathcal{L}^1$ RVs. Set $\mathcal{A}_n := \sigma(Y_1, \ldots, Y_n)$ for $n \geq 1$ and $\mathcal{A}_0 = \{\emptyset, \Omega\}$. Then

$$X_n := \sum_{i=1}^{n}(Y_i - E[Y_i]) \quad , \quad X_0 = 0$$

is a martingale with respect to $\mathcal{A}$. In general, partial sums of independent, centered $\mathcal{L}^1$ RVs form a martingale (with respect to their own filtration). Note, as well, that $\mathcal{A}_n = \sigma(Y_1, \ldots, Y_n) = \sigma(X_1, \ldots, X_n)$.

Example 6.5 (Successive prognosis). Let $X \in \mathcal{L}^1(\mathcal{F})$ and $\mathcal{A}$ be given. Then

$$X_n := E[X|\mathcal{A}_n]$$

is a martingale. To see why, notice that

$$E[X_{n+1}|\mathcal{A}_n] = E[E[X|\mathcal{A}_{n+1}]|\mathcal{A}_n] = E[X|\mathcal{A}_n] = X_n$$

6.1 Gambling Systems and Stopping Times

Let (X_n) be a martingale with respect to $\mathcal{A}$ and let $(V_n)_{n\geq 1}$ be previsible such that $V_n(\Delta_n X) = V_n(X_n - X_{n-1}) \in \mathcal{L}^1$. Set

$$(V.X)_n = X_0 + \sum_{k=1}^{n} V_k \cdot \Delta_k X$$

This is called a "gambling system" (or a martingale transform or a discrete stochastic integral).

Example 6.6. 1. Let X_n be SSRW (i.e. $X_0 = x_0, \Delta_n X$ are i.i.d. ± 1, centered), let $V_n = 1$ and $(\mathcal{A}_n) = \sigma(X_1, \ldots, X_n)$. Since V_n is previsible (*** why?) then $(V.X)$ is a gambling system.

Interpretation Since $\Delta_n X = \pm 1$ with probability $\frac{1}{2}$, you bet on 1 each time with \$1. You start with $\$x_0$ and continue betting. Then $(V.X)_n$ is your balance after the nth bet.

Theorem 6.7. *If $(V.X)$ is a gambling system then $V.X$ is a martingale (with respect to $\mathcal{A}_n = \sigma(X_1, \ldots, X_n)$).*

Proof. To show $(V.X)$ is adapted, observe that

$$V \cdot X_n = \underbrace{V_n}_{\in \mathcal{A}_{n-1}} \cdot \underbrace{(X_n - X_{n-1})}_{\in \mathcal{A}_n} \in \mathcal{A}_n$$

and is $\in \mathcal{L}^1$ (by assumption). To show the martingale property, note that

$$\begin{aligned} E[(V.X)_n - (V.X)_{n-1} | \mathcal{A}_{n-1}] &= E[\overbrace{V_n}^{\in \mathcal{A}^{n-1}} \cdot \Delta_n X | \mathcal{A}_{n-1}] \\ &= V_n \cdot \underbrace{E[\Delta_n X | \mathcal{A}_{n-1}]}_{=0 \text{ a.s.}} = 0 \text{ a.s.} \end{aligned}$$

□

In particular,

$$E[(V.X)_n] = E[(V.X)_0] = E[X_0] = x_0$$

so there is nothing to gain (on average).

2. Let $X_0 = x_0$ and set $\Delta_k X = \pm 1$ with probability $\frac{1}{2}$ (independent) and let $\mathcal{A}_n = \sigma(X_1, \ldots, X_n)$. Let

$$V_k = \begin{cases} 1 & \text{if } (V.X)_{k-1} \leq x_0 \\ 0 & \text{if } (V.X)_{k-1} > x_0 \end{cases}$$

Note

$$V_k \in \sigma((V.X)_{k-1}) \subseteq \sigma(X_1, \ldots, X_{k-1}) = \mathcal{A}_{k-1}$$

since

$$(V.X)_{k-1} = x_0 + \sum_{j=1}^{k-1} \underbrace{V_j}_{\in \mathcal{A}_{j-1}} \cdot \underbrace{(X_j - X_{j-1})}_{\in \mathcal{A}_j \subseteq \mathcal{A}_{k-1}}$$

and therefore V_k is previsible, so $(V \cdot X)$ is a gambling system. Note: since RV oscillates between $\pm\infty$ (*** later) we a.s. win \$1 but, unfortunately, the expected time to win $= +\infty$! (Also, expected loss before winning is $+\infty$. Yikes!)

3. (A version with shorter waiting time) Take the same SRW as X and set

$$V_k = \begin{cases} 2^{k-1} & \text{if } \Delta_1 X = \Delta_2 X = \cdots = \Delta_{k-1} X = -1 \\ 0 & \text{otherwise} \end{cases}$$

Clearly, V_k is $\sigma(X_1, \dots, X_{k-1})$ measurable and thus predictable. In general, $(V.X)_n = x_0 + 1$ after we've won. If $T(\omega)$ is the time we win (the first time), then $T \sim \text{geom}(1/2)$, and in practice $E[T] < \infty$. However, $(V.X)_n$ is *not* uniformly integrable; there may be big losses before making even \$1!

Definition 6.8 (Stopping time). *A RV $T : \Omega \to \bar{\mathbb{N}}$ such that $\{T = n\} \in \mathcal{A}_n$ for every $n = 0, 1, \dots$ is called a stopping time.*

Remark 6.9. The property in the definition above is equivalent to saying $\{T \leq n\} \in \mathcal{A}_n$ for all n. Notice that

$$\{T \leq n\} = \underbrace{\{T = 0\}}_{\in \mathcal{A}_0 \subseteq \mathcal{A}_n} \cup \underbrace{\{T = 1\}}_{\in \mathcal{A}_1 \subseteq \mathcal{A}_n} \cup \cdots \cup \underbrace{\{T = n\}}_{\in \mathcal{A}_n} \in \mathcal{A}_n$$

and

$$\{T = n\} = \underbrace{\{T \leq n\}}_{\in \mathcal{A}_n} \setminus \underbrace{\{T > n-1\}}_{\in \mathcal{A}_{n-1}} \in \mathcal{A}_n$$

So an interpretation of a stopping time is that at time n, we know whether $T(\omega) \leq n$ or $T > n$. (What we can't tell, in general, is whether $T > n+1$, for instance, and other similar things.)

Example 6.10. 1. Let $A \in \mathcal{B}_{\mathbb{R}}$, and let (X_n) be adapted on $\mathcal{A}_n$. The first entrance (or hitting) time of A is given by

$$T_A(\omega) = \inf\{n \geq 0 | X_n(\omega) \in A\} (\leq +\infty)$$

and it *is* a stopping time. To see why, observe that

$$\{T_A \leq n\} = \bigcup_{k=0}^{n} \underbrace{\{X_k \in A\}}_{\in \mathcal{A}_k \subseteq \mathcal{A}_n} \in \mathcal{A}_n$$

2. Let (X_n) be a SRW and set $\mathcal{A}_n := \sigma(X_0, \dots, X_n)$. A *run of length* r is a segment of the walk consisting of successive upwards steps. Let $r \geq 1$ be fixed. Then

$$T(\omega) = T_r = \inf\{n | (n-r, n-r+1, \dots, n) \text{ is a run}\}$$

represents the first time that a run of length r has been completed, and it *is* a stopping time. To see why, let $k \geq r$ and set

$$R_k = \{\Delta_k X = \Delta_{k-1} X = \cdots = \Delta_{k-r+1} X = +1\}$$

Note that $R_k \in \mathcal{A}_k$. Then

$$\{T \leq n\} = \bigcup_{r \leq k \leq n} R_k \in \mathcal{A}_n$$

Example 6.11. Here are two examples that are *not* stopping times:

1. $L_A = \sup\{n \geq 0 | X_n \in \mathcal{A}\}$, i.e. the "last visit" in A
2. the beginning of the first run of length r

Definition 6.12. *If X is a process and T is a random time, then*

1. *$X_\cdot^T$ is a "stopped process". For all n, let $X_n^T(\omega) := X_{n \wedge T(\omega)}(\omega)$.*
2. *The "process at (time) T" is defined by $X_T(\omega) := X_{T(\omega)}(\omega)$ (a RV).*

Note: after T (i.e. $n \geq T(\omega)$), $X_n^T(\omega) = X_T(\omega)$.

Theorem 6.13. *Let X be a martingale and T a stopping time with respect to $(\mathcal{A})$. Then $(X_\cdot^T)$ is a martingale with respect to $(\mathcal{A})$.*

Proof. Let $V_n := \mathbf{1}_{\{T \geq n\}}$, so $V.$ is previsible. Then

$$\underbrace{V_n}_{\text{bdd}} \cdot (\underbrace{X_n - X_{n-1}}_{\in \mathcal{L}^1}) \in \mathcal{L}^1 \ \Rightarrow \ (V.X) \text{ is a martingale}$$

But $(V.X) = X^T$. This proves the claim. To see why $(V.X) = X^T$, observe that

$$(V.X)_n = X_0 + \sum_{k=1}^{n} \mathbf{1}_{\{T \geq k\}} \cdot \Delta_k X = X_0 + \sum_{k=1}^{T \wedge n} \Delta_k X = X_n^T$$

□

Theorem 6.14 (Optional Stopping)**.** *Let $X.$ be a $\mathcal{A}$-martingale and T be a stopping time. Then*

1. *$X_\cdot^T$ is a martingale and $E[X_{T \wedge n}] = E[X_0]$*

2. *If T is bounded (i.e. $T \leq N$ a.s.) then*

$$E[X_T] = E[X_{T \wedge n}] = E[X_N^T] = E[X_0]$$

3. *If $T < \infty$ a.s. and $(X_n^T)_{n \geq 0}$ is uniformly integrable, then $E[X_T] = E[X_0]$.*

Proof. We only prove (3). Apply uniform integrability and a.s. convergence to write

$$E[X_T] = E\left[\lim_{n\to\infty} X_{T\wedge n}\right] = \lim_{n\to\infty} E[X_{T\wedge n}] = E[X_0]$$

□

Example 6.15. **Application: classical ruin problem** (gambling fairly to make $(b - x_0)$\$ with credit level a). Let $X_n = x + \sum_{i=1}^n Y_i$ where the (Y_i) are i.i.d. 1 with probability p and -1 with probability $1 - p$. Define

$$T(\omega) = \min\{n \geq 0 | X_n(\omega) \notin (a, b)\}$$

which is a stopping time. By Borel-Cantelli, $T < \infty$ a.s. (***). Define

$$r(x) := P[X_T = a]$$

to be the "ruin probability".

1. $p = \frac{1}{2}$. Then $X.$ is a martingale and $(X_{n\wedge T})$ is bounded and therefore uniformly integrable. Thus,

$$x = E[X_0] = E[X_T] = b \cdot \overbrace{P[X_T = b]}^{=1-P[X_T=a]} + aP[X_T = a]$$

   and so

$$x = b(1 - r(x)) + ar(x) \;\Rightarrow\; r(x) = \frac{b - x}{b - a}$$

2. $p \neq \frac{1}{2}$. Let $h(x) := \left(\frac{1-p}{p}\right)^x$. Then $h(X_n)$ is a martingale (*** HW), and so

$$E[h(X_0)] = h(x) = E[h(X_T)] = [h(X)]_T = h(b)(1 - r(x)) + h(a)r(x)$$

   Thus,

$$r(x) = \frac{h(b) - h(x)}{h(b) - h(a)} = \frac{1 - \left(\frac{p}{1-p}\right)^{b-x}}{1 - \left(\frac{p}{1-p}\right)^{b-a}}$$

3. $p < \frac{1}{2}$. Then $r(x) \geq 1 - \left(\frac{p}{1-p}\right)^{b-x}$ and this bound doesn't depend on a! For instance, if $p = \frac{18}{37}$ then $b - x = 128$ is sufficient to have $r(x) \geq 0.999$! That is, before winning 128, you are ruined no matter how much reserves you have (assuming finite reserves, of course).

Example 6.16. **Application**: How long do you have to wait for the occurrence of a fixed binary text $[a_1, \ldots, a_N]$ in a random binary sequence (with $p = 1/2$)?

Let $(Y_k)_{k\geq 1}$ be i.i.d. ± 1 with $p = \frac{1}{2}$ and set $\mathcal{A}_k = \sigma(Y_1, \ldots, Y_k)$. Let the stopping time be

$$T(\omega) = \inf\{n \geq 1 | Y_{n-N+1}(\omega) = a_1, \ldots, Y_n(\omega) = a_N$$

By Borel-Cantelli, $T < \infty$ a.s. What is $E[T]$? We estimate

$$\frac{T}{N} \leq T' := \inf\{k : [a_1, \ldots, a_N] \text{ occurs in the } k\text{-th block}\}$$

so T' is a geometric RV with parameter 2^{-N}. Thus, $E[T'] = \frac{1}{2^{-N}} = 2^N$ and so $E[T] \leq N2^N < \infty$.

At each (fixed) time k with $0 \leq k \leq T - 1$, start a *game* (i.e. gambling system) with the martingale $X_n = \sum_{k=1}^n Y_k$ as follows:

1. We bet 1 on seeing a_1 next. If we lose, we lost 1 and the entire game is over. If we win, we get back 2 and we continue.

2. We bet 2 on seeing a_2 next. If we lose, we lost 2 and the game is finished. If we win, we get 4 and continue.

3. We bet 4 on seeing a_3 next, ….

N We bet 2^{N-1} on seeing a_N next. If we lose, finish the game with overall loss 1. If we win we get back 2^N and finish, with a net win of $2^N - 1$.

Note: up to time k, there are k games. *Each* game consists of a random number (at least one, at most N) of bets, and each game is self-financing after paying the initial \$1. The balance of each finished game is either $0 - 1$ if we lost or $2^N - 1$ if we won. For an unfinished game with k winning bets, the balance is $2^k - 1$. What is our balance at time T? (i.e. the first time we win an entire game) We have

$$\begin{aligned}(V.X)_T &= \text{price of all } T \text{ games } + \text{ amount won} \\ &= -T + \text{ amount won in the last } N \text{ games } + 0 \\ &= -T + \sum_{k=1}^{N} 2^{N-k+1} \cdot W_k(\omega)\end{aligned}$$

where

$$W_k(\omega) = \begin{cases} 1 & \text{if } T - N + k\text{-th game is won at time } T(\omega) \\ 0 & \text{otherwise} \end{cases}$$

Note: $W_k(\omega)$ is deterministic! This is because we know the end of the sequence $Y_{T-N+1}, \cdots, Y_T$! In particular, $k = 1 \Rightarrow W_k = 1$. In general $W_k = 1 \iff$
********* insert picture *********

Note again that if $W_k = 1$ then the final payoff is 2^{N-k+1} and so

$$(V.X)_T = (_T) + \sum_{k=1}^{N} 2^{N-k+1} \cdot W_k$$

since W_k depends only on $[\alpha_1, \ldots, \alpha_N]$ (i.e. it is deterministic). We will that $(V.X)_{T\wedge n}$ is uniformly integrable, specifically

$$\underbrace{-T}_{\in \mathcal{L}^1} \leq (V.X)_{T\wedge n} \leq \underbrace{2^{N+1}}_{\in \mathcal{L}^1}$$

so then

$$E[(V.X)_T] = \lim_{n\to\infty} E[\underbrace{(V.X)_{T\wedge n}}_{=0}] = 0$$

and finally

$$0 = -E[T] + \sum_{k=1}^{N} 2^{N-k+1} W_k \;\Rightarrow\; E[T] = \sum_{k=1}^{N} 2^{N-k+1} W_k$$

The RHS is larger the more "repetitive" the text $[\alpha_1, \ldots, \alpha_N]$ is. For instance,

$[a_1, \ldots, a_N]$	$E[T]$
000000	126
001100	70
011111	64

Finally, we show the estimate works:

$$(V.X)_{T\wedge n} \geq -(T \wedge n) \geq -T \in \mathcal{L}^1$$

and

$$(V.X)_{T\wedge n} \leq \sum_{k=1}^{N} 2^k = 2^{N+1}$$

since only in the last N games can we win.

6.2 Martingale Convergence

Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{A}.$ a filtration, and $X.$ a martingale. For $a < b$ and $N \in \mathbb{N}$ fixed, we define

$$U_{a,b}^{N}(\omega) = \text{ \# of upcrossings of } [a, b] \text{ during time } [0, N]$$

More precisely, set $S_0 = T_0 = 0$ and

$$S_k(\omega) = \inf\{n \geq T_{k-1}(\omega) : X_n(\omega) \leq a\} = \text{ beginning of } k\text{-th upcrossing}$$

and

$$T_k(\omega) = \inf\{n \geq S_k(\omega) : X_n(\omega) \geq b\} = \text{ end of } k\text{-th upcrossing}$$

and then define

$$U_{a,b}^N(\omega) = \max\{k \geq 0 : T_k(\omega) \leq N\} = \text{ \# of upcrossings during } [0, N]$$

Lemma 6.17 (Upcrossing inequality)**.**

$$E[U_{a,b}^N] \leq \frac{E[(X_N - a)^-]}{b-a}$$

and this implies, in particular,

$$E[U_{a,b}^\infty] \leq \frac{1}{b-a} \sup_N E[(X_N - a)^-]$$

Proof. Since S_k, T_k are stopping times (they are only defined in terms of information before them), then the Stopping Theorem 6.14 implies $E[Z_N] = 0$, where

$$Z_N = \sum_{k=1}^N (X_{T_k \wedge N} - X_{S_k \wedge N})$$

On the other hand,

$$\begin{aligned} Z_N &= \sum_{k=1}^{U^N} (X_{T_k} - X_{S_k}) + (\underbrace{X_N}_{=N \wedge T_{U^N+1}} - X_{N \wedge S_{U^N+1}}) \\ &\geq U_{a,b}^N \cdot (b-a) + \underbrace{(X_N - X_{N \wedge S_{U^N+1}})}_{:=\star} \end{aligned}$$

and since

$$\star = \begin{cases} 0 & \text{if } S_{U^N+1} \geq N \\ \geq \underbrace{X_N - a}_{\leq 0} & \text{if } S_{U^N+1} < N \geq -(X_N - a)^- \end{cases}$$

we can say

$$Z_N \geq U_{a,b}^N \cdot (b-a) + -(X_N - a)^-$$

so

$$E[Z_N] = 0 \geq E[U_{a,b}^N] \cdot (b-a) - E[(X_N - a)^-]$$

□

Theorem 6.18 (Martingale convergence)**.** *Let $X.$ be an $\mathcal{L}^1$-bounded martingale. Then*

$$X_\infty(\omega) = \lim_{n\to\infty} X_n(\omega) \textit{ exists a.s., and } X_\infty \in \mathcal{L}^1$$

Remark 6.19. If $X.$ is a martingale ($X \in \mathfrak{M}$), then TFAE

1. X is $\mathcal{L}^1$-bounded ($\sup_n E[|X_n|] < \infty$)
2. $\sup_n E[X_n^+] < \infty$
3. $\sup_n E[X_n^-] < \infty$

To see why, note that

$$E[|X_n|] = E[X_n^-] + E[X_n^+] = (E[X_n^+] - \underbrace{E[X_n]}_{=E[X_0]}) + E[X_n^+] = 2E[X_n^+] - E[X_0]$$

and take sups ...

Proof. Observe that

$$\left\{\liminf_{n\to\infty} X_n < \limsup_{n\to\infty} X_n\right\} \subseteq \bigcup_{\substack{a,b\in\mathbb{Q}\\ a<b}} \{U_{a,b} = \infty\}$$

where $U_{a,b} =: \lim_{N\to\infty} U_{a,b}^N$. This implies

$$P[X_n(\cdot) \text{ doesn't converge}] \leq \sum_{\substack{a,b\in\mathbb{Q}\\ a<b}} P[U_{a,b} = \infty]$$

and we know $P[U_{a,b} = \infty] = 0$ provided $E[U_{a,b}] < \infty$. But this is the case since

$$E[U_{a,b}] \leq \frac{1}{b-a} \sup_n E[\overbrace{(X_n - a)^-}^{\leq |X_n - a|}] \leq \frac{1}{b-a} \sup_n E[|X_n| + |a|] < \infty$$

which is finite since $X.$ is $\mathcal{L}^1$-bounded. Therefore, X_∞ exists a.s. Next,

$$E[|X_\infty|] \leq \liminf_{n\to\infty} E[|X_n|] \leq \sup_n E[|X_n|] < \infty$$

by Fatou's Lemma 1.45. Note: this does *not* imply that $X_n \to X_\infty$ in $\mathcal{L}^1$! □

Example 6.20. **Random walk**. ********************

Example 6.21. **Dirichlet problem / harmonic functions**. **********

6.3 Uniformly Integrable Martingales

Theorem 6.22. *Let $X.$ be a stochastic process on $(\Omega, \mathcal{F}, \mathcal{A}_n, P)$ and set $\mathcal{A}_\infty := \sigma(\bigcup_{n\geq 0} \mathcal{A}_n)$. Then*

1. *(X_n is an $\mathcal{A}_n$-martingale and X_n is uniformly integrable) $\iff$ $\exists X \in \mathcal{L}^1(\mathcal{F})$ such that $X_n = E[X|\mathcal{A}_n]$ a.s.*
2. *In the case that the above (equivalent) conditions hold, then $X_n \to X_\infty$ a.s. (and in $\mathcal{L}^1$); moreover, $X_\infty = E[X|\mathcal{A}_\infty]$ a.s.*

Proof. We prove ($1 \Rightarrow$) first. Suppose $(X_n)_{n\geq 0}$ is uniformly integrable $\Rightarrow (X_n)$ is $\mathcal{L}^1$-bounded, so by the Martingale Convergence Theorem 6.18 $\lim X_n =: X_\infty$ exists a.s. and $X_n \to X_\infty$ in $\mathcal{L}^1$. WWTS $X_n = E[X_\infty|\mathcal{A}_n]$ a.s.

Let $A_n \in \mathcal{A}_n$ (for n fixed). Then

$$\begin{aligned} E[\mathbf{1}_{A_n} \cdot X_\infty] &= E\left[\mathbf{1}_{A_n} \cdot \lim_{k\to\infty} X_k\right] = E\left[\lim_{k\to\infty} (\mathbf{1}_{A_n} \cdot X_k)\right] \\ &= \lim_{k\to\infty} E[\mathbf{1}_{A_n} X_k] = \lim_{k\to\infty} E[\underbrace{E[\mathbf{1}_{A_n} X_k|\mathcal{A}_n]}_{=\mathbf{1}_{A_n} X_n}] \\ &= E[\mathbf{1}_{A_n} X_n] \end{aligned}$$

and so $X_n = E[X_\infty|\mathcal{A}_n]$.

Next, we prove ($1 \Leftarrow$). Let $X \in \mathcal{L}^1(\mathcal{F})$ and set $X_n := E[X|\mathcal{A}_n]$. Then (X_n) is a martingale; WWTS (X_n) is unif. int. Observe

$$|X_n| \leq |E[X|\mathcal{A}_n]| \leq E[|X||\mathcal{A}_n] \text{ a.s.}$$

which implies

$$\begin{aligned} E[|X_n|; |X_n| \geq c] &\leq E[E[|X||\mathcal{A}_n]; \underbrace{|X_n| \geq c}_{\in \mathcal{A}_n}] = E[|X|; |X_n| \geq c] \\ &= E[|X|; |X_n| \geq c, |X| \geq a] + E[|X|; |X_n| \geq c, |X| < a] \\ &\leq E[|X|; |X| \geq a] + a\underbrace{P[|X_n| \geq c]}_{\leq E[|X_n|]\cdot\frac{1}{c}} \\ &\leq E[|X|; |X| \geq a] + \frac{a}{c}E[|X|] < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \end{aligned}$$

for a large enough and then for c large enough (given fixed a). This all implies $(|X_n|)_{n\geq 0}$ is unif. int.

Next, we prove (2). WWTS $X_\infty = E[X|\mathcal{A}_\infty]$ a.s. This is true $\iff$ $E[\mathbf{1}_A X] = E[\mathbf{1}_A X_\infty]$ for all $A \in \mathcal{A}_\infty$. To show this, let $A_k \in \mathcal{A}_k$. Then

$$E[X\mathbf{1}_{A_k}] = E[E[X|\mathcal{A}_n]\mathbf{1}_{A_k}] = E[X_n \mathbf{1}_{A_k}]$$

which implies

$$E[X\mathbf{1}_{A_k}] = \lim_{n\to\infty} E[X_n \mathbf{1}_{A_k}] = E[X_\infty \mathbf{1}_{A_k}]$$

since $(X_n \mathbf{1}_{A_k})_{n\geq 0}$ is unif. int. Next, set

$$\mathcal{D} = \{A \in \mathcal{A}_\infty : E[X\mathbf{1}_A] = E[X_\infty \mathbf{1}_A]\}$$

Notice $\mathcal{D}$ is a Dynkin system, $\bigcup_k \mathcal{A}_k \subseteq \mathcal{D}$ and $\cap$-closed, so $\mathcal{D} = \sigma(\bigcup_k \mathcal{A}_k) = \mathcal{A}_\infty$. Thus, $E[X|\mathcal{A}_\infty] = X_\infty$ a.s. □

Corollary 6.23 (0-1 Law of Levy)**.** *Let $A \in \mathcal{A}_\infty$. Then*

$$\lim_{n\to\infty} P[A|\mathcal{A}_n] = \mathbf{1}_A \textit{ a.s.}$$

Proof. Set $X = \mathbf{1}_A$ and $X_n = E[X|\mathcal{A}_n]$. THen (X_n) is a uniformly integrable martingale, so $X_n \to X_\infty$ a.s. with $X_\infty \in \mathcal{A}_\infty$, and $E[X|\mathcal{A}_\infty] = X_\infty$. Thus, $\mathbf{1}_A = X_\infty$ a.s., since $X \in \mathcal{A}_\infty$. □

Remark 6.24. The 0-1 Law of Kolmogorov 1.20 follows! Let $\mathcal{B}_1, \mathcal{B}_2, \ldots$ be independent σ-fields and

$$A \in \bigcap_{n\geq 0} \sigma\left(\bigcup_{k\geq n} \mathcal{B}_k\right) =: \tau \quad \text{(tail field)}$$

Then $P[A] = 0$ or $P[A] = 1$. To see why, set

$$\mathcal{A}_n := \sigma\left(\bigcup_{k=1}^{n} \mathcal{B}_k\right)$$

Then $A \in \tau \subseteq \mathcal{A}_\infty$. Thus,

$$\lim_{n\to\infty} \underbrace{P[A|\mathcal{A}_n]}_{=P[A]} = \mathbf{1}_A \text{ a.s.}$$

But since $A \in \sigma\left(\bigcup_{k\geq n+1} \mathcal{B}_k\right)$ (which is *independent* of $\mathcal{B}_1, \ldots, \mathcal{B}_n$) we have $P[A] = \mathbf{1}_A(\omega)$ with $P[A]$ constant! This is only possible if $P[A] = 0$ or 1.

Theorem 6.25. *1. If (X_n) is $\mathcal{L}^p$-bounded for $p > 1$ then (X_n) is unif. int and X_∞ exists a.s. and $X_n \to X$ in $\mathcal{L}^p$.*

2. Also, if $X \in \mathcal{L}^p(\mathcal{F})$ then $X_n := E[X|\mathcal{A}_n]$ is a $\mathcal{L}^p$-bounded martingale.

Proof. (1) is in the text. (2) is proven by Jensen. □

6.4 Further Applications of Martingale Convergence

6.4.1 Martingales with $\mathcal{L}^1$-dominated increments

Theorem 6.26. *Let X be a martingale such that*

$$\sup_n |\Delta_n X| \leq Y \in \mathcal{L}^1$$

Set

$$\begin{aligned} C &:= \{\omega : X_n(\omega) \text{ converges to a real } \#\} \\ O &:= \left\{\omega : \liminf_{n\to\infty} X_n(\omega) = -\infty, \limsup_{n\to\infty} X_n(\omega) = +\infty\right\} \\ &= \{\omega : \inf X_n(\omega) = -\infty, \sup X_n(\omega) = +\infty\} \quad \text{(for discrete time)} \end{aligned}$$

Then $P[C \dot\cup O] = 1$.

Proof. Let $a \in \mathbb{Z}$ and set $T_a = \inf\{n \geq 0 : X_n \leq a\}$. Then

$$X_{T_a \wedge n} = \begin{cases} X_0 & \text{on } \{X_0 \leq a\} \\ X_n > a & \text{on } \{X_0 > a, n < T_a\} \\ X_{T_n} \geq a - \sup_k |\Delta_k X| & \text{on } \{X_0 > a, n \geq T_a\} \end{cases}$$
$$\geq X_0 \wedge (a - \underbrace{\sup_n |\Delta_n X|}_{\in \mathcal{L}^1}) \in \mathcal{L}^1$$

and so $X_{T_a \wedge n}$ is an $\mathcal{L}^1$-bounded martingale; therefore, $X_n^{T_a} \to$ finite limit a.s. □

Claim:

$$\left\{\inf_n X_n > -\infty\right\} \subseteq C \text{ a.s.}$$

WWTS

$$\left\{\inf_n X_n > a\right\} \subseteq C \text{ a.s.} \quad \forall a$$

which implies

$$C \supseteq \bigcup_k \left\{\inf_n X_n > -k\right\} = \left\{\inf_n X_n > -\infty\right\}$$

If $\inf_n X)n(\omega) > a$ then $T_a(\omega) = +\infty$, so

$$X_{T_a \wedge n}(\omega) = X_n(\omega) \xrightarrow[n\to\infty]{} X_\infty(\omega) \in \mathbb{R} \text{ for a.e. } \omega$$

so $\omega \in C$ (for a.e. ω). Similarly,

$$\left\{\sup_n X_n < \infty\right\} \subseteq C \text{ a.s.}$$

which implies

$$C^c \subseteq \left\{\inf_n X_n = -\infty\right\} \cap \left\{\sup_n X_n = +\infty\right\}$$

6.4.2 Generalized Borel-Cantelli II

This is a more general statement than Lemma 1.18.

Lemma 6.27. *Suppose $\mathcal{A}$ is a filtration with $A_n \in \mathcal{A}_n$. Define*

$$A_\infty := \bigcap_{n\geq 1} \bigcup_{k \geq n} A_k = \left\{\omega : \sum_{k\geq 1} \mathbf{1}_{A_k}(\omega) = +\infty\right\}$$

to be the event that ∞-many of the A_ks occur. Set

$$A'_\infty := \{\omega : \sum_{k\geq 1} \underbrace{P[A_k|\mathcal{A}_{k-1}]}_{=E[\mathbf{1}_{A_k}|\mathcal{A}_{k-1}]}(\omega) = \infty\}$$

Then $A_\infty = A'_\infty$ a.s.

Note: no *independence* is required!

Remark 6.28. If A_k independent of $\mathcal{A}_{k-1}$ for all k, then $P[A_k|\mathcal{A}_{k-1}] = 0$ or 1 (constant), so then

$$A_\infty = A'_\infty = \Omega \text{ or } \emptyset \text{ a.s.}$$

By the original Borel-Cantelli Lemmas, $P[A_\infty] = 1$ if $\sum P[A_k] = \infty$ and $P[A_\infty] = 0$ if $\sum P[A_k] < \infty$. (Typically, $\mathcal{A}_k = \sigma(\mathbf{1}_{A_1}, \dots, \mathbf{1}_{A_k})$).

Proof. Let $X_0 = 0$ and $\mathcal{A}_0 = (\emptyset, \Omega)$. Define

$$X_n := \sum_{k=1}^{n} (\mathbf{1}_{A_k} - E[\mathbf{1}_{A_k}|\mathcal{A}_{k-1}])$$

This *is* a martingale with respect to $\mathcal{A}.$, since

$$E[\Delta_n X|\mathcal{A}_{n-1}] = E[\mathbf{1}_{A_n} - \underbrace{E[\mathbf{1}_{A_n}|\mathcal{A}_{n-1}]}_{\in \mathcal{A}_{n-1}}|\mathcal{A}_{n-1}] \equiv 0$$

Since X has bounded increments ($-1 \le \Delta_n X \le 1$) then $P[C \cup O] = 1$. WWTS that A_∞ and A'_∞ agree a.s. on C and on O.

- On C, we have

$$\sum_k \mathbf{1}_{A_k} = \infty \iff \sum_k P[A_k|\mathcal{A}_{k-1}] = \infty$$

  since otherwise $X_n \not\to \cdot \in \mathbb{R}$. Thus, $C \cap A'_\infty = C \cap A_\infty$ a.s.

- On O, we have $\sum_k \mathbf{1}_{A_k}(\omega) = \infty$ (since otherwise $\sup X_k \neq \infty$) and similarly $\sum_k P[A_k|\mathcal{A}_{k-1}] = \infty$ (otherwise $\inf X_k \neq -\infty$). Thus, $O \subseteq A_\infty$ and $O \subseteq A'_\infty$ so $O \cap A_\infty = O = O \cap A'_\infty$.

Since $C \cup O = \Omega$ a.s., then $A_\infty = A'_\infty$ a.s. □

Example 6.29. James' example revisited. Set $X_0 = X_1 = 1$ and $X_k = 0$ or 1 and $S_k = \sum_{i=0}^{k} X_i$, with

$$P[X_{k+1} = 1|\sigma(X_0, \dots, X_k)](\omega) = \frac{1}{S_k(\omega)} \quad \forall k \ge 0$$

Recall: this is an example of a sequence such that $X_k \to 0$ in probability but *not* a.s. ($X_k = 1$ for ∞-many k a.s.) We showed this by explicit calculations (by using a discrete process with geometric waiting times ...).

Now, we let

$$A_k = \{X_k = 1\}; \text{ and } A_\infty = \{X_k = 1 : \text{for } \infty\text{-many } k\}$$

and

$$A'_\infty = \left\{\omega : \sum_{k \ge 0} P[X_k = 1|\mathcal{A}_{k-1}](\omega) = \infty\right\} = A_\infty$$

Observe that

$$P[X_k = 1|\mathcal{A}_{k-1}](\omega) = \frac{1}{S_{k-1}(\omega)}$$

and so

$$\sum_{k=0}^{\infty} P[X_k = 1|\mathcal{A}_{k-1}](\omega) = 1 + \sum_{k=1}^{\infty} \underbrace{\frac{1}{S_{k-1}(\omega)}}_{\geq 1/k} = \infty$$

Thus, $A'_\infty = \Omega = A_\infty$ a.s. Thus, $X_k = 1$ for ∞-many k a.s.

6.4.3 Branching processes

Model: Let $Y_{n,k} \in \mathbb{N}$ $(k, n = 1, 2, \dots)$ be independent RVs on $(\Omega, \mathcal{F}, P)$, all i.i.d. with distribution μ (with $\mu \neq \delta_k$ for $k = 0, 1, 2, \dots$) Assume $\infty > m = \sum_k k\mu_k$ to be the finite mean. Set $X_0 = 1$ and

$$X_n = Y_{n,1} + Y_{n,2} + \cdots + Y_{n,X_{n-1}}$$

We think of X_n as the number of individuals in the nth generation, where $Y_{n,k}$ is the number of children that the kth individual in the previous generation produced, and that all individuals die when the next generation is produced. Set

$$\mathcal{A}_n = \sigma\left(Y_{\ell,k} : 1 \leq \ell \leq n, k = 1, 2, \dots\right)$$

and note that this is bigger than just knowing the n children.

Lemma 6.30. *$M_n := \frac{X_n}{m^n}$ is a martingale with $M_n \geq 0$, so $M_n \to M_\infty$ a finite limit P-a.s.*

Proof. Assuming $M_n \in \mathcal{L}^1$, then

$$\begin{aligned} E\left[\frac{X_{n+1}}{m^{n+1}}|\mathcal{A}_n\right](\omega) &= \frac{1}{m^{n+1}} E\left[\sum_{k=1}^{X_n} Y_{n+1,k}|\mathcal{A}_k\right](\omega) \\ &= \frac{1}{m^{n+1}} E\left[\sum_{k=1}^{X_n(\omega)} Y_{n+1,k}\right] \text{ a.s.} \\ &= \frac{1}{m^{n+1}} \cdot m \cdot X_n(\omega) = \frac{X_n(\omega)}{m^n} = M_n(\omega) \end{aligned}$$

so $M.$ is indeed a martingale.

Why is $M_n \in \mathcal{L}^1$? This happens $\iff$ $X_n \in \mathcal{L}^1$, and for $n \geq 1$

$$E[X_n] = E\left[\sum_{k=1}^{X_{n-1}} Y_{n,k}\right] = m \cdot E[X_{n-1}]$$

by Wald's Identity since the Ys are independent from X_{n-1}. Thus, $E[X_n] = m^n$. $\square$

Definition 6.31. *Let*

$$T(\omega) = \min\{n \geq 0 : X_n(\omega) = 0\}$$

be the "time of extinction".

Using this definition, we have

$$\begin{aligned} E[S_T^2] &= E[E[S_T^2|\sigma(T)]] \\ &= E\left[E\left[\right]\right] \end{aligned}$$

6.5 Sub and supermartingales

Let (X_n) be an $\mathcal{A}_n$-adapted $\mathcal{L}^1$ process. Set $\Delta_k X = X_k - X_{k-1}$ so then $X_n = X_0 + \sum_{k=1}^n \Delta_k X$.

Lemma 6.32. *(M_n) given by*

$$M_n := X_0 + \sum_{k=1}^n (\underbrace{\Delta_k X - E[\Delta_k X|\mathcal{A}_{k-1}]}_{\Delta_k M})$$

is a martingale.

Proof. Notice $M_0 = X_0 \in \mathcal{L}61$ and $E[\Delta_k M|\mathcal{A}_{k-1}] \equiv 0$ a.s. □

Theorem 6.33 (Doob decomposition)**.** *Let X_n be adapted and $\mathcal{L}^1$. Then $\exists!$ decomposition $X_n = M_n + A_n$ where $M.$ is a martingale and $A.$ is previsible with $A_0 \equiv 0$.*

Proof. First, existence. Using M from the previous lemma:

$$M_n = \underbrace{X_0 + \sum_{k=1}^n \Delta_k X}_{=X_n} - \underbrace{\sum_{k=1}^n E[\Delta_k X|\mathcal{A}_{k-1}]}_{=:A_n}$$

and one can check that A_n is indeed previsible. Note: $\Delta_n A = E[\Delta_n X|\mathcal{A}_{n-1}]$.

Second, uniqueness. Suppose $X_n = \bar{M}_n + \bar{A}_n$. Then

$$E[\Delta_k X|\mathcal{A}_{k-1}] = \underbrace{E[\Delta_k \bar{M}|\mathcal{A}_{k-1}]}_{=0} + \underbrace{E[\Delta_k \bar{A}|\mathcal{A}_{k-1}]}_{=\Delta_k \bar{A}}$$

which implies $\Delta_k \bar{A} = E[\Delta_k X|\mathcal{A}_{k-1}]$. Thus, we have no choice for $\bar{A}$! Then $\bar{M} = X - \bar{A}$ is also uniquely determined. □

Definition 6.34. *Let (X_n) be a stochastic process. It is called a sub (resp. super) martingale if $X_k \in \mathcal{L}^1$ and adapted, and*

$$X_n \leq E[X_{n+1}|\mathcal{A}_n] \text{ a.s.} \iff 0 \leq E[\Delta_{n+1} X|\mathcal{A}_n] \text{ a.s.}$$

(resp. $\geq$).

Note: the inequality condition is equivalent to

$$0 \le A_1 \le A_2 \le \cdots \le A_n \text{ in Doob decomposition}$$

for submartingale ($\ge$ for super).

Example 6.35.

- (X_n) is both a sub and supermartingale $\iff$ it's a martingale.
- (X_n) is a submartingale $\iff$ $(-X_n)$ is a supermartingale.
- If X_n is a martingale and u is convex (resp. concave) then $u(X_n)$ is a sub (resp. super) martingale. To see why, observe that

  $$E[u(X_{n+1})|\mathcal{A}_n] \ge u(E[X_{n+1}|\mathcal{A}_n]) = u(X_n) \text{ a.s}$$

  by Jenesen.
- Let (X_n) be an adapted process. If $\exists\lambda \in \mathbb{R}$ with $\exp(\lambda X_0) \in \mathcal{L}^1$ and $E[\exp(\lambda\Delta_k X)|\mathcal{A}_{k-1}] \le 1$ for every k, then $\exp(\lambda X_n)$ is a supermartingale. Additionally, if X_n is a martingale then $\exp(\lambda X_n)$ is a martingale. To prove the first claim, observe that

  $$E[\exp(\lambda X_{n+1}|\mathcal{A}_n] = \exp(\lambda X_n)E[\underbrace{\exp(\lambda\Delta_{n+1}X)}_{\le 1}|\mathcal{A}_n] \le \exp(\lambda X_n)$$

  To prove the second claim, just notice that $\exp(\lambda X_n)$ would also be a submartingale since $\exp(\lambda t)$ is convex.

Theorem 6.36 (Supermartingale convergence)**.** *Let (X_n) be a supermartingale with $\sup E[X_n^-] < \infty$. Then $\lim X_n =: X_\infty$ exists a.s. and $X_\infty \in \mathcal{L}^1$.*

Proof. Let $X_n = M_n - A_n$ (where $0 = A_0 \le A_1 \le A_2 \le \cdots$). Then $M_n = X_n + A_n$ implies $M_n \ge X_n$ so $M_n^- \le X_n^-$ and thus $\sup_n E[M_n^-] < \infty$. This implies $M_n \to M_\infty \in \mathcal{L}^1$ a.s.

Next, $A_n = M_n - X_n \le M_n + X_n^-$, so

$$E[A_n] \le E[M_0] + E[X_n^-] \;\Rightarrow\; E[\underbrace{\lim \nearrow A_n}_{=A_\infty}] \le E[X_0] + \underbrace{\liminf E[X_n^-]}_{<\infty} \in \mathcal{L}^1$$

by Fatou. Thus, $X_\infty := M_\infty - A_\infty \in \mathcal{L}^1$. □

More on stopping times: If S, T are stopping times, then $S\wedge T, S\vee T, T\wedge n, S+T$ are all stopping times (with respect to the same filtration).

Definition 6.37. *Let T be a stopping time with respect to $\mathcal{F}_\cdot$. Then*

$$\mathcal{F}_T := \{A \in \mathcal{F} : A \cap \{T-k\} \in \mathcal{F}_k \;\forall k \ge 0\}$$

is the collection of "up to time T observable events".

Note: $A \in \mathcal{F}_T \iff \forall k,\ A \cap \{T \le k\} \in \mathcal{F}_k$. Interpretation: $A \cap \{T \le k\} \in \mathcal{F}_k\ \forall k$ if $\forall k$ that part of the event A where the stopping time occurred before time k is $\in \mathcal{F}_k$ (i.e.is observable at time k).

Lemma 6.38. *$\mathcal{F}_{S \wedge T} = \mathcal{F}_S \cap \mathcal{F}_T$. $X_\cdot$ adapted $\Rightarrow X_T$ is $\mathcal{F}_T$-measurable.*

Proof. (***) homework. □

Theorem 6.39 (Stopping time for unif. int. martingales)**.** *Let $S \le T$ be stopping times and $X_\cdot$ a martingale.*

1. *If $X_\cdot$ is unif. int. then $X_\cdot^T$ is unif. int. and $E[X_0] = E[X_T]$.*
2. *If X is unif. int. with $E[X|\mathcal{F}_n] = X_n$ a.s. $\forall n$, then $E[X|\mathcal{F}_T] = X_T$ a.s.*
3. *If $X_\cdot^T$ is unif. int. and $S \le T$ then $X_\cdot^S$ is unif. int.*

Proof.

1. Assume (X_n) is unif. int. $\Rightarrow \exists X \in \mathcal{F}, X \in \mathcal{L}^1$ such that $X_n = E[X|\mathcal{F}_n]$ a.s. WWTS $(X^T)_n = X_{T\wedge n} = E[X|\mathcal{F}_{T\wedge n}]$ a.s. which implies $X_\cdot^T$ is also a unif. int. (by successive prognosis) martingale (with respect to another filtration $(\mathcal{F}_{T\wedge n})_{n\ge 0}$) and that, in particular, it is unif. int. To show this, it suffices to prove (2) and then apply it to the stopping time $T \wedge n$ (instead of T).

2. Let $A \in \mathcal{F}_T$. WWTS $E[X\mathbf{1}_A] = E[X_T\mathbf{1}_A]$. To see why, notice that

$$E[X\mathbf{1}_A] = \sum_{k=0}^{\infty} E[\underbrace{\underbrace{\mathbf{1}_{\{T=k\}}\mathbf{1}_A}_{=\mathbf{1}_{A\cap\{T=k\}}}}_{\in \mathcal{F}_k} X]$$

$$= \sum_{k=0}^{\infty} E[\mathbf{1}_{A\cap\{T=k\}} \underbrace{E[X|\mathcal{F}_k]}_{=X_k}]$$

$$= \sum_{k=0}^{\infty} E[\underbrace{\mathbf{1}_{\{T=k\}}(\mathbf{1}_A X_k)}_{=\mathbf{1}_{\{T=k\}}\mathbf{1}_A X_T}]$$

$$= \sum_{k=0}^{\infty} E[\mathbf{1}_{\{T=k\}}(\mathbf{1}_A X_T)] = E[\mathbf{1}_A X_T]$$

3. $Y := X^T$ is unif. int. $\Rightarrow Y^S$ is unif. int. by (1), so

$$(Y^S)_n = Y_{S\wedge n} = X_{T\wedge(S\wedge n)} = X_{(T\wedge S)\wedge n} = X_{S\wedge n} = (X^S)_n$$

□

Theorem 6.40 (Optional Stopping Theorem)**.** *Let $(X_n) = M_n - A_n$ be a supermartingale and T, S be stopping times with $T \le S$.*

1. *If (T, S bounded) OR ($S < \infty$a.s. and M^S, M^T are unif. int.) then*

$$E[X_0] \geq E[X_T] \geq E[X_S]$$

2. *If $X_n \geq Y \in \mathcal{L}^1$ for all n and T is any stopping time then $E[X_0] \geq E[X_T]$ (where $X_T := X_\infty$ on $\{T = \infty\}$).*

Proof. Write $X_n = M_n - A_n$ (with $A_n \geq 0$). Then $E[M_T] = E[M_0] = E[M_S]$ implies

$$E[X_T] = E[M_T] - E[A_T] \geq E[M_S] + E[A_S]$$

Next,

$$E[X_T] = E[\lim X_{T\wedge n}] \leq \liminf \underbrace{E[X_{T\wedge n}]}_{=E[M_{T\wedge n} - E[A_{T\wedge n}]} \leq E[X_0]$$

□

Example 6.41 (Applications to microeconomics). Following a game $g(x)$ given. A random walk starts at $\bar{x} \in (a, b) \cap \mathbb{Z}$ and will be stopped at the boundary a, b. Write $X_n := \bar{x} + Y_1 + \cdots + Y_n$ as the random walk and set

$$S = \min\{n \geq 0 : X_n \in \{a, b\}\}$$

Our process is $X_n^S := X_{S\wedge n}$ which is a martingale. We are looking for a stopping time T such that $E[g(X_T^S)]$ is maximal.

The solution is to let h be the concave envelope of g (i.e. the smallest concave majorant of g). Then $h(X_n^S)$ is a supermartingale which implies

$$h(\bar{x}) = h(X_0^S) \geq E[h(X_T^S)] \geq E[g(X_T^S)]$$

for each stopping time T, so $h(\bar{x})$ is an upper bound on the expected gain with any strategy T.

Claim: if we set

$$T^\star = \min\{n \geq 0 : X_n \in \{g = h\}\}$$

then $E[g(X_{T^\star}^S)] = h(\bar{x})$. (The optimal solution!) To prove this, if $h \equiv g$ then we stop at $t = 0$ and we have $h(\bar{x})$ deterministically. So let $h(\bar{x}) > g(\bar{x})$. Then $\exists \bar{x} \in [c, d] \subseteq [a, b]$ such that $h(y) > g(y)$ on $y \in (c, d)$ and $h(c) = g(c)$ and $h(d) = g(d)$. Thus, $T^\star$ is an exit time from (c, d). Now h is linear (convex and concave) on $[c, d]$ and

$$\Lambda_{T^\star\wedge n}^S = X_{T^\star\wedge n}$$

is a martingale, so $h(X_{T^\star\wedge n}^S)$ is a martingale so

$$h(\bar{x}) = E[h(X_{T^\star\wedge n})] = E[h(X_{T^\star})] = E[g(X_{T^\star})]$$

****

6.6 Maximal inequalities

Lemma 6.42. *If $X_k \geq 0$ is a* super*martingale then*

$$P\left[\sup_n \geq 0 X_n \geq c\right] \leq \frac{1}{c} E[X_0]$$

If $X_k \geq 0$ is a sub*martingale then*

$$P\left[\max_{k \leq N} X_k \geq c\right] \leq \frac{1}{c} E\left[X_N; \max_{k \leq N} X_k \geq c\right] \leq \frac{1}{c} E[X_N]$$

Proof. Assume X_k is a supermartingale. We know $X_n \to X_\infty$ a.s. and $X_\infty \geq 0$ and $\mathcal{L}^1$. Let

$$T_c := \min\{n \geq 0 : X_n \geq c\}$$

Then

$$E[X_0] \geq E[X_{T_c}] \geq E[X_{T_c}; T_c < \infty] \geq cP[T_c < \infty]$$

and so

$$\begin{aligned} E[X_0] &\geq \left(c - \frac{1}{n}\right) P\left[T_{c-\frac{1}{n}} < \infty\right] \ (n \to \infty) \\ &\geq cP[\bigcap_n \underbrace{\{T_{c-\frac{1}{n}} < \infty\}}_{\supseteq \{\sup_n X_n \geq c\}}] \\ &\geq cP[\sup_n X_n \geq c] \end{aligned}$$

Next, assume X_k is a submartingale. Then

$$\begin{aligned} cP\left[\max_{k \leq N} X_k \geq c\right] &= cP[T_c \leq N] = E[c; T_c \leq N] \\ &\leq E[X_{T_c}; T_c \leq N] = \sum_{k=0}^{n} E[\underbrace{X_k}_{\leq E[X_N|\mathcal{A}_k] \text{ a.s.}}; T_c = k] \\ &\leq \sum_{k=0}^{n} E\left[E[X_N|\mathcal{A}_k] \cdot \mathbf{1}_{\{T_c = k\}}\right] \\ &= \sum_{k=0}^{n} E[X_N; T_c = k] = E[X_N; T_c \leq N] \\ &= E[X_n; \max_{k \leq N} X_k \geq c] \end{aligned}$$

□

Corollary 6.43. *Let (M_n) be a martingale with $M_n \in \mathcal{L}^p$ for $p \geq 1$. Then*

$$P\left[\max_{k \leq N} |M_k| \geq c\right] \leq \frac{1}{c^p} E[|M_N|^p] \ \forall c > 0$$

Proof. $|M_n|^p$ is a submartingale (by Jensen), so

$$P\left[\max_{k\le N}|M_k|\ge c\right]=P\left[\max_{k\le N}|M_k|^p\ge c^p\right]$$

etc. □

Example 6.44. Application for insurances. How expensive should an insurance policy be? Let x_0 be the starting capital of the company (deterministic), c_n be the deterministic income, $Y_n(\omega)$ be the stochastic loss, and

$$X_n(\omega)=X_{n-1}(\omega)+\underbrace{c_n-Y_n(\omega)}_{=:\Delta_n X(\omega)}$$

be the balance of the company. Let $R(\omega)$ be the time of ruin. Then $P[R<\infty]\le ?$

Let c_n be big enough such that for some $\lambda>0$,

$$E\left[\exp(\lambda(Y_n-c_n))|\mathcal{A}_{n-1}\right]\le 1$$

Is this realistic? If Y_n is independent of $\mathcal{A}_{n-1}$ then

$$E[\exp(\lambda(Y_n-c_n))|\mathcal{A}_{n-1}]=E[\exp(\lambda(Y_n-c_n))]$$

and so the condition above means $E[\exp(\lambda Y_n)]\le\exp(\lambda c_n)$.

With the condition above, then $(\exp(-\lambda X_n))$ is a supermartingale. To see why, notice that

$$\begin{aligned}E[\exp(-\lambda X_n)|\mathcal{A}_{n-1}]&=\exp(-\lambda\Delta_n X)E[\exp(-\lambda X_{n-1})|\mathcal{A}_{n-1}]\\&\le 1\cdot\exp(-\lambda X_{n-1})\end{aligned}$$

Note that $X_n=0\iff\exp(-\lambda X_n)=1$. Then

$$\begin{aligned}P[R<\infty]=P[T_1<\infty]&\le P\left[\sup_n\exp(-\lambda X_n)\ge 1\right]\\&\le\frac{1}{1}E[\exp(-\lambda X_0)]=\exp(-\lambda X_0)\end{aligned}$$

by the maximal inequality 6.42, so choosing λ large enough (or X_0) will make $P[R<\infty]$ small.

Theorem 6.45. *Let (X_n) be an $\mathcal{L}^p$-bounded martingale for $p>1$ and let $X^\star=\sup_n|X_n|$. Then*

$$\|X^\star\|_p\le\frac{p}{p-1}\cdot\sup_n\|X_n\|_p$$

If X_n is a martingale with bounded entropy, *i.e.*

$$\sup_n E[|X_n|\log|X_n|]<\infty$$

then $X^\star\in\mathcal{L}^1$.

Before the proof we have a lemma.

Lemma 6.46. *Let* $X, Y \geq 0$. *If* $\forall c \geq 0$, $c[PY \geq c] \leq E[X; Y \geq c]$, *then* $\forall f \geq 0$ *with* $F(y) = \int_0^y f(x)\,dx$, *we have*

$$E[F(Y)] \leq E[X \cdot \int_0^Y \frac{1}{c} f(c)\,dc]$$

Proof. Observe that

$$\begin{aligned} E[F(Y)] &= E[\int_0^\infty \underbrace{\mathbf{1}_{[0,Y(\omega)]}(c)}_{=\mathbf{1}_{\{Y \geq c\}}(\omega)} f(c)\,dc] \\ &= \int_0^\infty P[Y \geq c] f(c)\,dc \\ &\leq \int_0^\infty f(c) \frac{1}{c} E[X; Y \geq c]\,dc \\ &= E[X \cdot \int_0^Y \frac{1}{c} f(c)\,dc] \end{aligned}$$

□

In particular, for $F(y) = y^p$, we have $f(c) = pc^{p-1}$ for $p > 1$, and so

$$\begin{aligned} E[Y^p] &\leq E[X \int_0^Y \frac{1}{c} pc^{p-1}\,dc] = E[X \int_0^Y pc^{p-2}\,dc] \\ &= E[XY^{p-1} \frac{p}{p-1}] = \frac{p}{p-1} E[Y^{p-1} X] \\ &\leq \frac{p}{p-1} \|X\|_p \|Y^{p-1}\|_q \end{aligned}$$

by Hölder, where $q = \frac{p}{p-1}$. Notice $\|Y^{p-1}\|_q = E[Y^p]^{\frac{p-1}{p}}$, so we can divide through by this factor and obtain

$$\|Y\|_p = E[Y^p]^{1/p} \leq \frac{p}{p-1} \|X\|_p$$

Now we're ready to prove the theorem above.

Proof. First, notice $Z_n = |X_n|$ is a submartingale (≥ 0) so

$$cP[\max_{k \leq N} |X_k| \geq c] \leq E[\underbrace{|X_N|}_{X}; \underbrace{\max_{k \leq N} X_k \geq c}_{Y \geq c}]$$

and thus

$$\left\| \max_{k \leq N} |X_k| \right\|_p \leq q \, \|X_N\|_p$$

Then,

$$\begin{aligned}\|X^\star\|_p &= E\left[\left(\lim_N \max_{k\le N}|X_k|\right)^p\right]^{1/p} \\ &= E\left[\lim_N \left(\max_{k\le N}|X_k|\right)^p\right]^{1/p} \\ &= \lim_N E\left[\left(\max_{k\le N}|X_k|\right)^p\right]^{1/p} \\ &\le q\sup_N \|X_N\|_p\end{aligned}$$

□

6.7 Backwards martingales

Let $(\mathcal{A}_n) \nearrow$ for $n \le 0$ (!) on $(\Omega, \mathcal{F}, P)$, i.e.

$$\cdots \subseteq \mathcal{A}_{-2} \subseteq \mathcal{A}_{-1} \subseteq \mathcal{A}_0 \subseteq \mathcal{F}$$

Then X_n is a martingale provided

$$E[X_{n+k}|\mathcal{A}_n] = X_n \iff X_n = E[X_0|\mathcal{A}_n]$$

Theorem 6.47. *Set $\mathcal{A}_{-\infty} := \bigcap_{n\ge 0}\mathcal{A}_n$. Then $X_n \to X_{-\infty}$ as $n \to -\infty$ a.s. and in $\mathcal{L}^1$ and $X_{-\infty} = E[X_0|\mathcal{A}_{-\infty}]$.*

Proof. For $N < 0$,

$$E\left[U_{a,b}^{(N,0]}\right] \le \frac{E\left[(X_0-a)^-\right]}{b-a}$$

so

$$E[U_{a,b}] = E\left[\lim_{N\to-\infty} U_{a,b}^{(N,0]}\right] = \lim_{N\to-\infty} E\left[U_{a,b}^{(N,0]}\right] \le \frac{E[(X_0-a)^-]}{b-a} < \infty$$

by monotone integrability and since $X_0 \in \mathcal{L}^1$. Thus, $P[U_{a,b} < \infty] = 1$ so $X_n \to X_{-\infty}$ a.s. By Fatou,

$$E[|X_{-\infty}|] \le \liminf_{n\to\infty} E[|X_n|] \le E[|X_0|]$$

and so $X_{-\infty} \in \mathcal{L}^1$. But also, (X_n) is unif. int. so $X_n \to X_{-\infty}$ in $\mathcal{L}^1$. Moreover, for $A \in \mathcal{A}_{-\infty}$,

$$E[X_{-\infty}\mathbf{1}_A] = \lim_{n\to\infty} E[X_{-n}\mathbf{1}_A] = E[X_0\mathbf{1}_A]$$

and so $X_{-\infty} = E[X_0|\mathcal{A}_{-\infty}]$ a.s. □

Corollary 6.48 (Law of large numbers)**.** *Let $Y_1, Y_2, \dots \in \mathcal{L}^1$ be i.i.d. and set $S_n = \sum_{i=1}^n Y_i$. Then $\frac{1}{n}S_n \to E[Y_1]$ a.s.*

Proof. By symmetry $E[Y_i|S_n] = E[Y_j|S_n]$, so $\frac{1}{n}S_n = E[Y_1|\sigma(S_n)]$. Also,

$$\frac{1}{n}S_n = E[Y_1|\sigma(S_n, Y_{n+1}, \dots,)] = E[Y_1|\underbrace{\sigma(S_n, S_{n+1}, \dots)}_{=:\mathcal{A}_{-n}}]$$

Then $\mathcal{A}_{-n} \searrow$ as $n \nearrow$. Therefore,

$$X_{-n} := E[Y_1|\mathcal{A}_{-n}] = \frac{1}{n}S_n$$

is a martingale, which implies $\lim_{n\to\infty} \frac{1}{n}S_n = X_{-\infty}$ exists a.s. and is in $\mathcal{L}^1$. Thus,

$$\lim_{n\to\infty} \frac{1}{n}S_n = X_{-\infty} \in \tau = \bigcap_{n\geq 1} \sigma\left(\bigcup_{k\geq n} \sigma(Y_1)\right)$$

so by Kolmogorov's 0-1 law 1.20, $X_{-\infty}$ is constant a.s. $\Rightarrow X_{-\infty} = E[X_{-\infty}] = E[X_1]$ by uniform integrability. $\square$

Example 6.49. Next application: Hewitt-Savage 0-1 Law. If X_1, X_2 are i.i.d. and $A \in \mathcal{E}$ then $P[A] = 0$ or 1 where $\mathcal{E}$ is the exchangeable σ-field (**** definition)

6.8 Concentration inequalities: the Martingale method

Let X be a $\mathcal{L}^1$ RV on some filtered probability space and assume $X_n \in \mathcal{F}_n$. Set $X_k := E[X|\mathcal{F}_k]$, an $\mathcal{F}_\cdot$ martingale. *Assume* that $\forall k$, $\|\Delta_k X\|_\infty =: c_k < \infty$, i.e. the martingale has bounded increments. Then

$$P[(X_n - E[X]) \geq t] \leq \exp\left(-\frac{t^2}{2\sum_{k=1}^n c_k^2}\right) \tag{7}$$

This inequality also holds for $P[(X - E[X]) \leq -t]$. This is known as Azuma's Inequality.

Proof. Set $\mathcal{F}_0 = \{\emptyset, \Omega\}$ so $X_0 = E[X]$ and $D_k := \Delta_k X$. Then

$$\begin{aligned} E[\exp(\lambda X)] &= E[\exp(\lambda X_{n-1}) \overbrace{E[\exp(\lambda D_n)|\mathcal{F}_{n-1}]}^{\leq \|E[\exp(\lambda D_n)|\mathcal{F}_{n-1}]\|_\infty}] \\ &\leq \|E[\exp(\lambda D_n)|\mathcal{F}_{n-1}]\|_\infty \cdot E[\exp(\lambda X_{n-1})] \\ &\vdots \\ &\leq \prod_{k=1}^n \|E[\exp(\lambda D_k)|\mathcal{F}_{k-1}]\|_\infty \cdot E[\exp(\lambda X_0)] \end{aligned}$$

Dividing both sides by $E[\exp(\lambda X_0)] = E[\exp(\lambda E[X])]$ tells us

$$E[\exp(\lambda(X - E[X]))] \leq \prod_{k=1}^n \|E[\exp(\lambda D_k)|\mathcal{F}_{k-1}]\|_\infty$$

Now, if $\|D_k\|_\infty < \infty$, then

$$\|E[\exp(\lambda D_k)|\mathcal{F}_{k-1}]\|_\infty \le \exp\left(\frac{\lambda^2}{2}\|D_k\|_\infty^2\right)$$

(proof of this claim given below) and so

$$E[\exp(\lambda(X - E[X]))] \le \exp\left(\frac{\lambda^2}{2} \cdot \sum_{k=1}^{n} \|D_k\|_\infty^2\right)$$

By Chebyshev 1.47 and optimal choice of λ, we get Azuma's inequality.

Here, we prove the claim from a few lines above. Note that $-c_k \le D_k(\omega) \le c_k$ a.s. where $c_k = \|D_k\|_\infty < \infty$. Write $D_k(\omega)$ as a convex combination of $-c_k, c_k$ (and now we drop the index k): $D(\omega) = p(\omega)(-c) + (1 - p(\omega))c$, so $p = \frac{1}{2} - \frac{D}{2c}$ and $1 - p = \frac{1}{2} + \frac{D}{2c}$. Since $\exp(\lambda(\cdot))$ is *convex*, then $\exp(\lambda D) \le p\exp(-\lambda c) + (1-p)\exp(\lambda c)$, so

$$\begin{aligned}
&E[\exp(\lambda D)|\mathcal{F}_{k-1}] \\
&\quad \le \exp(-\lambda c)E[p|\mathcal{F}_{k-1}] + \exp(\lambda c)E[1-p|\mathcal{F}_{k-1}] \\
&\quad = \exp(-\lambda c)\Bigg(\frac{1}{2} - \frac{1}{2c}\underbrace{E[D|\mathcal{F}_{k-2}]}_{=0}\Bigg) + \exp(\lambda c)\Bigg(\frac{1}{2} + \frac{1}{2c}\underbrace{E[D|\mathcal{F}_{k-2}]}_{=0}\Bigg) \\
&\quad = \cosh(\lambda c) \le \exp\left(\frac{\lambda^2 c^2}{2}\right)
\end{aligned}$$

The claim follows.

Now, we prove the step using Chebyshev and optimizing λ. We have

$$P[(X - E[X]) > t] \le \exp(-\lambda t)E[\exp(\lambda(X - E[X]))] \le \exp\left(-\lambda t + \frac{\lambda^2}{2}\sum_{k=1}^{n} c_k^2\right)$$

where the first inequality is by Chebyshev with $\varphi(t) = \exp(\lambda T)$ and both inequalities hold for all λ. To find the optimal λ, we set

$$\left(-\lambda t + \frac{\lambda^2}{2}\sum_{k=1}^{n} c_k^2\right)' = -t + \lambda \sum c_k^2 = 0$$

which implies $\lambda = \frac{t}{\sum c_k^2}$. Then

$$P[(X - E[X]) > t] \le \exp\left(-\frac{t^2}{\sum c_k^2} + \frac{t^2}{2(\sum c_k^2)}(\sum c_k^2)\right) = \exp\left(-\frac{t^2}{2} \cdot \frac{1}{\sum c_k^2}\right)$$

□

The inequality in the proof involving cosh follows from the expansion of cosh:

$$\cosh x = \sum_{k\ge 0} \frac{x^{2k}}{(2k)!} \le \sum_{k\ge 0} \frac{1}{k!}\frac{x^{2k}}{2^k} = \exp(x^2/2)$$

6.8.1 Applications

Definition 6.50. *A function $\varphi(x_1, \ldots, x_n)$ is called* discrete-Lipschitz *provided* $\forall k$

$$\sup_{\vec{x}} \sup_{y} |\varphi(x_1, \ldots, x_{k-1}, x_k, \ldots, x_n) - \varphi(x_1, \ldots, x_{k-1}, y, x_{k+1}, \ldots, x_n)| \\ =: c_k < \infty$$

Note: it is "Lipschitz" with respect to discrete metrics.

Theorem 6.51. *Let $\vec{Y} := Y_1, Y_2, \ldots, Y_n$ be* independent *RVs and $\varphi(x_1, \ldots, x_n)$ be discrete-Lipschitz. Then*

$$P\left[\varphi(Y_1, \ldots, Y_n) - E[\varphi(\vec{Y})] \geq t\right] \leq \exp\left(-\frac{1}{2}\frac{t^2}{\sum c_k^2}\right)$$

The inequality also holds for $\leq -t$.

Remark 6.52. The above theorem gives a *concentration inequality* around the mean. ***** something about arbitrary $t, \varphi, \ldots$ ******

Definition 6.53. *Let $Y_1, \ldots, Y_n$ b RVs on $(\Omega, \mathcal{F}, P)$ and $\varphi : \mathbb{R}^n \to \mathbb{R}$ measurable. We say that $\varphi(Y_1, \ldots, Y_n)$ has* bounded variation in every argument a.s. *provided $\exists \bar{\Omega} \in \mathcal{F}$ with $P[\bar{\Omega}] = 1$ such that $\forall \omega, \omega' \in \bar{\Omega}$, $\forall k = 1, \ldots, n$, we have*

$$\sup_{\vec{x} \in \mathbb{R}^n} |\varphi\left(x_1, \ldots, x_{k-1}, Y_k(\omega), x_{k+1}, \ldots, x_n\right) \\ - \varphi\left(x_1, \ldots, x_{k-1}, Y_k(\omega'), x_{k+1}, \ldots, x_n\right)| =: c_k < \infty$$

Example 6.54. For $\vec{Y}$ arbitrary and φ discrete Lipschitz, the condition holds. If Y_k is bounded a.s. for every k and φ is continuous, the condition holds.

Theorem 6.55. *If $\varphi(Y_1, \ldots, Y_n)$ satisfies the condition (i.e. has bounded variation in every argument a.s.) and $Y_1, \ldots, Y_n$ are* independent*, then*

$$P\left[\varphi(\vec{Y}) - E[\varphi(\vec{Y})] \geq t\right] \leq \exp\left(-\frac{t^2}{2\sum_{k=1}^n c_k^2}\right)$$

Also true for $P[\cdot \leq -t]$.

Proof. Set $\mathcal{F}_k = \sigma(Y_1, \ldots, Y_k)$ and $\mathcal{F}_0 = (\emptyset, \Omega)$ and let our martingale be $X_k :=$

$E[\varphi(Y_1,\ldots,Y_n)|\mathcal{F}_k]$. WWTS it has bounded increments. Fix $\omega \in \Omega$. Then

$$\begin{aligned}
&|X_k(\omega) - X_{k-1}(\omega)| \\
&\qquad = |E[\varphi(Y_1,\ldots,Y_n)|\mathcal{F}_k](\omega) - E[\varphi(\vec{Y})|\mathcal{F}_{k-1}](\omega)| \\
&\qquad \text{(for a.e. } \omega) \\
&\qquad = |E[\varphi(Y_1(\omega),\ldots,Y_k(\omega),Y_{k+1}(\cdot),\ldots,Y_n(\cdot))] \\
&\qquad\qquad - E[\varphi(Y_1(\omega),\ldots,Y_{k-1}(\omega),Y_k(\cdot),\ldots,Y_n(\cdot))]| \\
&\qquad \text{(cond. on } F_{k+1}) \\
&\qquad = |E_{\omega'}[E[\varphi(Y_1(\omega),\ldots,Y_k(\omega),Y_{k+1},\ldots,Y_n)|\mathcal{F}_{k+1}](\omega')] \\
&\qquad\qquad - E_{\omega'}[E[\varphi(Y_1(\omega),\ldots,Y_{k-1}(\omega),Y_k,\ldots,Y_n)|\mathcal{F}_{k+1}](\omega')]| \\
&\qquad = |E_{\omega'}[E_{\omega''}[\varphi(Y_1(\omega),\ldots,Y_k(\omega),Y_{k+1}(\omega'),\ldots,Y_n(\omega'))]] \\
&\qquad\qquad - E_{\omega'}[E_{\omega''}[\varphi(Y_1(\omega),\ldots,Y_{k-1}(\omega),Y_k(\omega''),Y_{k+1}(\omega'),\ldots,Y_n(\omega'))]]| \\
&\qquad \leq E_{\omega'}[E_{\omega''}[|\varphi(Y_1(\omega),\ldots,Y_k(\omega),Y_{k+1}(\omega'),\ldots,Y_n(\omega')) \\
&\qquad\qquad - \varphi(Y_1(\omega),\ldots,Y_{k-1}(\omega),Y_k(\omega''),Y_{k+1}(\omega'),\ldots,Y_n(\omega'))|]] \\
&\qquad \leq c_k \text{ for a.e. } \omega
\end{aligned}$$

The theorem follows by Azuma (7). □

Example 6.56 (Directed first passage percolation). Take Γ a directed path from A to B. For every edge e, Y_e is a $U[0,1]$ distributed RV representing the time we need to pass through edge e. Let

$$\varphi(\vec{Y}(\omega)) := \min_{\Gamma:\text{ path } A\to B} \left\{ \sum_{e\in\Gamma} Y_e \right\}$$

be the passage time from $A \to B$. **Questions**: Expected value $E[\varphi(\vec{Y})]$? Concentration? Variance? Large dev.? CLT?

6.9 Large Deviations: Cramer's Theorem

Let $(X_k)_{k\geq 1}$ be an i.i.d. sequence of RVs with $X_k \sim \mu$. The *logarithmic moment generating function* is defined to be

$$\Lambda(\lambda) := \log E[\exp(\lambda X)] \in (-\infty, \infty]$$

and its *Legendre transform* is

$$\Lambda^\star(x) := \sup_\lambda \{\lambda x - \Lambda(\lambda)\}$$

Let

$$\mu_n = P \circ \bar{S}_n^{-1} = P \circ \left(\frac{1}{n} \sum_{k=1}^n X_k \right)^{-1}$$

be the distribution of the average ($\leq n$) of the X_is.

Theorem 6.57 (Cramer's Thm)**.** *Let $A \subseteq \mathbb{R}$ be Borel. Then*

$$-\inf_{A^\circ} \Lambda^\star \le \liminf_{n\to\infty} \frac{1}{n} \log \mu_n(A^\circ) \le \limsup_{n\to\infty} \frac{1}{n} \log \mu_n(\bar{A}) \le -\inf_{\bar{A}} \Lambda^\star$$

Note: $\mu_n(A) = P[\bar{S}_n \in A]$, A° is interior of A, $\bar{A}$ is closure of A. Also,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(\bar{A}) \le -\inf_{\bar{A}} \Lambda^\star \iff$$
$$\forall \varepsilon > 0 \quad \mu_n(\bar{A}) \le \exp\left(-n \inf_{\bar{A}} \Lambda^\star - \varepsilon\right) \ \forall n \ge n_0(\varepsilon)$$

Note: if $m = \exp(-n\Lambda)$ for some Λ, then $\frac{1}{n}\log m = -\Lambda$ (rate of exponential decay).

Properties of $\Lambda, \Lambda^\star$:

1. $\Lambda, \Lambda^\star$ are both convex and lower semi-continuous
2. $\Lambda_X(-\lambda) = \Lambda_{-X}(\lambda)$ for $\lambda \ge 0$
3. $\Lambda(0) = 0$, $\Lambda^\star(x) \ge 0$

To prove these, notice

$$\Lambda = \lim_{c\to\infty} \nearrow \log \underbrace{E[\exp(\lambda(X \wedge c))]}_{\text{const.}} \Rightarrow \text{ l.s.c.}$$

and

$$\begin{aligned}
\Lambda(p\lambda_1 + (1-p)\lambda_2) &= \log E\left[\exp(\lambda_1 X)^p \exp(\lambda_2 X)^{1-p}\right] \\
&\le \log\left(E[\exp(\lambda_1 X)]^p E[\exp(\lambda_2 X)]^{1-p}\right) \\
&= p\Lambda(\lambda_1) + (1-p)\Lambda(\lambda_2)
\end{aligned}$$

by Hölder, which implies Λ is convex. Then $\Lambda^\star$ is convex and l.s.c. as a pointwise supremum of linear functions.

From now on, we assume "Cramer's condition", i.e. that X has *some exponential moment*:

$$\exists \lambda_0 > 0 \text{ such that } E[\exp(\lambda_0 |X|)] < \infty$$

which implies that $\Lambda(\lambda) < \infty$ for $|\lambda| < \lambda_0$. Set

$$D_\Lambda = \{\lambda \in \mathbb{R} : \Lambda(\lambda) < \infty\}$$

6.9.1 Further properties under Cramer's condition

1. $E[X] =: \bar{x}$ is finite and $\Lambda^\star(\bar{x}) = 0$.

2. $\forall x \geq \bar{x}$,

$$\Lambda^\star(x) = \sup_{\lambda \geq 0} \{\lambda x - \Lambda(\lambda)\}$$

and for $x \leq \bar{x}$, the sup is over $\lambda \leq 0$.

3. $\Lambda^\star$ is $\nearrow$ on $[\bar{x}, \infty)$ and $\searrow$ on $(-\infty, \bar{x}]$.

Proof of (1): $\forall \lambda$,

$$\Lambda(\lambda) = \log E[\exp(\lambda X)] \geq E[\log \exp(\lambda X)] = \lambda \bar{x}$$

by Jensen, so

$$\sup_\lambda \underbrace{\lambda \bar{x} - \Lambda(\lambda)}_{\leq 0} = \Lambda^\star(\bar{x}) \leq 0 \ \Rightarrow \ \Lambda^\star(\bar{x}) = 0$$

Proof of (2): for $x \geq \bar{x}$, $\forall \lambda > 0$,

$$\lambda x - \Lambda(\lambda) \leq \lambda \bar{x} - \Lambda(\lambda) \leq \Lambda^\star(\bar{x}) = 0$$

Proof of (3): let $\bar{x} \leq x \leq y$, so then

$$\Lambda^\star(x) = \sup_{\lambda \geq 0} \underbrace{\lambda x - \Lambda(\lambda)}_{\leq \lambda y - \Lambda(\lambda)} \leq \sup \lambda y - \Lambda(\lambda) = \Lambda^\star(y)$$

etc.

Lemma 6.58. *Λ is differentiable in D_Λ° with*

$$\Lambda'(\lambda) = \frac{1}{E[\exp(\lambda X)]} E[X \exp(\lambda X)]$$

(finite) and

$$\Lambda'(\lambda_0) = q \ \Rightarrow \ \Lambda^\star(q) = \lambda_0 q - \Lambda(\lambda_0)$$

(i.e. λ_0 is the optimizer for $\Lambda^\star(q)$).

Proof. First statement is straightforward application of dominated convergence. To prove the second statement, let $g(y) := \lambda y - \Lambda(\lambda)$. Since $g'(\lambda_0) = y - \Lambda'(\lambda_0) = 0$ and $g(\cdot)$ is concave, we have that λ_0 is a global max; that is,

$$\lambda_0 y - \Lambda(\lambda_0) = g(\lambda_0) = \sup_\lambda g(\lambda) = \sup_\lambda \lambda y - \Lambda(\lambda) = \Lambda^\star(y)$$

□

Now we're ready to prove Cramer's Theorem 6.57.

Proof. **Upper bound**: Let $x > \bar{x}$ (proof for $x < \bar{x}$ is analogous). Then

$$P[\bar{S}_n \in [x, \infty)] = E[\mathbf{1}_{[x,\infty)}(\bar{S}_n)] \leq E\left[\exp(-n\lambda x) \exp(n\lambda \bar{S}_n)\right] = (\star)$$

by Chebyshev and the fact that $\lambda \geq 0 \Rightarrow \mathbf{1}_{[x,\infty)}(\cdot) \leq \exp(-n\lambda x)\exp(n\lambda(\cdot))$. Continuing, we have

$$\begin{aligned}(\star) &= \exp(-n\lambda x)\prod_{k=1}^{n} E[\exp(\lambda X)] = \exp(-n\lambda x)E[\exp(\lambda X)]^n \\ &= \exp\left(-n(\lambda x - \Lambda(\lambda))\right) \\ &\leq \inf_{\lambda\geq 0}\exp(-n(\cdots)) = \exp\left(-n\sup_{\lambda\geq 0}(\lambda x - \Lambda(\lambda))\right) \\ &= \exp(-n\Lambda^\star(x))\end{aligned}$$

Then, to get

$$\limsup_{n\to\infty}\frac{1}{n}\log P[\bar{S}_n \in F] \leq -\inf_F \Lambda^\star$$

we notice that if $\bar{x} \in F$ then $\inf_F \Lambda^\star = 0$ (since $\Lambda^\star(\bar{x}) = 0$) so there's nothing to show. Let $\bar{x} \notin F$ with F closed. Then

$$P[\bar{S}_n \in F] \leq \underbrace{P[S_n \geq x^+]}_{\leq \exp(-n\Lambda^\star(x^+))} + \underbrace{P[S_n \leq x^-]}_{\leq \exp(-n\Lambda^\star(x^-))}$$

Note

$$\inf_F \Lambda^\star = \min\left\{\Lambda^\star(x^+), \Lambda^\star(x^-)\right\}$$

since $\Lambda^\star$ is $\nearrow$ on $[\bar{x},\infty)$ and $\searrow$ on $(-\infty,\bar{x}]$. WOLOG $\Lambda^\star(x^+)$ is the minimum, so

$$\exp(-na) = \exp(-n\Lambda^\star(x^+)) \geq \exp(-n\Lambda^\star(x^-)) = \exp(-nb)$$

Then,

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\log P[\bar{S}_n \in F] &\leq \limsup_{n\to\infty}\frac{1}{n}\log\left(\exp(-na)\left(1+\frac{\exp(-nb)}{\exp(-na)}\right)\right) \\ &\leq \limsup_{n\to\infty}\frac{1}{n}(-na) + 0 = -\Lambda^\star(x^+) = -\inf_F \Lambda^\star\end{aligned}$$

since $\log\left(\exp(-na)\left(1+\frac{\exp(-nb)}{\exp(-na)}\right)\right) \leq -na + \log 2$.

Lower bound: we will show that $\forall\delta > 0, \mu(\sim X)$ (with Cramer's condition), the following $(\star)$ holds:

$$\liminf_{n\to\infty}\frac{1}{n}\log\underbrace{\mu_n((-\delta,\delta))}_{=P[|S_n|<\delta]} \geq -\Lambda^\star(0)$$

This will, in turn, imply that

$$\liminf_{n\to\infty}\frac{1}{n}\mu_n((q-\delta,q+\delta)) \geq -\Lambda^\star(q); \forall q$$

after a shift $Y := X - q$. Morevoer, if $G \subseteq \mathbb{R}$ is open and $q \in G$ then $\exists \delta > 0$ such that $(q - \delta, q + \delta) \subseteq G$, so

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n((q-\delta, q+\delta)) \geq \sup_{q\in G}(-\Lambda^\star(q)) = -\inf_{q\in G} \Lambda^\star(q)$$

Now, WWTS $(\star)$. Assume, first, that $\mu(-\infty, 0) > 0$ and $\mu(0, \infty) > 0$ and X is bounded (μ has compact support). From these assumptions, we have

$$\lim_{|\lambda|\to\infty} \Lambda(\lambda) = +\infty \text{ and } \Lambda(\lambda) < \infty \; \forall \lambda \in \mathbb{R}$$

Then $D_\Lambda^\circ = \mathbb{R}$ and Λ is differentiable, which implies $\exists$ a global min λ_0 where $\Lambda'(\lambda_0) = 0$ (which implies $\Lambda^\star(0) = \lambda_0 \cdot 0 - \Lambda(\lambda_0) = -\Lambda(\lambda_0)$). We use λ_0 to define a new probability measure $\tilde{\mu}$ on $\mathbb{R}$ by

$$\tilde{\mu}(dx) := \exp(\lambda_0 x - \Lambda(\lambda_0))\mu(dx)$$

Note that

$$\int_{\mathbb{R}} \tilde{\mu}(dx) = \underbrace{\exp(-\Lambda(\lambda_0))}_{=1/E[\exp(\lambda_0 X)]} \underbrace{\int \exp(\lambda_0 x)\mu(dx)}_{=E[\exp(\lambda_0 X)]} = 1$$

Moreover, $\int_{\mathbb{R}} x\tilde{\mu}(dx) = 0$ since

$$\begin{aligned}\int_{\mathbb{R}} x\tilde{\mu}(dx) &= \int x \exp(\lambda_0 x)\mu(dx) \cdot \frac{1}{E[\exp(\lambda_0 x)]} \\ &= \frac{E[X\exp(\lambda_0 X)]}{E[\exp(\lambda_0 X)]} = \Lambda'(\lambda_0) = 0\end{aligned}$$

by the previous lemma. Let $\tilde{\mu}_n$ be the joint distribution of $\bar{S}_n$ where $X_i \sim \tilde{\mu}$ are i.i.d. Then

$$\begin{aligned}\mu_n(-\delta, \delta) &= \int_{\mathbb{R}^n} \mathbf{1}_{\{\frac{1}{n}|\frac{1}{n}\sum x_i|<\delta\}}(x)\, \underbrace{\mu(dx_1)} \ldots \mu(dx_n) \\ &= \int_{\mathbb{R}^n} \mathbf{1}_{\{\cdots\}} \exp(-\lambda_0 \sum x_i)) \exp(n\Lambda(\lambda_0))\tilde{\mu}(dx_1)\ldots\tilde{\mu}(dx_n) \\ &\geq \exp(-n\delta)\exp(n\Lambda(\lambda_0))\, \underbrace{\tilde{\mu}_n(-\delta, \delta)}_{\to 1}\end{aligned}$$

since under $\tilde{\mu}_n$, $\bar{S}_n \to 0$ weakly. Thus,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(-\delta, \delta) \geq -\delta + \Lambda(\lambda_0) = -\delta - \Lambda^\star(0)$$

and let $\delta \to 0$. $\square$