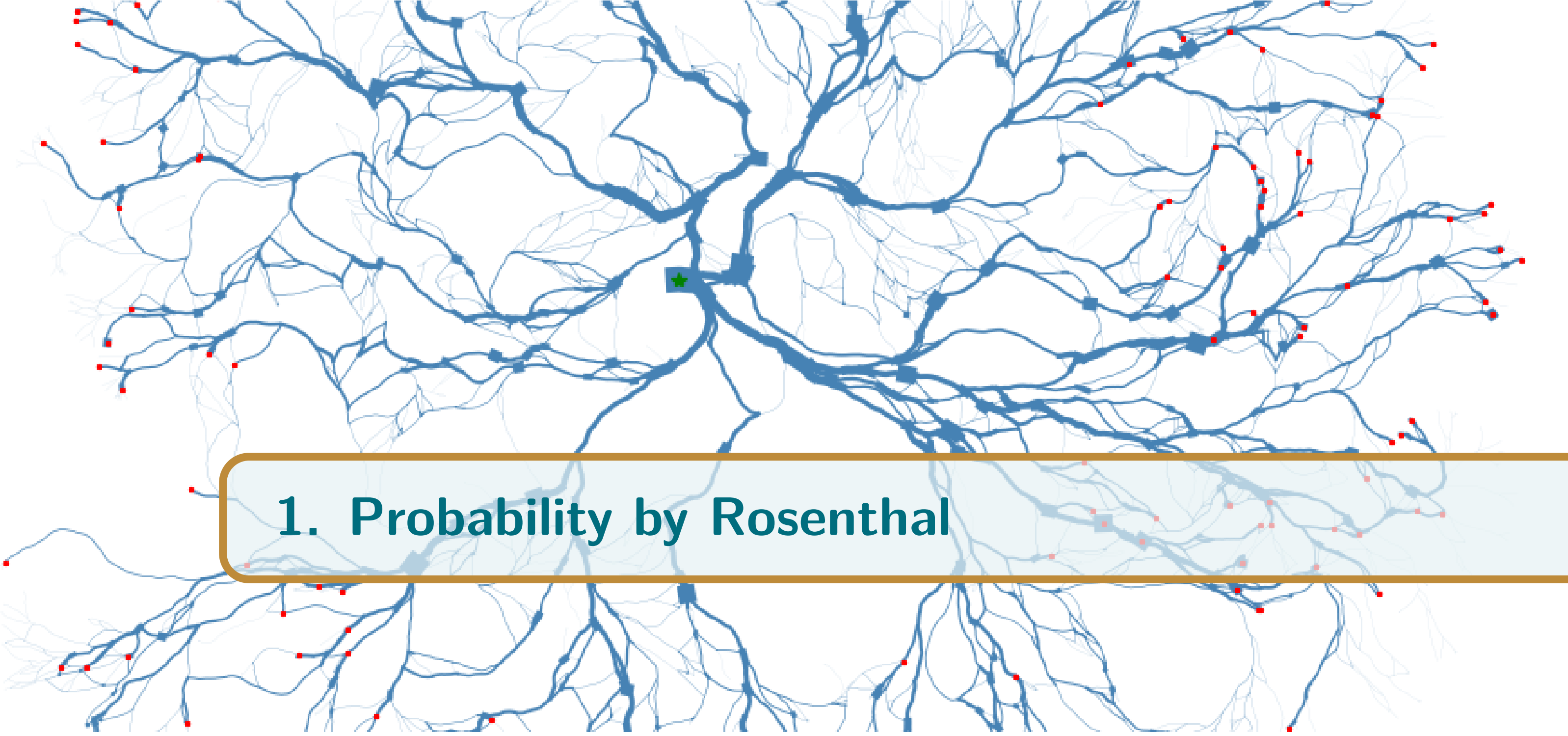
Lecture Notes For: Probability and Stochastic Process

Ali Fele Paranj
alifele@student.ubc.ca

December 16, 2024

0. Contents



1. Probability by Rosenthal

In this chapter, I will include sporadic notes during my study of probability from the Rosenthal book. Also, I will try to compile a set of solutions for the problems in this book.

1.1 Probability Triples

Definition 1.1 — Semialgebra. Let X be a set. A Semialgebra $\mathcal{I}$ of the subsets of X is a collection of the subsets of such that

(a) $\varnothing, X \in \mathcal{I}$.

(b) $\mathcal{I}$ is closed *finite* **intersection**.

(c) For $E \in \mathcal{I}$ it complement E^c can be written as a *finite disjoint* **union** of sets in $\mathcal{I}$.

■ **Remark** One canonical example for a semialgebra is the set of all intervals in $\mathbb{R}$, where the term interval contains all open, closed and half open intervals, as well as the empty set, singletons, and the whole space.

Definition 1.2 — Algebra. Let $\mathcal{M}$ be a collection of sets. Then $\mathcal{M}$ is an algebra if

(a) $\Omega, \varnothing \in \mathcal{M}$

(b) $\mathcal{M}$ is closed under complements.

(c) $\mathcal{M}$ is closed under finite intersection.

(d) $\mathcal{M}$ is closed under finite union.

Proposition 1.1 Let $\mathcal{I}$ be a semialgebra, and $\mathcal{F} = \sigma(\mathcal{I})$. Let $\mathbb{P}, \mathbb{Q}$ be two probability measures defined on $\mathcal{F}$. Then if $\mathbb{P}$ agrees with $\mathbb{Q}$ on $\mathcal{I}$, then they agree on $\mathcal{F}$.

■ **Remark** The condition that $\mathcal{I}$ is a semialgebra is crucial. See 1.10 for an example.

1.1.1 Solved Problems

■ **Problem 1.1** Let $\Omega = \{1, .2, 3, 4\}$. Determine whether or not each of the following is a σ-algebra.

(a) $\mathcal{F}_1 = \{\varnothing, \{1, 2\}, \{3, 4\}, \{1, 2, 3, 4\}\}$.

(b) $\mathcal{F}_2 = \{\varnothing, \{3\}, \{4\}, \{1, 2\}, \{3, 4\}, \{1, 2, 3\}, \{1, 2, 4\}, \{1, 2, 3, 4\}\}$.

(c) $\mathcal{F}_3 = \{\varnothing, \{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}, \{3, 4\}, \{1, 2, 3, 4\}\}$.

Solution

(a) $\mathcal{F}_1$ is a σ-algebra and the set of its atoms are $\{\{1, 2\}, \{3, 4\}\}$.

(b) $\mathcal{F}_2$ is a σ-algebra and the set of its atoms are $\{\{3\}, \{4\}, \{1, 2\}\}$.

(c) $\mathcal{F}_3$ is **not** a σ-algebra because $\{1, 2\}, \{2, 3\} \in \mathcal{F}_3$ but $\{1, 2\} \cap \{2, 3\} = \{2\} \notin \mathcal{F}_3$.

■ **Problem 1.2** Let $\Omega = \{1, 2, 3, 4\}$, and let $\mathcal{I} = \{\{1\}, \{2\}\}$. Describe explicitly the σ-algebra $\sigma(\mathcal{I})$ (i.e. the smallest σ-algebra containing the collection $\mathcal{I}$).

Solution The smallest σ-algebra containing the collection $\mathcal{I}$ is

$$\sigma(\mathcal{I}) = \{\{1\}, \{2\}, \{3, 4\}, \{1, 2\}, \{2, 3, 4\}, \{1, 3, 4\}, \{1, 2, 3, 4\}, \varnothing\}.$$

One way to check to see if this is really the smallest σ-algebra is to first observe that the cardinality of σ-algebra of a finite set should always be of the form 2^n for some $n \in \mathbb{N}$, where n is the number of atoms (or the number of the non-empty sets the the σ-algebra does not contain any of its subsets). Observe that $\{1\}$ and $\{2\}$ are already the atoms of the σ-algebra. Thus the size of $\sigma(\mathcal{I})$ must be at least four. However, we know that $\sigma(\mathcal{I})$ contains at least 5 elements (i.e. $\{1\}, \{2\}, \{1, 2\}, \{1, 2, 3, 4\}, \varnothing$). This suggests that there should be at least one other atom in the set. Choosing that atom to be $\{3, 4\}$ will yield that $\sigma(\mathcal{I})$ that contains 8 elements. Since this already includes that collection $\mathcal{I}$, and we can not have any smaller σ-algebra then we are sure that this is the smallest σ-algebra.

■ **Problem 1.3** Suppose $\mathcal{F}$ is a collection of subsets of Ω, such that $\Omega \in \mathcal{F}$.

(a) Suppose $\mathcal{F}$ is an algebra. Prove that $\mathcal{F}$ is a semialgebra.

(b) Suppose that whenever $A, B \in \mathcal{F}$, then also $A \backslash B \equiv A \cap B^c \in \mathcal{F}$. Prove that $\mathcal{F}$ is an algebra.

(c) Suppose that $\mathcal{F}$ is closed under complement, and also closed under finite *disjoint* unions. Give a counter example to show that $\mathcal{F}$ might not be an algebra.

Solution (a) Firstly, Since $\mathcal{F}$ is an algebra, then it is closed under complement, hence $\varnothing \in \mathcal{F}$. Secondly, Since it is closed under finite intersection, then it meets the closedness under finite intersection property of a semialgebra. Lastly, let $E \in \mathcal{F}$. Since $\mathcal{F}$ is an algebra then $E^c \in \mathcal{F}$. So we can trivially write $E^c = E^c$ as a finite disjoint union of sets in $\mathcal{F}$. Thus $\mathcal{F}$ is a semialgebra.

(b) Firstly, since $\Omega \in \mathcal{F}$, then by hypothesis $\Omega \backslash \Omega = \varnothing \in \mathcal{F}$. Secondly, let $A \in \mathcal{F}$. Then by hypothesis $\Omega\ A = A^c \in \mathcal{F}$, thus $\mathcal{F}$ is closed under complement. Lastly, Let $A, B \in \mathcal{F}$. By the reasoning above $B^c \in \mathcal{F}$. And by hypothesis $A \backslash B^c \in \mathcal{F}$. This implies that $A \cap B \in \mathcal{F}$

(c) One simple counter example can be constructed when we let $\Omega = \{1,2,3,4\}$ and then let

$$\mathcal{F} = \{\Omega, \varnothing, \{1,2\}, \{1,3\}, \{1,4\}, \{2,3\}, \{2,4\}, \{3,4\}\}.$$

This collection is closed under finite disjoint union as well as complement. But it fails to be an algebra. For instance $\{1,2\}, \{2,3\} \in \mathcal{F}$, but their intersection is not in the collection.

■ **Problem 1.4** Let $\mathcal{F}_1, \mathcal{F}_2, \cdots$ be a sequence of collections of subsets of Ω, such that $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$ for each n.

(a) Suppose that each $\mathcal{F}_i$ is an algebra. Prove that $\bigcup_{i=1}^{\infty} \mathcal{F}_i$ is also an algebra.

(b) Suppose that each $\mathcal{F}_i$ is a σ-algebra. Show (by counterexample) that $\bigcup_{i=1}^{\infty} \mathcal{F}_i$ need not be a σ-algebra.

Solution (a) Let $\mathcal{G} = \bigcup_{i=1}^{\infty} \mathcal{F}_i$. First, observe that since $\Omega, \varnothing \in \mathcal{F}_i$ for all $i \in \mathbb{N}$ (since all of them are algebra), then it follows that $\Omega, \varnothing \in \mathcal{G}$. Furthermore, let $A \in \mathcal{G}$. Then $A \in \mathcal{F}_i$ for some $i \in \mathbb{N}$. Since $\mathcal{F}_i$ is an algebra, then $A^c \in \mathcal{F}_i$, hence $A^c \in \mathcal{G}$. Lastly, let $A, B \in \mathcal{G}$. Then $A \in \mathcal{F}_i$ and $B \in \mathcal{F}_j$ for some $i, j \in \mathbb{N}$. WLOG we can assume $i \leq j$. Then $\mathcal{F}_i \subset \mathcal{F}_j$, hence $A, B \in \mathcal{F}_j$. Since $\mathcal{F}_j$ is an algebra, then $A \cap B \in \mathcal{F}_j$. Thus $A \cap B \in \mathcal{G}$. This proves that $\mathcal{G}$ is an algebra.

(b) Let $\Omega = \mathbb{N}$. Let $\mathcal{F}_n$ be the smallest σ-algebra that contains the collection $\{\{1\}, \cdots, \{n\}\}$. On other way to think about $\mathcal{F}_n$ is the σ-algebra that contains the power set of $\{1, \cdots, n\}$ as well as all of their complements (with respect to Ω). For instance, we have

$$\mathcal{F}_1 = \{\varnothing, \{1\}, \mathbb{N}, \{1\}^c\}.$$

Similarly

$$\mathcal{F}_2 = \{\varnothing, \{1\}, \{2\}, \{1,2\}, \mathbb{N}, \{1\}^c, \{2\}^c, \{1,2\}^c\},$$

and etc. Let $A_i = \{2i\}$. Clearly $A_i \in \bigcup_i \mathcal{F}_i$. However, $\bigcup_i A_i \notin \bigcup_i \mathcal{F}_i$ as it does not belong to any $\mathcal{F}_k$. Thus $\bigcup_i \mathcal{F}_i$ is not a σ-algebra.

■ **Problem 1.5** Suppose that $\Omega = \mathbb{N}$ is the set of positive integers, and $\mathcal{F}$ is the set of all subsets A such that either A or A^c is finite, and $\mathbb{P}$ is defined by $\mathbb{P}(A) = 0$ if A is finite, and $\mathbb{P}(A) = 1$ if A^c is finite.

(a) Is $\mathcal{F}$ an algebra?

(b) Is $\mathcal{F}$ a σ-algebra?

(c) Is $\mathbb{P}$ finitely additive?

(d) Is $\mathbb{P}$ countably additive on $\mathcal{F}$, meaning that if $A_1, A_2, \cdots \in \mathcal{F}$ are disjoint, and if it happens that $\bigcup_n A_n \in \mathcal{F}$, then $\mathbb{P}(\bigcup_n A_n) = \sum_n \mathbb{P}(A_n)$?

Solution (a) Yes. First, observe that $\varnothing, \Omega \in \mathcal{F}$ as $\varnothing$ is finite, and Ω has a finite complement. Further, let $A \in \mathcal{F}$. Then either it is finite or it has a finite complement, where for both cases we have $A^c \in \mathcal{F}$. Let $A_1, \cdots, A_n$ be a finite collection of sets in $\mathcal{F}$. If all A_i for $i = 1, \cdots, n$ are finite, then since the finite intersection and complement of any finite collection of finite sets is finite, $\bigcap_{i=1}^{n} A_i$ as well as $\bigcup_{i=1}^{n} A_i$ are finite as well, thus belongs to $\mathcal{F}$. If A_i are all infinite, then since they all belong to $\mathcal{F}$ then they have finite complement, hence $\bigcup_{i=1}^{n} A_i^c$ and $\bigcap_{i=1}^{n} A_i^c$ are finite as well, thus belongs to $\mathcal{F}$. If the collection is not in any of the case above, then there is $1 \leq j \leq n$ such that A_j is finite. Thus $\bigcup_i A_i$ is finite, thus belongs to $\mathcal{F}$. Being closed under finite intersection and complements implies being closed under finite union.

(b) No. Let $A_n = \{2n\}$. Then $A_n \in \mathcal{F}$ for all $n \in \mathbb{N}$. However, $\bigcup_n A_n \notin \mathcal{F}$ as it is neither finite nor has a finite complement.

(c) Yes. First observe that if A, B are both infinite with A^c, B^c finite (this $A, B \in \mathcal{F}$), then $A \cap B \neq \varnothing$. Otherwise, $A^c \cup B^c = \mathbb{N}$ which implies that either of them is infinite which is a contradiction. Thus given A, B, if both are finite then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) = 0$ as $A \cup B$ is also finite. If both are infinite, then based on our argument above then they are not disjoint, so the argument of additivity does not apply to them. However if WLOG A is finite and B is infinite and $A \cap B = 0$, then $A \cup B$ is also infinite thus $1 = \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) = 0 + 1$.

(d) No. Let $A_n = \{n\}$. Then

$$\mathbb{P}(\bigcup_n A_n) = \mathbb{P}(\mathbb{N}) = 1 \neq \sum_n \mathbb{P}(A_n) = 0.$$

Proposition 1.2 Let $\Omega = \mathbb{N}$ and let $\mathcal{F}$ be the collection of all subsets of Ω that is *countable* or has countable *complement*. Then $\mathcal{F} = \sigma(\mathcal{A})$ where $\mathcal{A} = \{\{1\}, \{2\}, \cdots\}$, i.e. the set of all singletons.

Proof. Let $E \in \mathcal{F}$. First, note that the collection $\mathcal{F}$ is a σ-algebra. Then, notice that $\mathcal{F}$ contain $\mathcal{A}$ as singletons are finite, hence countable. Since $\sigma(\mathcal{A})$ is the smallest σ-algebra that contain $\mathcal{A}$ then $\sigma(\mathcal{A}) \subset \mathcal{F}$. Let $E \in \mathcal{F}$. Then E is either countable or has a countable complement. If E is countable then it can be written as a countable union of singletons in which each singleton contains one element of E. Thus $E \in \sigma(\mathcal{A})$. If E^c is countable, then E^c can be written as a countable union of singleton of its element. By applying De Morgan's law E can be written as a countable union of the complements of singletons (which belong to $\sigma(\mathcal{A})$). Thus case also implies $E \in \sigma(\mathcal{A})$. Thus $\mathcal{F} = \sigma(\mathcal{A})$. $\square$

■ **Problem 1.6** Suppose that $\Omega = [0, 1]$ is the unit interval, and $\mathcal{F}$ is the set of all subsets A such that either A or A^c is finite, and $\mathbb{P}$ is defined by $\mathbb{P}(A) = 0$ of A is finite and $\mathbb{P}(A) = 1$ if A^c is finite.

(a) Is $\mathcal{F}$ an algebra?

(b) Is $\mathcal{F}$ a σ-algebra?

(c) Is $\mathbb{P}$ finitely additive?

(d) Is $\mathbb{P}$ countably additive on $\mathcal{F}$?

Solution (a) Yes. Being closed under complement is immediate from the definition. Thus $\varnothing, \Omega \in \mathcal{F}$. Let $A, B \in \mathcal{F}$. Then if A, B are both finite, then $A \cap B$ is also finite thus $A \cap B \in \mathcal{F}$. If A, B are both infinite, then A^c, B^c are both finite, so it is $A^c \cup B^c$. Being closed under complement it implies that $A \cup B \in \mathcal{F}$. If one of A, B is infinite and the other one is finite, then $A \cap B$ is finite, hence $A \cap B \in \mathcal{F}$. Thus $\mathcal{F}$ is a σ-algebra.

(b) No. Consider the collection $\{A_q\}_{q \in \mathbb{Q}}$ where $q \in A$. Each of these sets are finite, hence $A_q \in \mathcal{F}$ for all $q \in \mathbb{Q}$. However $\bigcup_q A_q$ is not finite and its complement is also not finite. Thus $\mathcal{F}$ is not closed under countable union.

(c) Yes. First observe that if $A, B \in \mathcal{F}$ both infinite, then their intersection can not be empty, otherwise $A^c \cup B^c = [0, 1]$ which means that at least one of them is infinite which is a contradiction. With this in mind let $A, B \in \mathcal{F}$. If A, B both finite with empty intersection then $0 = \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) = 0 + 0$ as $A \cup B$ is also finite. If WLOG A is infinite and B is finite and $A \cup B = \varnothing$, then $A \cup B$ is also infinite and we have $1 = \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) = 0 + 1$. Thus $\mathbb{P}$ is finitely additive.

(d) Yes. First observe that we can not find any two $A, B \in \mathcal{F}$ disjoint and infinite with empty intersection since then $A^c \cup B^c = [0, 1]$ that implies at least one of them is infinite. So let $A_1, A_2, \cdots$ be a sequence of *finite* sets in $\mathcal{F}$. Then

$$0 = \mathbb{P}(\bigcup_i A_i) = \sum_i \mathbb{P}(A_i) = 0.$$

In case if just one of A_i is infinite (no more than two can be infinite at the same time) then $\bigcup_i A_i$ is infinite and

$$1 = \mathbb{P}(\bigcup_i A_i) = \sum_i \mathbb{P}(A_i) = 1.$$

■ **Problem 1.7** Suppose that $\Omega = [0, 1]$ is the unit interval, and $\mathcal{F}$ is the set of all subsets A such that either A or A^c is countable (i.e. finite or countable), and $\mathbb{P}$ is defined by $\mathbb{P}(A) = 0$ if A is countable, and $\mathbb{P}(A) = 1$ if A^c is countable.

(a) Is $\mathcal{F}$ an algebra?

(b) Is $\mathcal{F}$ a σ-algebra?

(c) Is $\mathbb{P}$ finitely additive?

(d) Is $\mathbb{P}$ countably additive?

Solution (a) Yes. Being closed under complement follows immediately from the definition. On the other hand, since $\varnothing$ is finite, then Ω (complement of the empty set) also belongs to Ω. Let $A, B \in \mathcal{F}$. If both are countable then $A \cap B$ is also countable, thus $A \cap B \in \mathcal{F}$. If for both their complement is countable, then $A^c \cup B^c = (A \cap B)^c$ is countable. Thus $A \cap B \in \mathcal{F}$. If one of them is countable, WLOG A, then $A \cap B$ is also countable, thus $A \cap B \in \mathcal{F}$. Thus it is an algebra.

(b) Yes. Being closed under complement follows from the definition and from this it follows that $\Omega \in \mathcal{F}$. Let $E_1, E_2, \cdots$ be a sequence of sets in $\mathcal{F}$. If at least one of them is countable, then $\bigcap_i E_i$ is also countable hence belonging to $\mathcal{F}$. If all is uncountable, then $\bigcup_i E_i^c$ is countable. We can write $\bigcup_i E_i^c = (\bigcap_i E_i)^c$ that is finite. Thus $\bigcap +iE_i \in \mathcal{F}$. This shows that $\mathcal{F}$ is closed under countable union. Being closed under union follows from being closed under intersection and complement. Thus $\mathcal{F}$ is a σ-algebra.

(c) Yes. We will show additivity for two sets and finite additivity will follow by induction. Let $A, B \in \mathcal{F}$. If A, B both are disjoint countable then $A \cup B$ is also countable. Thus $0 = \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) = 0 + 0 = 0$. If A, B are both uncountable (i.e. A^c, B^c are countable) then $A \cap B$ can not be non-empty, otherwise $A^c \cup B^c = [0, 1]$ which then implies that at least one of A^c or B^c be uncountable, which is a contradiction. When one of these sets is countable, WLOG A, then $A \cup B$ is also uncountable and we have $1 = \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) = 1 + 0$. Thus $\mathbb{P}$ is finitely countable.

(d) Yes. Let $A_1, A_2, \cdots$ be a sequence of disjoint sets in $\mathcal{F}$. Then at most one set can be uncountable, otherwise they will fail to be disjoint (see the reasoning in part (c)). If non of them are uncountable, then their union is also countable (countable union of countable sets is countable). Thus $0 = \mathbb{P}(\bigcup_i A_i) = \sum_i \mathbb{P}(A_i) = 0$. If one of them is uncountable, then the union is also uncountable and we will have $1 = \mathbb{P}(\bigcup_i A_i) = \sum_i \mathbb{P}(A_i) = 0 + \cdots + 0 + 1 + 0 + \cdots = 1$. Thus $\mathbb{P}$ is additive on $\mathcal{F}$.

■ **Problem 1.8** For the example of 1.7, is $\mathbb{P}$ uncountably additive?

Solution No. Otherwise we can write $\Omega = \bigcup_{x\in\Omega}\{x\}$. But we have

$$1 = \mathbb{P}(\Omega) = \sum_{x\in\Omega}\{x\} = 0.$$

■ **Problem 1.9** Let $\mathcal{F}$ be a σ-algebra, and write $|\mathcal{F}|$ for the total number of subsets in $\mathcal{F}$. Prove that if $|\mathcal{F}| < \infty$, i.e. $\mathcal{F}$ consists of just a finite number of subsets, then $|\mathcal{F}| = 2^m$ for some $m \in \mathbb{N}$. (*Hint: Consider those non-empty subsets in $\mathcal{F}$ which do not contain any other non-empty subset in $\mathcal{F}$. How can all subsets in $\mathcal{F}$ be build up from these particular subsets?*).

Solution Let $\mathcal{A}$ be the collection of all non-empty sets in $\mathcal{F}$ whose non of its subsets do not belong to $\mathcal{F}$. Then for any $E \in \mathcal{F}$ can be build up from these "atom" by union. For each atom there are two possibilities to be present in the union or not. Thus there are in total 2^m elements in $\mathcal{F}$.

■ **Problem 1.10** Let $\Omega = \{1, 2, 3, 4\}$, with $\mathcal{F}$ the collection of all subsets of Ω. Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on $\mathcal{F}$ such that $\mathbb{P}\{1\} = \mathbb{P}\{2\} = \mathbb{P}\{3\} = \mathbb{P}\{4\} = 1/4$, and $\mathbb{Q}\{2\} = \mathbb{Q}\{4\} = 1/2$, extended to $\mathcal{F}$ by linearity. Finally, let $\mathcal{I} = \{\varnothing, \Omega, \{1, 2\}, \{2, 3\}, \{3, 4\}, \{1, 4\}\}$.

(a) Prove that $\mathbb{P}(A) = \mathbb{Q}(A)$ for all $A \in \mathcal{I}$.

(b) Prove that there is $A \in \sigma(\mathcal{I})$ with $\mathbb{P}(A) \neq \mathbb{Q}(A)$.

(c) Why does this not contradict Proposition 2.5.8 (in Rosenthal)?

Solution (a) By a simple calculation we can show the identity above. For instance

$$\mathbb{P}\{1, 2\} = \mathbb{P}\{1\} + \mathbb{P}\{2\} = 1/4 + 1/4 = 1/2,$$

where as

$$\mathbb{Q}\{1, 2\} = \mathbb{Q}\{1\} + \mathbb{Q}\{2\} = 0 + 1/2 = 1/2.$$

By a similar computation we can show

$$\mathbb{P}\{2, 3\} = \mathbb{Q}\{2, 3\} = 1/2, \quad \mathbb{P}\{3, 4\} = \mathbb{Q}\{3, 4\} = 1/2, \quad \mathbb{P}\{1, 4\} = \mathbb{Q}\{1, 4\} = 1/2,$$

and so on.

(b) First, observe that $\{1, 2, 3\} \in \sigma(\mathcal{I})$. That is because $\{3\} = \{2, 3\} \cap \{3, 4\} \in \sigma(\mathcal{I})$. Thus $\{1, 2\} \cup \{3\} = \{1, 2, 3\} \in \sigma(\mathcal{I})$. But

$$\mathbb{P}\{1, 2, 3\} = 3/4, \qquad \mathbb{Q}\{1, 2, 3\} = 1/2.$$

Thus if we let $A = \{1, 2, 3\} \in \sigma(\mathcal{I})$ then $\mathbb{P}(A) \neq \mathbb{Q}(A)$.

(c) That is because the proposition 2.5.8 requires the collection $\mathcal{I}$ be a semialgebra, which is not here. For instance $\mathcal{I}$ is not closed under finite intersection as $\{1, 2\}, \{2, 3\} \in \mathcal{I}$ whereas their intersection is not in the collection.

■ **Problem 1.11** Let $(\Omega, \mathcal{M}, \lambda)$ be Lebesgue measure on the interval $[0, 1]$. Let

$$\Omega' = \{(x, y) \in \mathbb{R}^2 : 0 < x \leq 1, 0 < y \leq 1\}.$$

Let $\mathcal{F}$ be the collection of all subsets of Ω' of the form

$$\{(x, y) \in \mathbb{R}^2 : x \in A, 0 < y \leq 1\}$$

for some $A \in \mathcal{M}$. Finally, defined a probability $\mathbb{P}$ on $\mathcal{F}$ by

$$\mathbb{P}\{(x, y) \in \mathbb{R}^2 : x \in A, 0 < y \leq 1\} = \lambda(A).$$

(a) Prove that $(\lambda', \mathcal{F}, \mathbb{P})$ is probability space.

(b) Let $\mathbb{P}^*$ be the outer measure corresponding to $\mathbb{P}$ and $\mathcal{F}$. Define the subset $S \subseteq \Omega'$ by

$$S = \{(x, y) \in \mathbb{R}^2 : 0 < x \leq 1, y = 1/2\}.$$

(Note that $S \notin \mathcal{F}$.) Prove that $\mathbb{P}^*(S) = 1$ and $\mathbb{P}^*(S^c) = 1$.

Solution (a) $\mathcal{F}$ being a σ-algebra follows immediately from $\mathcal{M}$ being a σ-algebra. To see this, for instance let $H \in \mathcal{F}$. Then there exists some $A \in \mathcal{M}$ such that $H = A \times (0, 1]$, where $\times$ is the Cartesian product of two sets. Then $H^c = \Omega' \backslash H$ will be given as $H^c = A^c \times (0, 1]$. Since $A^c \in \mathcal{M}$ then it follows that $H^c \in \mathcal{F}$. With a similar reasoning we can show that $\mathcal{F}$ is a σ-algebra.

Furthermore, $\mathbb{P}$ being a probability measure follows immediately from the fact that λ is a probability measure. For instance $\mathbb{P}(\Omega') = \lambda(\Omega) = 1$, and $\mathbb{P}(B) \geq 0$ for all $B \in \mathcal{F}$ since we can write $B = A \times (0, 1]$ where $A \in \mathcal{M}$ and by definition $\mathbb{P}(A) = \lambda(A) > 0$. Countable additivity also follows from a similar line of reasoning.

(b) From the monotonicity of the outer measure and using the fact that $S \subset \Omega'$ one gets that $\mathbb{P}^*(S) \leq \mathbb{P}^*(\Omega') = \mathbb{P}(\Omega') = 1$. Furthermore, observe that we can write $S = \Omega \times \{1/2\}$. Let $\{I_k \in \mathcal{F}\}$ be any open cover for S. Thus for each I_k there exists $A_k \in \mathcal{M}$ such that $I_k = A_k \times (0, 1]$. Since $S = \Omega \times \{1/2\}$ the $\{B_k\}$ will be an open cover for Ω. Using the fact that $\mathbb{P}(I_k) = \lambda(B_k)$

$$1 = \lambda(\Omega) \leq \sum_k \lambda(B_k) = \sum_k \mathbb{P}(I_k)$$

Since the inequality above holds for any open cover $\{I_k\}$ for S then we can conclude that $1 \leq \mathbb{P}^*(S)$. So far

$$\mathbb{P}^*(S) \geq 1, \qquad \mathbb{P}^*(S) \leq 1.$$

Then it follows that

$$\mathbb{P}^*(S) = 1.$$

■ **Problem 1.12** (a) Where in the proof of Theorem 2.3.1 was assumption 2.3.3 used?

(b) How would the of Theorem 2.3.1 be modified?

Solution (a) It is only used with proving the equality $\mathbb{P}(A) = \mathbb{P}^*(A)$ for $A \in \mathcal{I}$. I.e. to show that $\mathbb{P}^*$ is an extension of $\mathbb{P}$ to a larger domain $\mathcal{M}$ which is a σ-algebra that contain the collection $\mathcal{I}$.

(b) Then the identity $\mathbb{P}^*(A) = \mathbb{P}(A)$ for $A \in \mathcal{I}$ should be replaced with $\mathbb{P}^*(A) \leq \mathbb{P}(A)$.

■ **Problem 1.13** Let $\Omega = \{1, 2\}$, and let $\mathcal{I}$ be the collection of all subsets of Ω, with $\mathbb{P}(\varnothing) = 0, \mathbb{P}(\Omega) = 1$, and $\mathbb{P}\{1\} = \mathbb{P}\{2\} = 1/3$.

(a) Verify that all assumptions of theorem 2.3.1 other than 2.3.3 are satisfied.

(b) Verify that the assumption 2.3.3 is not satisfies.

(c) Describe precisely the $\mathcal{M}$ and $\mathbb{P}^*$ that would result in this example from the modified version of Theorem 2.3.1 in 1.12.

Solution (a) First, notice that the power set of Ω finite, is always an algebra thus semialgebra. So $\mathcal{I}$ is a semialgebra. On the other hand by the hypothesis we have $\mathbb{P}(\varnothing) = 0$ and $\mathbb{P}(\Omega) = 1$ which satisfies some of the conditions in Theorem 2.3.1. For the super-additivity, it holds because

$$\mathbb{P}(\{1\} \cup \{2\}) = 1 > \mathbb{P}\{1\} + \mathbb{P}\{2\} = 2/3.$$

(b) It is easy to check as $\{1,2\},\{1\},\{2\} \in \mathcal{I}$ with $\{1,2\} \subseteq \{1\} \cup \{2\}$ but

$$\mathbb{P}\{1,2\} \not\leq \mathbb{P}\{1\} + \mathbb{P}\{2\}.$$

(c) TODO: TOBEADDED

■ **Problem 1.14** Let $\Omega = [0,1]$. Let $\mathcal{I}'$ be the set of all half-open intervals of the form $(a,b]$, for $0 \leq a < b \leq 1$, togheter with the sets $\varnothing, \Omega$, and $\{0\}$.

(a) Prove that $\mathcal{I}'$ is s semialgebra.

(b) Prove that $\sigma(\mathcal{I}') = \mathcal{B}$, i.e. that the σ-algebra generated by this $\mathcal{I}'$ is equal to the σ-algebra generated by the σ-algebra of (2.4.1) in Rosenthal.

(c) Let $\mathcal{B}_0'$ be the collection of all finite disjoint unions of elements of $\mathcal{I}'$. Is $\mathcal{B}_0'$ the same as the algebra $\mathcal{B}_0$ defined in (2.2.4) Rosenthal?

Solution (a) By definition $\mathcal{I}'$ contains $\varnothing$ and Ω. To show being closed under intersection let $A, B \in \mathcal{I}'$. If A, B are disjoint then $A \cap B \in \mathcal{I}'$. However if A, B are not disjoint, then WLOG we can assume that $A = (a_1, a_2], B = (b_1, b_2]$ where $a_1 < b_1 < a_2 \leq b_2$. Thus $A \cap B = (b_1, a_2]$ which is also at $\mathcal{I}'$. For the last property of a semialgebra, let $A \in \mathcal{I}'$. We can assume $A = (a, b]$ for $0 \leq a < b \leq 1$. Then $A^c = [0, a] \cup (b, 1] = \{0\} \cup (0, a] \cup (b, 1]$ which is a finite disjoint union of elements of $\mathcal{I}'$.

(b) By (2.4.1) Rosenthal, $\mathcal{B}$ is the smallest σ-algebra of all intervals in $[0,1]$ where the term intervals include all open, closed, half-open, intervals as well as the empty set, singletons, and the whole set $[0,1]$. First, notice that $\sigma(\mathcal{I}')$ contains all of the intervals in $[0,1]$. That is because by using complements, countable unions, as well as countable intersections one can construct any kind of intervals using the intervals of the type $(a,b]$. Since $\mathcal{B}$ is the smallest σ-algebra containing $\mathcal{I}$ thus $\sigma(\mathcal{I}) \subseteq \sigma(\mathcal{I}')$. To show the equality, observe that $\mathcal{I}' \subset \mathcal{I}$ thus $\sigma(\mathcal{I}') \subseteq \sigma(\mathcal{I})$. These two inequalities implies that $\sigma(\mathcal{I}') = \sigma(\mathcal{I}) = \mathcal{B}$.

(c) No it is not. Since $(1/3, 1/2) \in \mathcal{B}_0$ but not in $\mathcal{B'}_0$.

■ **Problem 1.15** Let K be the Cantor set. Let $D_n = K \oplus \frac{1}{n}$ be a shifted Cantor set by $1/n$. Let $B = \bigcup_n D_n$.

(a) Draw a rough sketch of D_3.

(b) What is $\lambda(D_3)$?

(c) Draw a rough image of B.

(d) What is $\lambda(B)$?

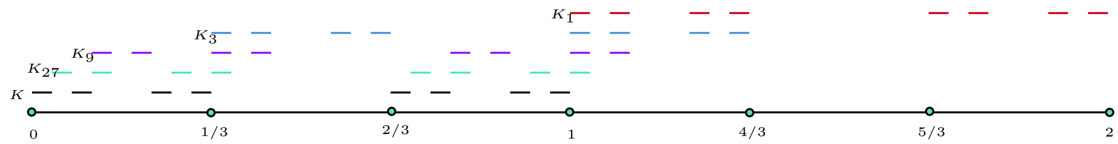
Solution (a) See figure below.

(b) Since $K \in \mathcal{B}$, then it is shift invariant, thus $\lambda(K_3) = \lambda(K) = 0$.

(c) See the figure below.

(d) From the countable sub-additivity we have

$$\lambda(B) = \lambda(\bigcup_n K_n) \leq \sum_n \lambda(K_n) = 0.$$



■ **Remark** Note that $K \oplus K = \bigcup_{x\in K}(K \oplus x) = [0, 2]$. I.e. the set of all numbers that can be created by adding two Cantor numbers is all the numbers in $[0, 2]$. Note that the Cantor set has Lebesgue measure zero, however $[0, 2]$ has measure 2. That is because $\bigcup_{x\in K})$ is in fact an uncountable union of sets (since a Cantor set is uncountable).

■ **Problem 1.16** Let Ω be a finite non-empty set, and let $\mathcal{I}$ consist of all singletons in Ω, together with $\varnothing$ and Ω. Let $p : \Omega \to [0, 1]$ with $\sum_{\omega\in\Omega} p(\omega) = 1$, and define $\mathbb{P}(\varnothing) = 0, \mathbb{P}(\Omega) = 1$, and $\mathbb{P}\{\omega\} = \mathbb{P}(w)$ for all $\omega \in \Omega$.

(a) Prove that $\mathcal{I}$ is a semialgebra.

(b) Prove that (2.3.2) and (2.3.3) are satisfied.

(c) Describe precisely the $\mathcal{M}$ and $\mathbb{P}^*$ that result from applying Theorem 2.3.1 in Rosenthal.

(d) Are these $\mathcal{M}$ and $\mathbb{P}^*$ the same as those described in Theorem 2.2.1 in Rosenthal?

Solution (a) By definition $\mathcal{I}$ contains $\varnothing$ and well as Ω. $\mathcal{I}$ is also closed under finite intersection as the intersection of two singletons is either a singleton or the empty set, and the intersection of Ω with any singleton is a singleton. Furthermore, the intersection of any singleton with empty set is the empty set that is contained in $\mathcal{I}$. Finally, let $E \in \mathcal{I}$. If E is either Ω or the empty set, then it complement can trivially be written as the disjoint union of $\varnothing$ or Ω respectively. If E is a singleton, then E^c can be written as the disjoint union of the singleton of its elements. Thus $\mathcal{I}$ is a semialgebra.

(b) To check (2.3.2) let $A_1, \cdots, A_k \in \mathcal{I}$ disjoint with $\bigcup_i A_i \in \mathcal{I}$. Then $\bigcup_i A_i = \Omega$ the collection A_i's are all of the singletons. Thus

$$1 = \mathbb{P}(\bigcup_i A_i) = \mathbb{P}(\Omega) = \sum_i \mathbb{P}(A_i) = \sum_{\omega\in\Omega} \mathbb{P}(\{\omega\}) = \sum_{\omega\in\Omega} p(\omega) = 1.$$

Thus 2.3.2 holds with equality

To verify (2.3.3) let $A, A_1, A_2, \cdots, A_k \in \mathcal{I}$ with $A \subset \bigcup_i A_i$. If A is empty set, then (2.3.3) holds as $0 \le a$ for all $a \in [0, 1]$. If A is Ω, then the the sets A_i are the sets of all singletons. Thus 2.3.3 holds as $1 \le 1$. Lastly, if A is a singleton, then at least one of A_i's should be the same as A. Then $\mathbb{P}(A) \le \mathbb{P}(A_1) + \cdots + \mathbb{P}(A_j) + \cdots + \mathbb{P}(A_k)$ for some $0 \le j \le n$. Since $\mathbb{P}(A_j) = \mathbb{P}(A)$ then 2.3.3 holds.

(c) The collection $\mathcal{M}$ will be the same as the power set of Ω. And the probability measure $\mathbb{P}^*$ will be give as

$$\mathbb{P}^*(A) = \sum_{\omega\in A} p(\omega).$$

(d) Although $\mathcal{M}$ is the same as in theorem 2.3.1, but $\mathbb{P}^*$ is not the same. The probability measure defined in Theorem 2.3.1 is the uniform probability measure, where here it is not. The probability measure $\mathbb{P}^*$ is a more general one and will be the same as probability measure in Theorem 2.3.1 if we choose $p(\omega) = 1/|\Omega|$.

■ **Problem 1.17** Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures defined on the same sample space Ω and σ-algebra $\mathcal{F}$.

(a) Suppose that $\mathbb{P}(A) = \mathbb{Q}(A)$ for all $A \in \mathcal{F}$ with $\mathbb{P}(A) \leq 1/2$. Prove that $\mathbb{P} = \mathbb{Q}$, i.e. that $\mathbb{P}(A) = \mathbb{Q}(A)$ for all $A \in \mathcal{F}$.

(b) Give an example where $\mathbb{P}(A) = \mathbb{Q}(A)$ for all $A \in \mathcal{F}$ with $\mathbb{P}(A) < 1/2$, such that $\mathbb{P} \neq \mathbb{Q}$, i.e. that $\mathbb{P}(A) \neq \mathbb{Q}(A)$ for some $A \in \mathcal{F}$.

Solution (a) Let $A \in \mathcal{F}$ with $\mathbb{P}(A) > 1/2$, hence $\mathbb{P}(A^c) \leq 1/2$. Also, let $E \in \mathcal{F}$ such that $\mathbb{P}(E) \leq 1/2$. By definition (by using 2.3.7 or using the fact that $\mathcal{F}$ is a σ-algebra thus closed under intersection and complement) we can write

$$\mathbb{P}(A) = \mathbb{P}(A \cap E) + \mathbb{P}(A \cap E^c).$$

Observe that $A \cap E \subseteq E$ thus by monotonicity $\mathbb{P}(A \cap E) \leq \mathbb{P}(E) \leq 1/2$. Further more, we can write $\mathbb{P}(A \cap E^c) = 1 - \mathbb{P}(A^c \cup E)$. For the second term in the RHS we have

$$\mathbb{P}(A^c \cup E) = \mathbb{P}(A^c) + \mathbb{P}(E) - \mathbb{P}(A^c \cap E).$$

Note that $\mathbb{P}(A^c) \leq 1/2$ as well as since $A^c \cap E \subset A^c$ thus by monotonicity $\mathbb{P}(A^c \cap E) \leq \mathbb{P}(A^c) \leq 1/2$. Thus

$$\mathbb{P}(A) = \mathbb{P}(A \cap E) + 1 - \mathbb{P}(A^c) - \mathbb{P}(E) + \mathbb{P}(A^c \cap E).$$

For all the terms in the RHS, since their measure with respect to $\mathbb{P}$ is less than oe equal to 1/2, thus $\mathbb{P}$ and $\mathbb{Q}$ agrees on them. Thus

$$\mathbb{P}(A) = \mathbb{Q}(A \cap E) + 1 - \mathbb{Q}(A^c) - \mathbb{Q}(E) + \mathbb{Q}(A^c \cap E) = \mathbb{Q}(A).$$

This completes the proof.

An easier solution. Let $A \in \mathcal{F}$ with $\mathbb{P}(A) > 1/2$. Then $\mathbb{P}(A) = 1 - \mathbb{P}(A^c) = 1 - \mathbb{Q}(A^c) = \mathbb{Q}(A)$, where we used the fact that $\mathbb{P}(A^c) \leq 1/2$ thus $\mathbb{P}(A^c) = \mathbb{Q}(A^c)$.

(b) Let $\Omega = \{1, 2\}$ with $\mathcal{F}$ being the power set of Ω. Then defined

$$\mathbb{P}(\varnothing) = 0, \quad \mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = 1/2, \quad \mathbb{P}(\{1, 2\}) = 1.$$

And

$$\mathbb{Q}(\varnothing) = 0, \quad \mathbb{Q}(\{1\}) = 1/10, \quad \mathbb{Q}(\{2\}) = 9/10, \quad \mathbb{Q}(\{1, 2\}) = 1.$$

■ **Remark** The hypothesis in part (a) in question above means that $\mathbb{P}$ and $\mathbb{Q}$ on all of the sets that has measure less than or equal to 1/2 w.r.t $\mathbb{P}$, must agree on all of the element of $\mathcal{F}$.

■ **Problem 1.18** Let $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ be Lebesgue measure on $[0, 1]$. Consider a second probability triple $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$, defined as follows: $\Omega_2 = \{1, 2\}$, $\mathcal{F}_2$ consists of all subsets of Ω_2, and $\mathbb{P}_2$ is defined by $\mathbb{P}_2\{1\} = 1/3$ and $\mathbb{P}_2\{2\} = 2/3$, and additivity. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the product measure of $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ and $(\Omega_2, \mathcal{F}_\in, \mathbb{P}_2)$.

(a) Express each of $\Omega, \mathcal{F}$ and $\mathbb{P}$ as explicitly as possible.

(b) Find a set $A \in \mathcal{F}$ such that $\mathbb{P}(A) = 3/4$.

Solution (a) The set Ω is

$$\Omega = \{1, 2\} \times [0, 1].$$

The collection $\mathcal{F}$ is given by

$$\mathcal{F} = \{\{1\} \times B : B \in \mathcal{B}\} \quad \cup \quad \{\{2\} \times B : B \in \mathcal{B}\} \quad \cup \quad \{\{1, 2\} \times B : B \in \mathcal{B}\}.$$

And $\mathbb{P}$ is given by

$$\mathbb{P}(\{1\} \times B) = \lambda(B)/3, \quad \mathbb{P}(\{2\} \times B) = 2\lambda(B)/3, \quad \mathbb{P}(\{1, 2\} \times B) = \lambda(B).$$

(b) One easy choice for such a set would be $A = \{1, 2\} \times (0, 3/4)$.

1.2 Further Probabilistic Foundations

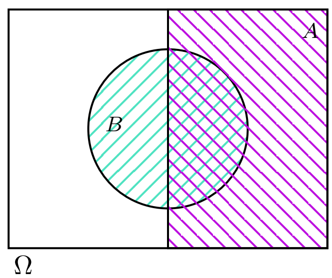
Definition 1.3 Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $E_1, E_2 \in \mathcal{F}$ be two events. Then E_1 and E_2 are said to be independent events if and only if we have

$$\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1)\mathbb{P}(E_2).$$

Another way to formulate this is to write

$$\mathbb{P}(E_1 | E_2) = \mathbb{P}(E_1).$$

■ **Remark** The following diagram is very suggestive to make an intuition to see how does two independent events (i.e. sets) look like.



The intuition is that the proportion of A in the restricted world B (i.e. $\mathbb{P}(A|B)$) is the same as the proportion of A in the whole world. I.e.

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)}{\mathbb{P}(\Omega)} = \frac{\mathbb{P}(A)}{1} = \mathbb{P}(A).$$

Proposition 1.3 Let A, B be two independent events, i.e. $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Then the pair of events, (A^c, B^c), (A^c, B), and (A, B^c) are also independent.

Proof. We start by showing that A^c, B are independent events. Observe that

$$\mathbb{P}(A^c \cap B) = 1 - \mathbb{P}(A \cup B^c) = 1 - (\mathbb{P}(B^c) + \mathbb{P}(A \cap B)) = \mathbb{P}(B) + \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A^c)\mathbb{P}(B).$$

Thus A^c and B are independent events. Similarly we can prove that A, B^c are independent events. To show that the event A^c, B^c are independent we have

$$\begin{aligned}\mathbb{P}(A^c \cap B^c) &= 1 - \mathbb{P}(A \cup B) = 1 - \mathbb{P}(A) - \mathbb{P}(B) + \mathbb{P}(A \cap B) = \mathbb{P}(A^c) - \mathbb{P}(B) + \mathbb{P}(A)\mathbb{P}(B) \\ &= \mathbb{P}(A^c) - \mathbb{P}(B)(1 - \mathbb{P}(A)) = \mathbb{P}(A^c)(1 - \mathbb{P}(B)) = \mathbb{P}(A^c)\mathbb{P}(B^c).\end{aligned}$$

Thus A^c and B^c are independent. □

Proposition 1.4 Let $A_1, A_2, \cdots$ be a sequence of independent events. Then the sequence of events $B_1, B_2, \cdots$ are also independent where B_i is either equal to A_i or A_i^c. In particular $A_1^c, A_2^c, \cdots$ is a sequence of independent events.

Proof. Use the result of Proposition 1.3 with Exercise 3.2.2 in Rosenthal. □

Observation 1.2.1 Consider the following problem. Let $A_1, A_2, \cdots, B_1, B_2, \cdots$ be events.

(a) Prove that
$$(\limsup A_n) \cap (\limsup_n B_n) \supseteq \limsup_n (A_n \cap B_n).$$

(b) Give an example where the above inclusion is strict, and another example where it holds with equality.

Here, I will give a very intuitive explanation of the meaning of $\limsup_n A_n$ as well as $\limsup_n B_n$. Consider $A_1, A_2, \cdots$ as sum lamps in a row that if $w \in A_i$ then then ith lamp turns on. Thus $\limsup_n A_n$ are those elements in Ω that if we evaluate its presence in the sequence of sets, a pattern will emerge where as you move further in the row of lamps you will still find a lamp that is on. Similarly for $\limsup_n B_n$. However, the meaning of $\omega \in (A_n \cap B_n)$ is that there is lamp position i such that this lamp is on for both A_n sequence and B_n sequence. Thus $\limsup_n (A_n \cap B_n)$ is the set of all elements that if you evaluate its presence in the A_n sequence and B_n sequence, no matter how far you move in the sequence, you will still find spots where both lamps for A_i and B_i are on.

Proposition 1.5 — Borel-Cantelli Lemma. Let $A_1, A_2, \cdots \in \mathcal{F}$.

(a) If $\sum_n \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(\limsup_n A_n) = 0$.

(b) If $\sum_n \mathbb{P}(A_n) = \infty$, and the events are *independent*, then $\mathbb{P}(\limsup_n A_n) = 1$.

1.2.1 Solved Problems

■ **Problem 1.19** Let X be a real-valued random variable defined on a probability triple $(\Omega, \mathcal{F}, \mathbb{P})$. Fill in the following blanks:

(a) $\mathcal{F}$ is a collection of subsets of ____________.

(b) $\mathbb{P}(A)$ is a well-defined element of ____________provided that A is an element of ____________.

(c) $\{X \leq 5\}$ is shorthand notation for the particular subset of ____________which is defined by ____________.

(d) If S is a subset of ____________, then $\{X \in S\}$ is a subset of ____________.

(e) If S is a ____________subset of ____________, then $\{X \in S\}$ must be a element of ____________.

Solution

1. Ω.
2. $\mathbb{R}$, $\mathcal{F}$.
3. Ω, $\{\omega \in \Omega : X(\omega) \leq 5\}$.
4. $\mathcal{B}$, $\mathcal{F}$.
5. Borel, $\mathbb{R}$, σ-algebra $\mathcal{F}$.

■ **Problem 1.20** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be Lebesgue measure on $[0, 1]$. Let $A = (1/2, 3/4)$ and $B = (0, 2/3)$. Are A and B independent events?

Solution Yes. It is easy to check that $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ holds for A, B as above.

■ **Problem 1.21** Give an example of events A, B, and C, each of probability strictly between 0 and 1, such that

(a) $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, $\mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C)$, and $\mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C)$; but it is not the case that $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$.

(b) $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B), \mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C)$, and $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$; but it is not the case that $\mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C)$.

Solution (a) Let $\Omega = \{a, b, c, d\}$ and $\mathbb{P}$ a uniform discrete distribution on Ω. Let $A = \{a, b\}, B = \{a, c\}, C = \{a, d\}$. Then we have

$$\mathbb{P}(A \cap B) = \mathbb{P}(A \cap C) = \mathbb{P}(B \cap C) = \mathbb{P}(\{a\}) = \frac{1}{4} = \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(C) = \mathbb{P}(B)\mathbb{P}(C).$$

However,

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\{a\}) \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) = \frac{1}{8}.$$

(b) Let $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8\}$ and $\mathbb{P}$ a uniform discrete distribution on Ω. Define

$$A = \{1, 2, 3, 4\}, \quad B = \{3, 4, 5, 6\}, \quad C = \{1, 3, 5, 6\}.$$

Then we have $\mathbb{P}(A \cap B) = \mathbb{P}(\{3, 4\}) = \mathbb{P}(A)\mathbb{P}(B) = \frac{1}{4}$. Also $\mathbb{P}(A \cap C) = \mathbb{P}(\{1, 3\}) = \mathbb{P}(A)\mathbb{P}(C) = \frac{1}{4}$. Furthermore $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) = \frac{1}{8}$. However $\mathbb{P}(B \cap C) = \mathbb{P}(\{3, 5, 6\}) = \frac{3}{8} \neq \mathbb{P}(B)\mathbb{P}(C) = \frac{1}{4}$.

■ **Problem 1.22** Suppose $\{A_n\} \nearrow A$. Let $f : \Omega \to \mathbb{R}$ be any function. Prove that $\lim_{n\to\infty} \inf_{\omega \in A_n} f(\omega) = \inf_{\omega \in A} f(\omega)$.

Solution Since $\{A_n\} \nearrow A$ then $A_1 \subseteq A_2 \subseteq \cdots$ with $\bigcup_n A_n = A$. Let $f(A_i) \subset \mathbb{R}$ denote the image of set A_i under the map f. Then we have $f(A_1) \subseteq f(A_2) \subseteq \cdots$. Thus from properties of the infimum we have $\inf f(A_1) \geq \inf f(A_2) \geq \cdots$. Observe that $A_i \subset A$ for all $i \in \mathbb{N}$, thus $\inf(A_i) \geq \inf(A)$ for all $i \in \mathbb{N}$. Thus the sequence of real numbers $\{\inf f(A_n)\}$ is a decreasing sequence bounded from below. By monotone convergence theorem we conclude that $\lim_{n\to\infty} f(A_n)$ exists. Then the next step is to show that this limit is the same as $\inf f(A)$. Let $\epsilon > 0$ given. Then there is $\alpha \in f(A)$ such that $\alpha < \inf f(A) + \epsilon$. Because $f(A) = f(\bigcup_n A_n) = \bigcup_n f(A_n)$, then $\exists m \in \mathbb{N}$ such that $\alpha \in f(A_m)$. Then $\inf f(A_m) \leq \alpha$. Thus $\inf f(A_m) \leq \inf f(A) + \epsilon$. Since we can find such A_m for every $\epsilon > 0$, and since $\inf f(A_n)$ is a decreasing sequence bounded below, then by the definition of limit we get

$$\lim_{n\to\infty} \inf_{A_n} f(A_n) = \inf_{A} f(A).$$

■ **Problem 1.23** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability triple such that Ω is countable, and $\mathcal{F} = 2^{\Omega}$. Prove that it is impossible for there to exist a sequence $A_1, A_2, \dots \in \mathcal{F}$ which is *independent*, such that $\mathbb{P}(A_i) = \frac{1}{2}$ for each i.

Solution Let $\omega \in \Omega$. Then define the sequence B_n as

$$B_n = \begin{cases} A_n & \omega \in A_n, \\ A_n^c & \omega \notin A_n. \end{cases}$$

Then by Proposition 1.4 the sequence of events B_n are also independent. and we have $\mathbb{P}(B_n) = 1/2$ for all n. By construction we have $\omega \in \bigcap_n B_n$. Thus

$$\mathbb{P}(\{\omega\}) \leq \mathbb{P}(\bigcap_n B_n) = \prod_n \mathbb{P}(B_n) = \frac{1}{2^n}$$

Since this is true for all $n \in \mathbb{N}$ then $\mathbb{P}(\{\omega\}) = 0$ for all $\omega \in \Omega$. However, since Ω is countable we have $\Omega = \bigcup_{\omega\in\Omega}\{\omega\}$ which is a disjoint countable union. By countable additivity of the probability measure we have

$$1 = \mathbb{P}(\Omega) = \sum_{\omega\in\Omega} \mathbb{P}(\omega) = 0,$$

which is a contradiction.

Solution — A second solution!. Since $\mathbb{P}(A_i) = 1/2$ for all i, we have $\sum_i \mathbb{P}(A_i) = \infty$. Noting that A_i's are independent then by Borel-Cantelli we have

$$\mathbb{P}(\{A_n \text{ i.o.}\}) = 1.$$

Observe that $\{A_n \text{ i.o.}\} \subseteq \Omega$, thus it is at most countable. So it can be written as a disjoint union of the singletons of its element. Applying the countable additivity of $\mathbb{P}$ will result in $1 = 0$ which is a contradiction.

■ **Problem 1.24** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the uniform distribution on $\Omega = \{1, 2, 3\}$ as in Example 2.2.2. Give an example of a sequence $A_1, A_2, \cdots \in \mathcal{F}$ such that

$$\mathbb{P}(\liminf_n A_n) < \liminf_n \mathbb{P}(A_n) < \limsup_n \mathbb{P}(A_n) < \mathbb{P}(\limsup_n A_n).$$

Solution An easy choice for such a sequence is

$$\{1\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1\}, \cdots .$$

It is easy to see that $\liminf_n A_n = \varnothing$, $\limsup_n A_n = \{1, 2, 3\}$, $\liminf_n \mathbb{P}(A_n) = 1/3$, and $\limsup_n \mathbb{P}(A_n) = 2/3$. Thus we will get

$$\mathbb{P}(\liminf_n A_n) < \liminf_n \mathbb{P}(A_n) < \limsup_n \mathbb{P}(A_n) < \mathbb{P}(\limsup_n A_n).$$

■ **Problem 1.25** Let λ be Lebesgue measure on $[0, 1]$, and let $0 \leq a \leq b \leq c \leq d \leq 1$ be arbitrary real numbers. Give an example of a sequence $A_1, A_2, \cdots$ of subsets of $[0, 1]$, such that $\lambda(\liminf_f A_n) = a$, $\liminf_f \lambda(A_n) = b$, $\limsup_n \lambda(A_n) = c$, and $\lambda(\limsup_n A_n) = d$. (*Hint: Start with the case $d = b + c - a$, which is easiest, and then carefully branch out from there.*)

Solution I am not sure how to use the hint, but one of the examples I could construct is considering the events

$$A_n = (\frac{1}{4} + \frac{1}{8}\sin(n), \frac{3}{4} - \frac{1}{8}\sin(n)).$$

It is easy to see that

$$\limsup_n A_n = (\frac{1}{8}, \frac{7}{8}), \quad \liminf_n A_n = (\frac{3}{8}, \frac{5}{8}).$$

Thus we have

$$\lambda(\limsup_n A_n) = \frac{3}{4}, \qquad \lambda(\liminf_n A_n) = \frac{1}{4}.$$

Furthermore

$$\limsup_n \lambda(A_n) = \frac{3}{4}, \quad \liminf_n \lambda(A_n) = \frac{1}{4},$$

Which satisfies the requirements.

■ **Problem 1.26** Let $A_1, A_2, \cdots, B_1, B_2, \cdots$ be events.

(a) Prove that
$$(\limsup A_n) \cap (\limsup_n B_n) \supseteq \limsup_n (A_n \cap B_n).$$

(b) Give an example where the above inclusion is strict, and another example where it holds with equality.

Solution (a) Let $\omega \in \limsup_n (A_n \cap B_n)$. Then $\forall N > 0$ there exists $n > N$ such that $\omega \in A_n \cap B_n$. By definition this implies that $\omega \in (\limsup_n A_n \cap \limsup_n B_n)$, and this completes the proof.

(b) Fix $\omega \in \Omega$. Let E be any set that contains ω. Let $A_{2n} = E$ and $A_{2n+1} = E^c$ for all $n \in \mathbb{N}$. Furthermore let $B_{2n} = E^c$ while $B_{2n+1} = E$. By this construction we have
$$\{\omega\} \in \limsup_n A_n, \quad \{\omega\} \in \limsup_n B_n, \quad \text{thus} \quad \{\omega\} \in (\limsup_n A_n) \cap (\limsup_n B_n).$$
However, since $A_n \cap B_n = \varnothing$ for all n we have
$$\omega \notin \limsup_n (A_n \cap B_n),$$
which shows the strict inequality of the inequality we proved in (a).

To show an example which which the equality works, let $E \subset \Omega$ by any subset. Define $B_{bn} = A_{an} = E$ where $(a, b) = 1$ (i.e. are relatively prime), and $\varnothing$ otherwise. Then any $\omega \in \limsup_n A_n$ will belong to one A_i every a sets, and by design will belong to B_i every b sets. Since $(a, b) = 1$, then these sets will be the same infinitely often, hence $\omega \in \limsup(A_n \cap B_n)$ as well.

■ **Problem 1.27** Let $A_1, A_2, \cdots$ be a sequence of events, and let $N \in \mathbb{N}$. Suppose there are events B, C such that $B \subseteq A_n \subseteq C$ for all $n \geq N$, and such that $\mathbb{P}(B) = \mathbb{P}(C)$. Prove that $\mathbb{P}(\liminf_n A_n) = \mathbb{P}(\limsup_n A_n) = \mathbb{P}(B) = \mathbb{P}(C)$.

Solution We claim
$$B \subseteq \limsup_n A_n \subseteq C, \qquad B \subseteq \liminf_n A_n \subseteq C.$$
To show the first statement, let $\omega \in B$. Then since for all n large enough we have $B \subseteq A_n \subseteq C$, we have $B \in A_n$, thus $B \subseteq \limsup_n A_n$. Furthermore, let $\omega \in \limsup_n A_n$. Then for all N we can find $n > N$ such that $\omega \in A_n$. By hypothesis $\omega \in C$. Thus $\limsup_n A_n \subseteq C$.
To show the second statement, let $\omega \in C$. Then since for all n large enough $B \subseteq A_n \subseteq C$ we see that $B \subseteq \liminf_n A_n \subseteq C$. By the monotonicity of the probability we have
$$\mathbb{P}(B) \leq \mathbb{P}(\limsup_n A_n) \leq \mathbb{P}(C), \qquad \mathbb{P}(B) \leq \mathbb{P}(\liminf_n A_n) \leq \mathbb{P}(C).$$
Since $\mathbb{P}(B) = \mathbb{P}(C)$ then we conclude that
$$\mathbb{P}(\limsup_n A_n) = \mathbb{P}(\liminf_n A_n) = \mathbb{P}(B) = \mathbb{P}(C).$$

■ **Problem 1.28** Let $\{X_n\}$ be independent random variables, with $\mathbb{P}(X_n = i) = 1/n$ for $i = 1, 2, 3, \cdots, n$. Compute $\mathbb{P}(X_n = 5\ i.o.)$, the probability that an infinite number of the X_n are equal to 5.

Solution Let $A_n = \{X_n = 5\}$. Then by hypothesis $\mathbb{P}(A_n) = 0$ for $n < 5$ and $\mathbb{P}(A_n) = 1/n$ for $n \geq 5$. Thus $\sum_n \mathbb{P}(A_n) = \infty$. As the random variables X_n are all independent, so is the events A_n. Thus using the Borel-Cantelli Lemma we conclude that $\mathbb{P}(A_n\ i.o.) = 1$.

■ **Problem 1.29** Let X be a random variable with $\mathbb{P}(X > 0) > 0$. Prove that there exists $\delta > 0$ such that $\mathbb{P}(X \geq \delta) > 0$. (*Hint: Use the continuity of the probability function*)

Solution Let $A = \{X > 0\}$, and consider the events $A_n = \{X > 1/n\}$. Observe that $\{A_n\} \nearrow A$. The sequence of real numbers $\{\mathbb{P}(A_n)\}$ is an increasing sequence (by the monotonicity) and is converging to $\mathbb{P}(A)$ (by the continuity). Let $\epsilon = \mathbb{P}(A)/2$. Then by the definition of the convergence of real numbers for all $N > 0$ we have $\epsilon < \mathbb{P}(A_n)$ for all $n > N$. Let $\delta = 1/N$. This completes the proof.

■ **Problem 1.30** Let $X_1, X_2, \cdots$ be defined jointly on some probability space $(\Omega, \mathcal{F}, \mathcal{P})$, with $\sum_{i=1}^{\infty} i^2 \mathbb{P}(i \leq X_n < i+1) \leq C < \infty$ for all n. Prove that $\mathbb{P}(X_n \geq n\ i.o.) = 0$.

Solution Observe that

$$\begin{aligned}
C &\geq \sum_{i=1}^{\infty} i^2 \mathbb{P}(i \leq X_n < i+1) \\
&\geq \sum_{i=n}^{\infty} i^2 \mathbb{P}(i \leq X_n < i+1) \\
&\geq n^2 \sum_{i=n}^{\infty} \mathbb{P}(i \leq X_n < i+1) \\
&= n^2 \mathbb{P}(X_n \geq n).
\end{aligned}$$

Thus we have $\mathbb{P}(X_n \geq n) \leq C/n^2$. Since C/n^2 is summable, and the probability function is positive, then $\mathbb{P}(X_n \geq n)$ is also summable. I.e.

$$\sum_n \mathbb{P}(X_n \geq n) < \infty.$$

Using Borel-Cantelli lemma we then have

$$\mathbb{P}(X_n \geq n\ i.o.) = 0.$$

■ **Problem 1.31** Let $\delta, \epsilon > 0$, and let $X_1, X_2, \cdots$ be a sequence of independent non-negative random variables such that $\mathbb{P}(X_i \geq \delta) \geq \epsilon$ for all i. Prove that with probability one, $\sum_{i=1}^{\infty} X_i = \infty$. I.e. $\mathbb{P}(\sum_{i=1}^{\infty} X_i = \infty) = 1$.

Solution Let $A_n = \{X_n \geq \delta\}$. Since $\mathbb{P}(A_n) \geq \epsilon$ for all n, then

$$\sum_n \mathbb{P}(A_n) = \infty.$$

Since the random variables $X_1, X_2, \cdots$ are independent, the events $A_1, A_2, \cdots$ are also independent. Thus by Borel-Cantelli we have

$$\mathbb{P}(A_n\ i.o.) = 1.$$

We claim

$$\limsup_n A_n \subseteq \{\sum_n X_n = \infty\}.$$

To see this let $\omega \in \limsup_n A_n$. Then it means that $\forall m > 0$ there exists $n_m > m$ such that $\omega \in A_{n_m}$, or equivalently $X_{n_m}(\omega) \geq \epsilon$. Considering the subsequence $\{X_{n_m}\}$ we see that $\sum_m X_{n_m}(\omega) = \infty$. Thus the inclusion above holds. By the monotonicity of the probability and using the result above

$$1 = \mathbb{P}(\limsup_n A_n) \leq \mathbb{P}(\{\sum_n X_n = \infty\}).$$

Thus we conclude that

$$\mathbb{P}(\{\sum_n X_n = \infty\}) = 1.$$

■ **Problem 1.32** Consider infinite, independent, fair coin tossing, and let H_n be the event that the n^{th} coin is heads. Determine the following probabilities.

(a) $\mathbb{P}(H_{n+1} \cap H_{n+2} \cap \cdots \cap H_{n+9} \ i.o.)$

(b) $\mathbb{P}(H_{n+1} \cap H_{n+2} \cap \cdots \cap H_{2n} \ i.o.)$

(c) $\mathbb{P}(H_{n+1} \cap H_{n+2} \cap \cdots \cap H_{n+2\log_2 n} \ i.o.)$

(d) Prove that $\mathbb{P}(H_{n+1} \cap H_{n+2} \cap \cdots \cap H_{n+\log_2 n} \ i.o.)$ must equal either 0 or 1.

(e) Determine $\mathbb{P}(H_{n+1} \cap H_{n+2} \cap \cdots \cap H_{n+\log_2 n} i.o.)$

Solution (a) Let $A_0 = H_1 \cap \cdots \cap H_9, A_1 = H_2 \cap \cdots \cap H_{10}, A_2 = H_3 \cap \cdots \cap H_{11}, A_3 = H_4 \cap \cdots \cap H_{12}$, and so on. Then $A_1, A_2, \cdots$ are not necessarily independent, but there is a subsequence $A_0, A_{10}, A_{20}, \cdots$ that are independent. Observe that $\mathbb{P}(A_n) = 1/2^{10}$. Thus

$$\infty = A_0 + A_{10} + A_{20} + \cdots \leq \sum_n \mathbb{P}(A_n)$$

Thus $\mathbb{P}(A_{10k} \ i.o.) = 1$. Since $\{A_{10k} \ i.o.\} \subseteq \{A_k \ i.o.\}$, by monotonicity of probability $\mathbb{P}(A_{10k} \ i.o.) \leq \mathbb{P}(A_k \ i.o.)$. This implies that $\mathbb{P}(A_k \ i.o.) = 1$, i.e.

$$\mathbb{P}(H_{n+1} \cap \cdots \cap H_{n+9} \ i.o.) = 1.$$

(b) Observe that $\mathbb{P}(H_{n+1} \cap \cdots \cap H_{2n}) = \frac{1}{2^n}$, which is summable.

$$\sum_n \mathbb{P}(H_{n+1} \cap \cdots \cap H_{2n}) < \infty.$$

This implies that $\mathbb{P}(H_{n+1} \cap \cdots \cap H_{2n}) = 0$.

(c) The probability $\mathbb{P}(H_{n+1} \cap \cdots \cap H_{n+2\log_2 n})$ is approximately $(\frac{1}{2})^{\log_2 n^2} = \frac{1}{n^2}$ which is summable. Thus by Borel-Cantelli $\mathbb{P}(H_{n+1} \cap \cdots \cap H_{n+2\log_2 n} \ i.o.) = 0$.

(d) Since $\{H_{n+1} \cap \cdots \cap H_{n+\log_2 n}\}$ is a tail even, then by Kolmogorov zero-one law the probability is either zero or one.

(e) It is suggestive to write down some of the event explicitly. Let

$$\begin{aligned}
&A_2 = H_3,\ A_3 = H_4, \\
&A_4 = H_5 \cap H_6,\ \cdots,\ A_7 = H_8 \cap H_9, \\
&A_8 = H_9 \cap H_{10} \cap H_{11},\ \cdots,\ A_{15} = H_{16} \cap H_{17} \cap H_{18} \\
&A_{16} = H_{17} \cap H_{18} \cap H_{19} \cap H_{20},\ \cdots,\ A_{31} = H_{32} \cap H_{33} \cap H_{34} \cap H_{35}, \\
&A_{32} = H_{33} \cap H_{34} \cap H_{35} \cap H_{36} \cap H_{37},\ \cdots,\ A_{63} = H_{64} \cap H_{65} \cap H_{66} \cap H_{67} \cap H_{67},
\end{aligned}$$

and so on.

Note that each A_{2^k} up to $A_{2^{k+1}}$ we have probability $\frac{1}{2^k}$. However, we can roughly find $2^k/k$ independent events among them. Thus

$$\infty = \sum_k \frac{2^k/k}{2^k} \leq \sum_n \mathbb{P}(A_n).$$

Thus $\mathbb{P}(A_n \ i.o.) = 1$.

1.3 Expected Values

1.3.1 Solved Problems

■ **Problem 1.33** Let X be a random variable with finite mean, and let $a \in \mathbb{R}$ be any real number. Prove that

$$\mathbb{E}\Big[\max(X,a)\Big] \geq \max(\mathbb{E}\Big[X\Big], a).$$

Hint: Consider separately the cases $\mathbb{E}\Big[X\Big] \geq a$ *and* $\mathbb{E}\Big[X\Big] < a$.

Solution First observe that we can write

$$X = X\mathbb{1}_{X\geq a} + a\mathbb{1}_{X<a}.$$

So by the linearity of the expectation we can write

$$\mathbb{E}\Big[\max(X,a)\Big] = \mathbb{E}\Big[X\mathbb{1}_{X\geq a}\Big] + \mathbb{E}\Big[a\mathbb{1}_{X<a}\Big].$$ (𝅗𝅥)

For the first term observe that

$$\mathbb{E}\Big[X\mathbb{1}_{X\geq a}\Big] \geq \mathbb{E}\Big[a\mathbb{1}_{X\geq a}\Big],$$ (♩)

while for the second term

$$\mathbb{E}\Big[a\mathbb{1}_{X<a}\Big] \geq \mathbb{E}\Big[X\mathbb{1}_{X<a}\Big].$$ (♪)

By (♩) and (𝅗𝅥) we get

$$\mathbb{E}\Big[\max(X,a)\Big] \geq \mathbb{E}\Big[a\mathbb{1}_{X\geq a}\Big] + \mathbb{E}\Big[a\mathbb{1}_{X<a}\Big] = \mathbb{E}\Big[a\Big] = a.$$

And by (♪) and (𝅗𝅥)

$$\mathbb{E}\Big[\max(X,a)\Big] \geq \mathbb{E}\Big[X\mathbb{1}_{X<a}\Big] + \mathbb{E}\Big[X\mathbb{1}_{X\geq a}\Big] = \mathbb{E}\Big[X\Big].$$

These two equations imply that

$$\mathbb{E}\Big[\max(X,a)\Big] \geq \max(\mathbb{E}\Big[X\Big], a).$$

1.4 Inequalities and Convergence

Proposition 1.6 — Markov's inequality. Let X be a *non-negative* random variable. Let $\alpha > 0$, then

$$\mathbb{P}(X \geq \alpha) \leq \frac{\mathbb{E}\Big[X\Big]}{\alpha}$$

■ **Remark** Note that the random variable being *non-negative* is the key for the inequality to hold.

Definition 1.4 — Almost Surely Convergence. Let $X_1, X_2, \cdots$ be a sequence of random variables. Then we say $\{X_n\}$ is converging to the random variable X *almost surely* if

$$\mathbb{P}(X_n \to X) = 1,$$

where the arrow notation show a point-wise convergence.

■ **Remark** The almost sure convergence is when we have a point-wise convergence on a set of measure 1.

Definition 1.5 — Convergence in probability. Let $X, X_1, X_2, \cdots$ be random variables. We say X_n converges to X in *probability* if for all $\epsilon > 0$ we have

$$\mathbb{P}(|X_n - X| \geq \epsilon) \to 0 \quad \text{as } n \to \infty.$$

Observation 1.4.1 For a given ϵ, the sequence of reals formed for each n by

$$\mathbb{P}(|X_n - X| \geq \epsilon)$$

is very important. Its convergence (and the rate of convergence up to the summability of the sequence) will draw lines between the almost sure convergence and the convergence in probability. To The following proposition and corollary will make this more clear.

The idea of the proof of the following Lemma is very important and will show up again in the future.

Lemma 1.1 Let $\{X_n\}$ be a sequence of random variables and X a random variable. If $\forall \epsilon > 0$ we have

$$\mathbb{P}(|X_n - X| \geq \epsilon \ i.o.) = 1,$$

then

$$\mathbb{P}(X_n \to X) = 1,$$

i.e. X_n converges to X almost surely.

Proof. Consider

$$\mathbb{P}(X_n \to X) = \mathbb{P}(\bigcup_{r=1}^{\infty} \{|X_n - X| \leq \epsilon \ a.a.\}) = 1 - \mathbb{P}(\bigcup_{r=1}^{\infty} \{|X_n - X| \leq \epsilon \ a.a.\}).$$

From the countable sub-additivity of the probability measure we know that

$$\mathbb{P}(\bigcup_{r=1}^{\infty} \{|X_n - X| \leq \epsilon \ a.a.\}) \leq \sum_{r=1}^{\infty} \mathbb{P}(\{|X_n - X| \leq \epsilon \ a.a.\}) = 0.$$

Thus

$$\mathbb{P}(X_n \to X) = 1 - \mathbb{P}(\bigcup_{r=1}^{\infty} \{|X_n - X| \leq \epsilon \ a.a.\}) = 1.$$

□

■ **Remark — Important!.** In the proof above we used the following important fact

$$\boxed{\{X_n \to X\} = \bigcup_{r=1}^{\infty} \{|X_n - X| \leq 1/r \ a.a.\}}$$

or equivalently

$$\boxed{\{X_n \to X\} = \bigcup_{r\in\mathbb{Q}^+} \{|X_n - X| \leq q \ a.a.\}}$$

This is literally the definition of the point wise convergence. To see this first remember that

$$\{|X_n - X| \leq \epsilon \ a.a.\} = \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} \{|X_n - X| \leq \infty\}.$$

So if we let $\omega \in \{X_n \to X\}$ then from the definition of the convergence we know that $\forall N \in \mathbb{N}$ we can find $n > N$ such that $|Z_n(\omega) - Z(\omega)| \leq \epsilon$. I.e. by the definition of union and intersection

$$\omega \in \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} \{|X_n - X| \leq \infty.\}$$

To show the converse let $\omega \in \bigcup_{n=N}^{\infty}\{|X_n - X| \leq \infty.\}$. So $\forall N \in \mathbb{N}$ there exists $n \leq N$ such that $\omega \in \{|X_n - X| \leq \epsilon\}$. I.e. $|X_n(\omega) - X(\omega)| \leq \epsilon$ which is precisely the definition of the point wise convergence $X_n(\omega) \to X(\omega)$.

Corollary 1.1 Let $\{X_n\}$ be a sequence of random variables and X be a random variable. Then if for all $\epsilon > 0$

$$\sum_n \mathbb{P}(|X_n - X| \geq \epsilon) < \infty,$$

then X_n converges to X almost surely.

Proof. Apply Borel-Cantelli along with Lemma 1.1. □

Summary 1.1 Let $X, X_1, X_2, \cdots$ be random variables. Consider the sequence of reals given by

$$a_n = \mathbb{P}(|X_n - X| \geq \epsilon).$$

If for any $\epsilon > 0$ we have

$$a_n \to 0$$

as $n \to \infty$ then by Definition 1.5 X_n converges to X is probability. However, if the sequence is also summable for any ϵ, i.e.

$$\sum_n a_n < \infty,$$

then X_n converges to X almost surely.

1.4.1 Solved Problems

■ **Problem 1.34** Give an example of a random variable X and $\alpha > 0$ such that $\mathbb{P}(X \geq \alpha) > \mathbb{E}\Big[X\Big]/\alpha$. (*Hint: Obviously X can not be non-negative*). Where does the proof of the Markov inequality break down in this case?

Solution Consider $(\Omega, \mathcal{F}, \mathbb{P})$ where $\Omega = \{1, 2, 3, 4\}$ with uniform probability measure $\mathbb{P}$, and $\mathcal{F} = 2^\Omega$. Define the random variable $X : \Omega \to \mathbb{R}$ as

$$X(1) = -1, \ X(2) = -2, \ X(3) = 3, \ X(4) = 4.$$

It is easy to calculate

$$\mathbb{E}\Big[X\Big] = (-1-2+3+4)\cdot\frac{1}{4} = 1.$$

However

$$\mathbb{P}(X \geq 3) = \mathbb{P}(X=3) + \mathbb{P}(X=4) = \frac{1}{2}.$$

It is clear that the Markov inequality does not hold.

■ **Problem 1.35** Suppose X is a non-negative random variable with $\mathbb{E}\Big[X\Big] = \infty$. What does Markov's inequality can say in this case?

Solution Then the Markov's inequality will be trivially true,

$$\mathbb{P}(X \geq \alpha) \leq \infty.$$

■ **Problem 1.36** For general jointly defined random variables X and Y prove that $|\operatorname{Corr}(X,Y)| \leq 1$. (*Hint: Don't forget the Cauchy-Schwartz inequality.*)

Solution Recall the formula for Corr

$$\operatorname{Corr}(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X), \operatorname{Var}(Y)}}.$$

Note that $\operatorname{Var}(X) = \operatorname{Cov}(X,X)$, and in general

$$\operatorname{Cov}(X,Y) = \mathbb{E}\Big[(X-\mu_X)(Y-\mu_Y)\Big].$$

From C-S inequality we have

$$|\operatorname{Cov}(X,Y)| = \mathbb{E}\Big[|(X-\mu_X)(Y-\mu_Y)|\Big] \leq \sqrt{\mathbb{E}\Big[(X-\mu_X)^2\Big]\mathbb{E}\Big[(Y-\mu_Y)^2\Big]} = \sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}.$$

Thus it immediately follows that

$$|\operatorname{Corr}(X,Y)| \leq 1.$$

■ **Remark** Note the similarities between Corr can cos as defined for vectors

$$\cos(\theta) = \frac{a\cdot b}{\sqrt{(a\cdot a)(b\cdot b)}}.$$

for which we also have

$$|\cos\theta| \leq 1.$$

■ **Problem 1.37** Let $\varphi(x) = x^2$.

(a) Prove that φ is a convex function.

(b) What does Jensen's inequality say for this choice of φ?

(c) Where in the text have we already see the result of part (b)?

Solution (a) This follows from

$$\begin{aligned}\Big(\lambda a^2 + (1-\lambda)b^2\Big) - \Big(\lambda a + (1-\lambda)b\Big)^2 &= \lambda a^2 + (1-\lambda)b^2 - \lambda^2 a^2 - (1-\lambda)^2 b^2 - 2ab\lambda(1-\lambda) \\ &= \lambda a^2(1-\lambda) + (1-\lambda)b^2(1-(1-\lambda)) - 2ab\lambda(1-\lambda) \\ &= \lambda(1-\lambda)(a^2+b^2-2ab) = \lambda(1-\lambda)(b-a)^2 \geq 0.\end{aligned}$$

Thus it follows that

$$\Big(\lambda a^2 + (1-\lambda)b^2\Big) \geq \Big(\lambda a + (1-\lambda)b\Big)^2$$

(b) Jensen's inequality with this choice of φ will result in

$$\mathbb{E}\Big[X\Big]^2 \leq \mathbb{E}\Big[X^2\Big].$$

(c) From above and using the definition of variance we have

$$\mathrm{Var}(X) = \mathbb{E}\Big[X^2\Big] - \mathbb{E}\Big[X\Big]^2 \geq 0.$$

Thus this implies that $\mathrm{Var}(X)$ is always positive.

■ **Problem 1.38** Let $X_1, X_2, \cdots$ be a sequence of random variables, with $\mathbb{E}\Big[X_n\Big] = 8$ and $\mathrm{Var}(X_n) = 1/\sqrt{n}$

Solution Using Chebychev's inequality we can write

$$a_n = \mathbb{P}(|X_n - X| \geq \epsilon) \leq \frac{\mathrm{Var}(X_n)}{\epsilon^2} = \frac{1}{\sqrt{n}\epsilon^2}.$$

We see that for any choice of ϵ the sequence $a_n \to 0$ as $n \to \infty$. Thus by definition the sequence converges to 0 is probability.

■ **Remark** Observe that the sequence a_n is not summable for any choice of $\epsilon > 0$. Thus by Summary 1.1 we see that X_n does *not* converge to X almost surely.

■ **Problem 1.39** Give (with proof) an example of two discrete random variables having the same mean and the same variance, but which are not identically distributed.

Solution We demonstrate this by giving an explicit example. Let $(\Omega, \mathcal{F}, \mathbb{P})$ where $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{F} = 2^\Omega$, and

$$\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \frac{4}{20}, \quad \mathbb{P}(\{3\}) = \mathbb{P}(\{4\}) = \mathbb{P}(\{5\}) = \mathbb{P}(\{6\}) = \frac{3}{20}.$$

Define the random variables X, Y as

$$X(1) = -2,\ X(2) = 2,\ X(3) = 0,\ X(4) = 0,\ X(5) = -1,\ X(6) = 1,$$

and

$$Y(1) = -1,\ Y(2) = 1,\ Y(3) = -2,\ Y(4) = 2,\ Y(5) = -1,\ Y(6) = 1.$$

It is easy to check that

$$\mathbb{E}\Big[X\Big] = 0, \qquad \mathbb{E}\Big[Y\Big] = 0.$$

And

$$\mathrm{Var}(X) = \frac{38}{20}, \qquad \mathrm{Var}(Y) = \frac{38}{20}.$$

Consider the following permutation

$$\sigma = \begin{pmatrix} -2 & -1 & 0 & 1 & 2 \\ 0 & 1 & -2 & 2 & -1 \end{pmatrix},$$

and define the function f be the extension of σ on $\mathbb{R}$ where it assume the value 0 for all points in its domain other than $\{-2, -1, 0, 1, 2\}$. The we will have

$$f(X(1)) = 0,\ f(X(2)) = -1,\ f(X(3)) = -2,\ f(X(4)) = -2,\ f(X(5)) = 1,\ f(X(6)) = 2,$$

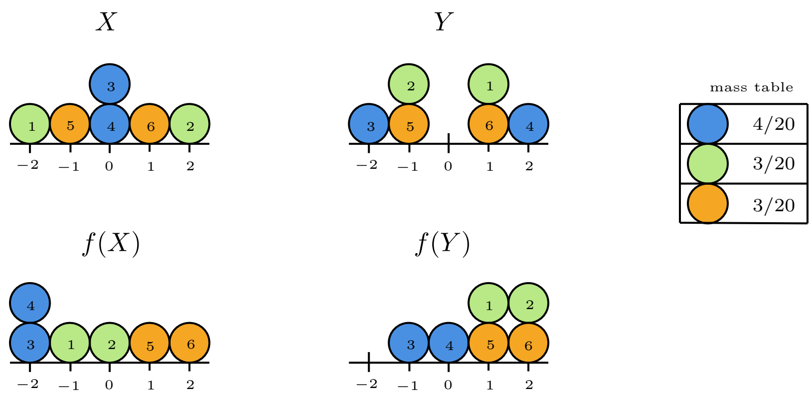
and

$$f(Y(1)) = 1,\ f(Y(2)) = 2,\ f(Y(3)) = 0,\ f(Y(4)) = -1,\ f(Y(5)) = 1,\ f(Y(6)) = 2.$$

It is easy to calculate

$$\mathbb{E}\Big[f(X)\Big] = \frac{-7}{20}, \qquad \mathbb{E}\Big[f(Y)\Big] = \frac{18}{20}.$$

The following diagram summarizes the whole idea! Note that we can also find other Borel



measurable function. For instance you can figure out a continuous function that behaves differently on positive numbers, vs. negative numbers.

■ **Problem 1.40** Prove that if $\{Z_n\}$ converges to X almost surely, then for each $\epsilon > 0$ we have $\mathbb{P}(|Z_n - Z| \geq \epsilon\ i.o.) = 0$.

Solution We start with

$$\begin{aligned} 1 = \mathbb{P}(Z_n \to Z) &= \mathbb{P}(\bigcap_{q\in\mathbb{Q}^+} \{|Z_n - Z| < q\ a.a.\}) \\ &= 1 - \mathbb{P}(\bigcup_{q\in\mathbb{Q}^+} \{|Z_n - Z| \geqq\ \ i.o.\}) \end{aligned}$$

Thus

$$\mathbb{P}(\bigcup_{q\in\mathbb{Q}^+} \{|Z_n - Z| \geq\ \ i.o.\}) = 0.$$

Observe that for all $\epsilon > 0$

$$\{|Z_n - Z| \geq \epsilon\ i.o.\} \subseteq \bigcup_{q\in\mathbb{Q}} \{|Z_n - Z| \geq q\ i.o.\}.$$

Thus by monotonicity of the probability function we will have

$$\mathbb{P}(\{|Z_n - Z| \geq \epsilon\ i.o.\}) \leq \mathbb{P}(\bigcup_{q\in\mathbb{Q}} \{|Z_n - Z| \geq q\ i.o.\}) = 0.$$

So

$$\mathbb{P}(\{|Z_n - Z| \geq \epsilon\ i.o.\}) = 0.$$

■ **Problem 1.41** Let $X_1, X_2, \cdots$ be a sequence of random variables that are independent. Then prove that

$$X_n \to X \; a.s. \implies \sum_n \mathbb{P}(|X_n - X| > \epsilon) < \infty.$$

Solution We do by the proof by contrapositive. We want to proof

$$\exists \epsilon > 0 \quad \sum_n \mathbb{P}(|X_n - X| > \epsilon) = \infty \implies P(X_n \to X) = 0$$

Since X_n's are independent, then by Borel-Cantelli we have

$$\mathbb{P}(|X_n - X| > \epsilon \; i.o.) = 1.. \tag{1}$$

This implies

$$\mathbb{P}(\bigcup_{q \in \mathbb{Q}^+} \{|X_n - X| > q \; i.o.\}) = 1. \tag{2}$$

(if this is not clear for you see the remark below)

Thus

$$\mathbb{P}(\bigcap_{r \in \mathbb{Q}^+} \{|X_n - X| \leq q \; a.a.\}) = 0.$$

which is equivalent to

$$\mathbb{P}(X_n \to X) = 0. \quad \square$$

■ **Remark** To see why (1) implies (2) see the below:

$$\{|X_n - X| > \epsilon \; i.o.\} \subseteq \bigcup_{q \in \mathbb{Q}} \{|X_n - X| > q \; i.o.\}.$$

By monotonicity of the probability measure

$$1 = \mathbb{P}(\{|X_n - X| > \epsilon \; i.o.\}) \leq \mathbb{P}(\bigcup_{q \in \mathbb{Q}^+} \{|X_n - X| > q \; i.o.\}).$$

Thus

$$\mathbb{P}(\bigcup_{q \in \mathbb{Q}^+} \{|X_n - X| > q \; i.o.\}) = 1.$$

■ **Problem 1.42** Let $X_1, X_2, \cdots$ be a sequence of independent random variable with $\mathbb{P}(X_n = 3^n) = \mathbb{P}(X_n = 3^{-n}) = \frac{1}{2}$. Let $S_n = X_1 + \cdots + X_n$.

(a) Compute $\mathbb{E}\Big[X_n\Big]$ for each n.

(b) For $n \in \mathbb{N}$, compute R_n defined as

$$R_n = \sup\{r \in \mathbb{R} : \mathbb{P}(|S_n| \geq r) = 1\},$$

i.e. the largest number such that $|S_n|$ is always at least R_n.

(c) Compute $\lim_{n \to \infty} R_n / n$.

(d) For which $\epsilon > 0$ (if any) is it the case that $\mathbb{P}(\frac{1}{n}|S_n| \geq \epsilon) \not\to 0$?

(e) Why does this result not contradict the various laws of large numbers?

Solution (a) Observe that X_n is a simple random variable, thus

$$\mathbb{E}\Big[X_n\Big] = \frac{1}{2}3^n - \frac{1}{2}3^n = 0.$$

(b) The value of R_n equal to the case where for S_n all the random variables $X_1, X_2, \cdots, X_{n-1}$ assume their lowest value i.e. $-3, -9, \cdots, -3^{n-1}$ respectively, and the random variable X_n assumes its largest value, i.e. 3^n. So we will have

$$S_n = 3^n - (3 + 9 + \cdots + 3^{n-1}) = 3^n - \frac{3^n - 3}{2} = \frac{3^n + 3}{2}.$$

(c) From what we had in part (b) we can see that

$$\lim_{n\to\infty} \frac{R_n}{n} = \lim_{n\to\infty} \frac{3^n + 3}{2n} = \infty.$$

(d) There is no $\epsilon > 0$ that works. That is because we know that $|S_n| \geq R_n$ almost everywhere. On the other hand, in part (b) we observed that $R_n/n \to \infty$ as $n \to \infty$. So for all $\epsilon > 0$ the sequence of reals $\mathbb{P}(|S_n|/n \geq \epsilon)$ will converge to zero.

(e) This does not contradicts any of the law of large number. That is because the variance of X_n is

$$\mathrm{Var}(X_n) = \mathbb{E}\Big[X^2\Big] - \mathbb{E}\Big[X\Big]^2 = 3^n,$$

which does not remain finite as $n \to \infty$.

■ **Problem 1.43** Suppose $\mathbb{E}\Big[2^X\Big] = 4$. Prove that $\mathbb{P}(X \geq 3) \leq 1/2$.

Solution — The first method. We can start with

$$\mathbb{E}\Big[2^X\Big] = \mathbb{E}\Big[2^X \mathbb{1}_{X\geq 3}\Big] + \mathbb{E}\Big[2^X \mathbb{1}_{X<3}\Big] = 4.$$

For the first and the second terms we have the following bounds

$$\mathbb{E}\Big[2^X \mathbb{1}_{X\geq 3}\Big] \geq 8\mathbb{P}(X \geq 3),$$

So

$$8\mathbb{P}(X \geq 3) \leq \mathbb{E}\Big[2^X \mathbb{1}_{X\geq 3}\Big] = 4 - \mathbb{E}\Big[2^X \mathbb{1}_{X<3}\Big].$$

This implies that

$$\mathbb{P}(X \geq 3) \leq \frac{1}{2} - \mathbb{E}\Big[2^X \mathbb{1}_{X<3}\Big] \leq \frac{1}{2},$$

where we have used the fact that the expectation of a non-negative random variable is positive.

■ **Remark — Connection to the conditional expectation.** If one has already used the conditional expectation before, then they can recover the definition of it using the answer above.

$$\mathbb{E}\Big[X\mathbb{1}_B\Big] = \mathbb{E}\Big[X|B\Big]\mathbb{E}\Big[\mathbb{1}_B\Big] = \mathbb{E}\Big[X|B\Big]\mathbb{P}(B).$$

In fact the conditional expectation is a way to factor the expectation when the random variable is of the form $X\mathbb{1}_A$.

Solution — The second method. Define the random variable

$$Y = 2^X.$$

Thus $X = \log(Y)$, where log is considered to be base 2 logarithm function. Thus

$$\mathbb{P}(X \geq 3) = \mathbb{P}(\log(Y) \geq 3) = \mathbb{P}(Y \geq 9) \leq \frac{\mathbb{E}\big[Y\big]}{9}\frac{4}{9} \leq \frac{4}{8} = \frac{1}{2},$$

where we have used the Markov's inequality. Note that the random variable Y is non-negative and that is the reason we could use the Markov's inequality.

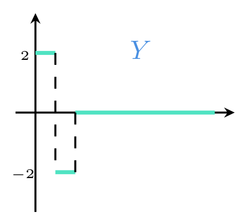
■ **Problem 1.44** Give examples of random variables Y with mean 0 and variance 1 such that

(a) $\mathbb{P}(|Y| \geq 2) = 1/4$.

(b) $\mathbb{P}(|Y| \geq 2) < 1/4$.

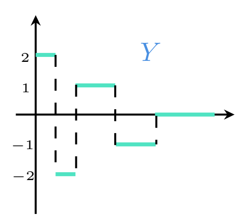
Solution For all of the examples below let $([0,1], \mathbb{B}, \lambda)$ be the probability space.

(a) An easy example is
$$X = 2\mathbb{1}_{[0,1/8)} - 2\mathbb{1}_{(1/8,2/8]}.$$
The graph of this random variable would look like as the following.



(b) The following example is a straightforward one.

$$X = 2\mathbb{1}_{[0,1/16)} - 2\mathbb{1}_{[1/16,2/16)} + \mathbb{1}_{(2/16,6/16]} - \mathbb{1}_{(6/16,10/16]}.$$



■ **Remark** Note that from the Chebychev inequality we have

$$\mathbb{P}(|Y - \mu_Y| \geq \alpha) \leq \frac{\mathrm{Var}(Y)}{\alpha^2}.$$

The problem above demonstrates the cases where the equality holds, and the strict inequality holds.

■ **Problem 1.45** Suppose Y is a random variable with finite mean μ_Y and $\mathrm{Var}(Y) = \infty$. What does Chebychev's inequality say in this case?

Solution The Chebychev's inequality states that for any random variable X we have

$$\mathbb{P}(|X - \mu_X| \geq \alpha) \leq \frac{\mathrm{Var}(X)}{\alpha^2}.$$

When $\mathrm{Var}(X) = \infty$, then the Chebychev's inequality will be the trivial inequality

$$\mathbb{P}(|X - \mu_X| \geq \alpha) \leq \infty.$$

■ **Problem 1.46** Give (with proof) an example of a sequence $\{Y_n\}$ of jointly defined random variables, such that as $n \to \infty$ we have all of three

(i) Y_n/n converges to 0 in probability.

(ii) Y_n/n^2 converges to 0 with probability 1.

(iii) Y_n/n does *not* converge to 0 with probability 1.

Solution Let $([0,1], \mathcal{B}, \lambda)$ be the probability space, and define the sequence of random variables $\{Y_n\}$ given as

$$Y_n = n\mathbb{1}_{A_n},$$

where $A_n \in \mathcal{B}$ with $\mathbb{P}(A_n) = 1/n$. Then we will have

$$Y_n/n = \mathbb{1}_{A_n}, \qquad Y_n/n^2 = \mathbb{1}_{A_n}/n.$$

Then

$$\mathbb{P}(|Y_n/n| \geq \epsilon) = \begin{cases} 0 & \epsilon > 0 \\ 1/n & 0 < \epsilon \leq 1 \end{cases}, \qquad \mathbb{P}(|Y_n/n^2| \geq \epsilon) = 0 \qquad \text{for } n \text{ large enough.}$$

Thus we can see that Y_n/n converges to 0 in probability, Y_n/n does not converge to 0 with probability one (as $\mathbb{P}(|Y_n/n| \geq \epsilon)$ is not summable). Lastly, Y_n/n^2 converges to 0 with probability 1.

■ **Problem 1.47** Let $r \in \mathbb{N}$. Let $X_1, X_2, \cdots$ be independently distributed random variables having finite mean m, which are r-dependent, i.e. such that $X_{k_1}, X_{k_2}, \ldots, X_{k_j}$ are independent whenever $k_{i+1} > k_i + r$ for each i (thus independent random variables are 0-dependent). Prove that with probability one $\frac{1}{n}\sum_{i=1}^n X_i \to m$ as $n \to \infty$. (*Hint: break up the sum into $r+1$ different sums.*)

Solution Let $S_n = \frac{1}{n}\sum_{i=1}^n X_i$ represent the partial sums of the infinite sum. Let $R = r + 1$. We are interested in the subsequence terms of the form S_{Rn}. These terms can be written as

$$\begin{aligned} S_{nR} &= \frac{1}{R} \cdot \frac{1}{n}(X_1 + X_{1+R} + X_{1+2R} + \cdots + X_{1+(n-1)R}) \\ &+ \frac{1}{R} \cdot \frac{1}{n}(X_2 + X_{2+R} + X_{2+2R} + \cdots + X_{2+(n-1)R}) \\ &\vdots \\ &+ \frac{1}{R} \cdot \frac{1}{n}(X_R + X_{R+R} + X_{R+2R} + \cdots + X_{R+(n-1)R}). \end{aligned}$$

Let

$$Y_{i,n} = \frac{1}{R} \cdot \frac{1}{n}(X_i + X_{i+R} + X_{i+2R} + \cdots + X_{i+(n-1)R}).$$

Note that the terms in each $Y_{i,n}$ are i.i.d. random variables by the hypothesis and all of them have mean n, thus by the SLLN we have

$$\mathbb{P}(Y_{i,n} \to m) = 1 \qquad \forall i \in \mathbb{N}.$$

On the other hand, the decomposition of the sum above we can write

$$S_{nR} = \frac{1}{R} \sum_{i=1}^{R} Y_{i,n}.$$

Each of the terms in the RHS goes to m with probability one, thus the LHS goes to $Rm/R = m$ with probability 1.

■ **Remark** This is a very clear generalization of the notion of the independent random variables. In this general notion the random variables are r-dependent if they are independent whenever they are r terms away from each other. This notion will be very useful in considering the systems who have an exponentially decaying memory through time.

1.5 Distribution of Random Variables

Proposition 1.7 — Cumulative distribution function is right-continuous. Let X be a random variable whose cumulative distribution function is defined as $F_X(x) = \mathbb{P}(X \leq x)$. Then $F_X(x)$ is right-continuous.

Proof. Fix $x \in \mathbb{R}$. Let $\{x_n\}$ be any sequence that $x_n \to x$ as $n \to \infty$ and for all n $x \leq x_n$. Observe that

$$\bigcap_n (-\infty, x_n] = (-\infty, x], \tag{♪}$$

and also

$$(-\infty, x_1] \supseteq (-\infty, x_2] \supseteq (-\infty, x_3] \supseteq \cdots. \tag{♫}$$

Thus we have $\{(-\infty, x_n]\} \downarrow (-\infty, x]$ as $n \to \infty$. Since (♪) and (♫) are both preserved by the pre-image X^{-1}, so we also have

$$\{X^{-1}(-\infty, x_n]\} \downarrow X^{-1}((-\infty, x]) \quad \text{as } n \to \infty.$$

From continuity of probability we have

$$\mathbb{P}(X \leq x_n) \to \mathbb{P}(X \leq x) \quad \text{as } n \to \infty.$$

Thus by the definition of cumulative distribution we have

$$F_X(x_n) \to F_X(x) \quad \text{as } n \to \infty.$$

□

■ **Remark** Note that we can not do a similar argument to show that $F_X(x)$ is left continuous. The obstacle is that if we choose $\{x_n\}$ such that $x_n \to x$ and for all n we have $x_n \leq x$, then the union

$$\bigcup_n (-\infty, x_n]$$

is not necessarily equal to $(-\infty, x]$. For instance if we let $x_n = x - 1/n$ then

$$\bigcup_n (-\infty, x_n] = (-\infty, x).$$

Proposition 1.8 Let X and Y be two random variables (possibly defined on different probability triples). Then $\mu_X = \mu_Y$ if and only if $\mathbb{E}\Big[f(X)\Big] = \mathbb{E}\Big[f(Y)\Big]$ for all Borel-measurable $f : \mathbb{R} \to \mathbb{R}$ for which either expectation is well-defined.

Proof. To show the $\boxed{\Longrightarrow}$, using the change of variable theorem we see that

$$\mathbb{E}\Big[f(X)\Big] = \int_{\mathbb{R}} f d\mu_X = \int_{\mathbb{R}} f d\mu_Y = \mathbb{E}\Big[f(Y)\Big].$$

To show the $\boxed{\Longleftarrow}$, let $B \in \mathcal{B}$, and let $f = \mathbb{1}_B$. Then

$$\mathbb{E}\Big[f(X)\Big] = \mathbb{E}\Big[\mathbb{1}_B(X)\Big] = \mathbb{E}\Big[\mathbb{1}_{X \in B}\Big] = \mathbb{P}(X \in B) = \mu_X.$$

with a similar reasoning we get $\mathbb{E}\Big[f(Y)\Big] = \mu_Y$. Then since $\mathbb{E}\Big[f(X)\Big] = \mathbb{E}\Big[f(Y)\Big]$,

$$\mu_X = \mu_Y.$$

□

Corollary 1.2 If X and Y are random variables with $\mathbb{P}(X = Y) = 1$, then

$$\mathbb{E}\Big[f(X)\Big] = \mathbb{E}\Big[f(Y)\Big]$$

for all Borel-measurable $f : \mathbb{R} \to \mathbb{R}$ for which either expectation is well-defined.

Proof. We claim that $\mathbb{P}(X = Y) = 1$ implies $\mu_X = \mu_Y$. To see this let $B \in \mathcal{B}$. Then

$$\begin{aligned}\mu_X(B) = \mathbb{P}(\{X \in B\}) &= \mathbb{P}(\{X \in B\} \cap \{X = Y\} \dot{\cup} \{X \in B\} \cap \{X \neq Y\}) \\ &= \mathbb{P}(\{X \in B\} \cap \{X = Y\}) + \mathbb{P}(\{X \in B\} \cap \{X \neq Y\})\end{aligned}$$

Note that $\{X \in B\} \cap \{X \neq Y\} \subseteq \{X \neq Y\}$ thus from the monotonicity $\mathbb{P}(\{X \in B\} \cap \{X \neq Y\}) \leq \mathbb{P}(\{X \neq Y\}) = 0$. Thus

$$\mu_X(B) = \mathbb{P}(\{X \in B\}) = \mathbb{P}(\{X \in B\} \cap \{X = Y\}) = \mathbb{P}(\{Y \in B\} \cap \{X = Y\}) = \mathbb{P}(\{Y \in B\}) = \mu_Y(B).$$

Now using the proposition above we get that

$$\mathbb{E}\Big[f(X)\Big] = \mathbb{E}\Big[f(Y)\Big].$$

□

1.5.1 Solved Problems

■ **Problem 1.48** Let μ have density $4x^3 \mathbb{1}_{0<x<1}$, and let ν have density $x/2 \mathbb{1}_{0<x<2}$.

(a) Compute $\mathbb{E}\Big[X\Big]$ where $\mathcal{L}(X) = \mu/3 + 2\nu/3$.

(b) Compute $\mathbb{E}\Big[Y^2\Big]$ where $\mathcal{L}(Y) = 1/6\mu + 1/3\delta_2 + 1/2\delta_5$.

(c) Compute $\mathbb{E}\Big[Z^3\Big]$ where $\mathcal{L}(Z) = 1/8\mu + 1/8\nu + 1/4\delta_3 + 1/2\delta_4$.

Solution (a) Denote $\mathcal{L}(X)$ with η. Then by definition and then using the change of variable principle we have

$$\mathbb{E}\Big[X\Big] = \int_\Omega X\mathbb{P}(d\omega) = \int_\mathbb{R} x\eta(dx) = 1/3\int_\mathbb{R} x\mu(dx) + 2/3\int_\mathbb{R} x\nu(dx)$$

where for the last equality we have used the proposition 6.2.1 Rosenthal. Now using proposition 6.2.3 Rosenthal we can write

$$\mathbb{E}\Big[X\Big] = 4/3\int_0^1 x^4\lambda(dx) + 1/3\int_0^2 x^2\lambda(dx) = 4/15 + 8/9 = 52/45.$$

where λ is the Lebesgue measure.

(b) Similar to the reasoning in part (a), we will get

$$\begin{aligned}\mathbb{E}\Big[X^2\Big] &= 1/6\int_\mathbb{R} x^2\mu(dx) + 1/3\int_\mathbb{R} x^2\delta_2(dx) + 1/2\int_R x^2\delta_5(dx)\\ &= 2/3\int_0^1 x^5\lambda(dx) + 1/3\int_{-\infty}^{+\infty} x^2\delta(x-2)\lambda(dx) + 1/2\int_{-\infty}^{+\infty} x^2\delta(x-5)\lambda(dx)\\ &= 2/18 + 4/3 + 25/2 = 251/18.\end{aligned}$$

(c) Similar to the part (a) we will get

$$\begin{aligned}\mathbb{E}\Big[X^3\Big] &= 1/8\int_\mathbb{R} x^3\mu(dx) + 1/8\int_\mathbb{R} x^3\nu(dx) + 1/4\int_\mathbb{R} x^3\delta_4(dx) + 1/2\int_\mathbb{R} x^3\delta_4(dx)\\ &= 1/2\int_0^1 x^6\lambda(dx) + 1/16\int_0^2 x^4\lambda(dx) + 1/4\int_{-\infty}^{+\infty} x^3\delta(x-3)\lambda(dx) + 1/2\int_{-\infty}^{+\infty} x^3\delta(x-4)\lambda(dx)\\ &= 1/14 + 32/80 + 27/4 + 32.\end{aligned}$$

■ **Problem 1.49** Suppose $\mathbb{P}(Z=0) = \mathbb{P}(Z=1) = \frac{1}{2}$, that $Y \sim \mathcal{N}(0,1)$, and that Y and Z are independent. Set $X = YZ$. What is the law of X?

Solution Using the conditional probability (see the proposition below), and by letting $B \in \mathbb{B}$ we can write

$$\mu_X(B) = \mathbb{P}(X \in B) = \mathbb{P}(YZ \in B) = \mathbb{P}(0 \in B)\mathbb{P}(Z=0) + \mathbb{P}(Y \in B)\mathbb{P}(Z=1) = \frac{\delta_0(B) + \mu_\mathcal{N}(B)}{2}.$$

Thus the law of X is given by

$$\mu_X = \frac{\delta_0 + \mu_\mathcal{N}}{2}.$$

Observation 1.5.1 — Conditional probability and conditional expectation. The notion of conditional probability and conditional expectation are so natural to work with that one often forgets what is the rigorous basis of these notions. Let $B, C \in \mathcal{F}$ be two sets that partitions the samples space Ω. Then for any $A \in \mathcal{F}$ we can write

$$A = (A\cap B)\dot{\cup}(A\cap C).$$

Then using the countable additivity of the probability measure we can write

$$\mathbb{P}(A) = \mathbb{P}(A\cap B) + \mathbb{P}(A\cap C).$$

We can then define the notion of the conditional probability that enables us to factor the terms in RHS in some useful sense. Define

$$\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B), \qquad \mathbb{P}(A|C) = \mathbb{P}(A \cap C)/\mathbb{P}(C).$$

Then we can write

$$\boxed{\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|C)\mathbb{P}(C).}$$

We can do the same thing with the notion of the indicator function that is a dual notion for the set theoretic operations we had above. Again, let B, C defined as above, and let X be any random variable. Then

$$X = X\mathbb{1}_B + X\mathbb{1}_C,$$

where we have used the fact that

$$\mathbb{1}_B + \mathbb{1}_C = 1 \quad \text{since } \Omega = B\dot{\cup}C.$$

Then

$$\mathbb{E}\Big[X\Big] = \mathbb{E}\Big[X\mathbb{1}_B\Big] + \mathbb{E}\Big[X\mathbb{1}_C\Big].$$

Now we define the notion of the conditional expectation that enables us to factor the terms in the RHS. Define

$$\mathbb{E}\Big[X|B\Big] = \mathbb{E}\Big[X\mathbb{1}_B\Big]/\mathbb{E}\Big[\mathbb{1}_B\Big] = \mathbb{E}\Big[X\mathbb{1}_B\Big]/\mathbb{P}(B),$$
$$\mathbb{E}\Big[X|C\Big] = \mathbb{E}\Big[X\mathbb{1}_C\Big]/\mathbb{E}\Big[\mathbb{1}_C\Big] = \mathbb{E}\Big[X\mathbb{1}_C\Big]/\mathbb{P}(C).$$

Then we can write

$$\boxed{\mathbb{E}\Big[X\Big] = \mathbb{E}\Big[X|B\Big]\mathbb{P}(B) + \mathbb{E}\Big[X|C\Big]\mathbb{P}(C).}$$

■ **Problem 1.50** Let $X \sim \text{Poisson}(5)$.

(a) Compute $\mathbb{E}\Big[X\Big]$.

(b) Compute $\text{Var}(X)$.

(c) Compute $\mathbb{E}\Big[3^X\Big]$.

Solution (a) To find $\mathbb{E}\Big[X\Big]$ we can start with

$$\begin{aligned}
\mathbb{E}\Big[X\Big] &= \sum_{n=0}^{\infty} n \frac{e^{-\lambda}\lambda^n}{n!} \\
&= e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^n}{(n-1)!} \\
&= \lambda e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} \\
&= \lambda e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{(n)!} \\
&= \lambda e^{-\lambda} e^{\lambda} \\
&= \lambda,
\end{aligned}$$

where we have used the fact that $e^\lambda = \sum_{n=0}^{\infty} \lambda^n/n!$.

(b) To harvest the possible cancellations, we start with

$$\begin{aligned}\mathbb{E}\Big[X(X-1)\Big] &= \sum_{n=1}^{\infty} n(n-1)\frac{e^{-\lambda}\lambda^n}{n!} \\ &= e^{-\lambda}\sum_{n=2}^{\infty}\frac{\lambda^n}{(n-2)!} \\ &= \lambda^2 e^{-\lambda}\sum_{n=2}^{\infty}\frac{\lambda^{n-2}}{(n-2)!} \\ &= \lambda^2 e^{-\lambda}\sum_{n=0}^{\infty}\frac{\lambda^n}{n!} \\ &= \lambda^2 e^{-\lambda}e^{\lambda} \\ &= \lambda^2.\end{aligned}$$

Thus $\mathbb{E}\Big[X(X-1)\Big] = \mathbb{E}\Big[X^2\Big] - \mathbb{E}\Big[X\Big] = \lambda^2$, which implies

$$\mathbb{E}\Big[X^2\Big] = \lambda_2 + \lambda.$$

Now using the formula for variance we will get

$$\mathrm{Var}(X) = \mathbb{E}\Big[X^2\Big] - \mathbb{E}\Big[X\Big]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

(c) We can write

$$\mathbb{P}(3^X) = \sum_{n=0}^{\infty} 3^n \mathbb{P}(X=n) = \sum_{n=0}^{\infty} 3^n \frac{e^{-\lambda}\lambda^n}{n!} = e^{-\lambda}\sum_{n=0}^{\infty}(3\lambda)^n/n! = e^{-\lambda}e^{3\lambda} = e^{2\lambda}.$$

■ **Problem 1.51** Compute $\mathbb{E}\Big[X\Big], \mathbb{E}\Big[X^2\Big]$ and ere $\mathrm{Var}(X)$ where the law of X is given by

(a) $\mathcal{L}(X) = \delta_1/2 + \lambda/2$, where λ is Lebesgue measure on $[0,1]$.

(b) $\mathcal{L}(X) = \delta_2/3 + 2\mu_{\mathcal{N}}/3$, where $\mu_{\mathcal{N}}$ is the standard normal distribution $\mathcal{N}(0,1)$.

Solution (a) For $\mathbb{E}\Big[X\Big]$ we can write

$$\mathbb{E}\Big[X\Big] = 1/2\int_{[0,1]} x\delta_1(dx) + 1/2\int_{[0,1]} x\lambda(dx) = 1/2\int_{[0,1]} x\delta(x-1)\lambda(dx) + 1/2\int_{[0,1]} x\lambda(dx) = 1/2+1+4 = 3/4.$$

Similarly, for $\mathbb{E}\Big[X^2\Big]$ we can write

$$\mathbb{E}\Big[X^2\Big] = 1/2\int_{[0,1]} x^2\delta_1(dx) + 1/2\int_{[0,1]} x^2\lambda(dx) = 1/2\int_{[0,1]} x^2\delta(x-1)\lambda(dx) + 1/2\int_{[0,1]} x^2\lambda(dx) = 1/2+1/6 = 2/3.$$

And using the formula for $\mathrm{Var}(X)$ we can compute

$$\mathrm{Var}(X) = \mathbb{E}\Big[X^2\Big] - \mathbb{E}\Big[X\Big]^2 = 2/3 - 9/16 = 5/48.$$

(b) Similar to part (a), for $\mathbb{E}\big[X\big]$ we have

$$\mathbb{E}\big[X\big] = 1/3\int_{[0,1]} x\delta_2(dx) + 2/3\int_{\mathbb{R}} x\mu_{\mathcal{N}}(dx) = 2/3 + 2/3\underbrace{\int_{\mathbb{R}} xf(x)\lambda(dx)}_{0} = 2/3,$$

where $f(x) = 1/2\sqrt{\pi}e^{-x^2/2}$ is the probability density function for the normal distribution. Also note that we have used the fact that mean of the normal distribution is zero. Similarly, for $\mathbb{E}\big[X^2\big]$ we can write

$$\mathbb{E}\big[X^2\big] = 1/3\int_{[0,1]} x^2\delta_2(dx) + 2/3\int_{\mathbb{R}} x^2\mu_{\mathcal{N}}(dx) = 4/3 + 2/3\underbrace{\int_{\mathbb{R}} x^2 f(x)\lambda(dx)}_{1} = 4/3 + 2/3 = 2,$$

where we have used the fact that for a random variable Z with the standard normal distribution $\mathrm{Var}(Z) = \mathbb{E}\big[Z^2\big] = 1$. And finally

$$\mathrm{Var}(X) = \mathbb{E}\big[X^2\big] - \mathbb{E}\big[X\big]^2 = 2 - 4/9 = 14/9.$$

■ **Problem 1.52** Let X and X be independent, with $X \sim \mathcal{N}(0,1)$, and with $\mathbb{P}(Z = 1) = \mathbb{P}(Z = -1) = 1/2$. Let $Y = XZ$.

(a) Prove that $Y \sim \mathcal{N}(0,1)$.

(b) Prove that $\mathbb{P}(|X| = |Y|) = 1$.

(c) Prove that X and Y are not independent.

(d) Prove that $\mathrm{Corr}(X,Y) = 0$.

(e) It is sometimes claimed that if X and Y are normally distributed random variables with $\mathrm{Corr}(X,Y) = 0$, then X, Y are independent. Is this correct?

Solution (a) Let $B \in \mathcal{B}$. Let μ_Y be the distribution of Y. Then

$$\mu_Y(B) = \mathbb{P}(Y \in B) = \mathbb{P}(X \in B|Z = 1)\mathbb{P}(Z = 1) + \mathbb{P}(-X \in B|Z = -1)\mathbb{P}(Z = -1) = \mu_X/2 + \mu_X/2 = \mu_X = \mu_{\mathcal{N}},$$

where we have used the fact that for the random variable $X \sim \mathcal{N}(0,1)$ we have $\mu_X = \mu_{-X}$. To see why the random variables X and $-X$ have the same law, observe that

$$F_{-X}(x) = \mathbb{P}(-X \le x) = \mathbb{P}(X > -x) = 1 - F_X(-x).$$

On the other hand since

$$F_X(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-x^2/2}dx$$

and using the fact

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-t} e^{-x^2/2}dx = \frac{1}{\sqrt{2\pi}}\int_{t}^{\infty} e^{-x^2/2}dx,$$

we get $F_{-X}(t) = 1 - F_X(t)$, which combining with the expression we found for $F_{-X}(x)$ above we can see that

$$F_{-X}(x) = F_X(x).$$

By proposition 6.0.2 Rosenthal, we can conclude that $\mu_X = \mu_{-X}$.

(a') There is also another easier solution to this. Let $B \in \mathcal{B}$. Then

$$\mu_Y = \mathbb{P}(Y \in B) = \mathbb{P}(XZ \in B) = \frac{1}{2}\mathbb{P}(XZ \in B|Z=1) + \frac{1}{2}\mathbb{P}(XZ \in B|Z=-1)$$
$$= \frac{1}{2}\mathbb{P}(X \in B) + \frac{1}{2}\mathbb{P}(X \in \bar{B}),$$

where $\bar{B}$ is the conjugate set to B where $\bar{B} = \{-x : x \in B\}$. Since X has standard normal distribution, $\mathbb{P}(X \in B) = \mathbb{P}(X \in \bar{B})$. This comes from the fact that the density function for $\mathcal{N}$ is an even function. So

$$\mathbb{P}(Y \in B) = \mathbb{P}(X \in B).$$

(b) We can write

$$\{|X| = |Y|\} = \{|X| = |Y|\} \cap \{Z = 1\} \dot{\cup} \{|X| = |Y|\} \cap \{Z = -1\}$$

Thus

$$\mathbb{P}(|X| = |Y|) = 1/2\mathbb{P}(|X| = |X|) + 1/2\mathbb{P}(|X| = |-X|) = \mathbb{P}(|X| = |X|) = 1.$$

(c) Let $B \in \mathcal{B}$. Then since $\mu_X = \mu_Y$ we have

$$\mathbb{P}(X \in B) = \mathbb{P}(Y \in B).$$

On the other hand,

$$\{X \in B\} \cap \{Y \in B\} = \{X \in B\} \cap \{XZ \in B\} = \{X \in B\} \cap \{Z = 1\}.$$

Since X, Z are independent

$$\mathbb{P}(\{X \in B\} \cap \{Z = 1\}) = 1/2\mathbb{P}(\{X \in B\}).$$

So we can see that $\mathbb{P}(\{X \in B\} \cap \{Y \in B\})$ is not equal to $\mathbb{P}(\{X \in B\})\mathbb{P}(\{Y \in B\})$ in general, unless $\mathbb{P}(B) = 0$. Thus X, Y are not independent.

(d) By the formula for Corr we have

$$\operatorname{Corr}(X,Y) = \mathbb{E}\Big[XY\Big] - \mathbb{E}\Big[X\Big]\mathbb{E}\Big[Y\Big] = \mathbb{E}\Big[X^2Z\Big] - \mathbb{E}\Big[X\Big]\mathbb{E}\Big[XZ\Big]$$
$$= \mathbb{E}\Big[X^2\Big]\mathbb{E}\Big[Z\Big] - \mathbb{E}\Big[X\Big]^2\mathbb{E}\Big[Z\Big] = \mathbb{E}\Big[Z\Big]\operatorname{Var}(X) = 0.$$

where we have used the fact that X, Z are independent, as well as X^2, Z. Thus $\mathbb{E}\Big[XZ\Big] = \mathbb{E}\Big[X\Big]\mathbb{E}\Big[Z\Big]$, and $\mathbb{E}\Big[X^2Z\Big] = \mathbb{E}\Big[X^2\Big]\mathbb{E}\Big[Z\Big]$.

(e) No it is not. All the calculations above demonstrates a counterexample for this claim.

Observation 1.5.2 The conclusion in the part (e) of the question above is very important. In this example we can see two random variables X, Y which both of them has the same distribution, $\mu_X = \mu_Y$ and their absolute value agrees almost everywhere $|X| = |Y|$ *a.e.*, and their correlation is zero $\operatorname{Corr}(X,Y) = 0$. However, we can see that they are *not* independent.

■ **Problem 1.53** Let X be a random variable, and let $F_X(x)$ be its cumulative distribution function. For fixed $x \in \mathbb{R}$, we know by right-continuoity that $\lim_{y \searrow x} F_X(y) = F_X(x)$.

(a) Give a necessary and sufficient condition that $\lim_{y\nearrow x} F_X(y) = F_X(x)$.

(b) More generally, give a formula for $F_X(x) - (\lim_{y\nearrow x} F_X(y))$, in terms of a simple property of X.

Solution (a) A necessary and sufficient condition is $\mu_X(\{x\}) = 0$, where μ_X is the law, or the distribution of the random variable X.

(b) The value of $F_X(x) - (\lim_{y\nearrow x} F_X(y))$ is precisely the value of μ_X evaluated on the singleton $\{x\}$. In other words
$$F_X(x) - (\lim_{y\nearrow x} F_X(y)) = \mu_X(\{x\}).$$

1.6 More Probability Theorems

Observation 1.6.1 — Almost sure convergence and the convergence of the expected values. Let $X, X_1, X_2, \cdots$ be random variables, and assume that $X_n \to X$ almost surely. Then the real sequence $\mathbb{E}\Big[X_n\Big]$ *does not* necessarily converge to $\mathbb{E}\Big[X\Big]$ as $n \to \infty$. One example is the following: Let $X_n = \mathbb{1}_{[0,1/n]}$. Then $X_n \to 0$ on pointwise on $(0,1]$. So we can say that $X_n \to X$ almost surely on $[0,1]$. However, a simple calculation reveals that $\mathbb{E}\Big[X_n\Big] = 1$ for all n, however $\mathbb{E}\Big[X\Big] = 0$.

1.7 Solved Problems

■ **Problem 1.54** For the "simple counter-example" with $\Omega = \mathbb{N}$ and $\mathbb{P}(\omega) = 2^{-\omega}$ for $\omega \in \Omega$ and $X_n(\omega) = 2^n\delta_{n,\omega}$, verify explicitly that the hypotheses of each of the monotone convergence theorem, the bounded convergence theorem, the dominated convergence theorem, and the uniform integrablity convergence theorem, are all violated.

Solution (i) The *Monotone Convergence Theorem* can not be used since the sequence $\{X_n\}$ is not monotone. It is not increasing as
$$0 = X_{n+1}(n) \le X_n(n) = 1,$$
and it is not decreasing as
$$1 = X_{n+1}(n+1) \ge X_n(n+1) = 0.$$

(ii) The *Bounded Convergence Theorem* can not be used since the sequence is not uniformally bounded. I.e. $\forall K \in \mathbb{R}$ we can find n large enough so that $|X_n| > K$.

(iii) The *Dominated Convergence Theorem* can not be used since there does not exists any random variable Y such that $|X_n| \le Y$ with $\mathbb{E}\Big[Y\Big] < \infty$. That is because to meet the first condition we need to have $Y(k) \ge 2^k$ for $k \in \Omega$, which then will imply
$$\mathbb{E}\Big[Y\Big] \ge \sum_{n=1}^{\infty} 1 = \infty.$$

(iv) The *Uniform Integrablity Convergence Theorem* can not be used since collection $\{X_n\}$ are not uniformally integrable. That is because for all $\alpha \in \mathbb{R}$, there exists $N \in \mathbb{N}$ such that $\forall n > N$ we have $\mathbb{E}\Big[|X|\mathbb{1}_{|X_n|\ge\alpha}\Big] = 1$.

■ **Problem 1.55** Give an example of a sequence of random variables which is unbounded but still uniformally integrable. For bonus points, make the sequence also be undominated, i.e. violate the hypothesis of the dominated convergence theorem.

Solution Let $\Omega = \mathbb{N}$, with $\mathbb{P}(\omega) = 2^{-\omega}$. Define $X_n = \frac{2^n}{n}\delta_{n,\omega}$. To see that the sequence is uniformally integrable first observe that $\mathbb{E}\Big[|X_n|\Big] = 1/n$, and since $|X_n|\mathbb{1}_{|X_n|\geq\alpha} \leq |X|$ for any $\alpha \in \mathbb{R}$, we have

$$\mathbb{E}\Big[|X_n|\mathbb{1}_{|X_n|\geq\alpha}\Big] \leq \mathbb{E}\Big[|X|\Big] = 1/n.$$

Consider the α of the form $\alpha = 2^k/k$ for $k \in \mathbb{N}$. Then for any fixed $k \in \mathbb{N}$ we have $|X_n| \geq \alpha(k)$ for all $n \geq k$. Thus

$$\mathbb{E}\Big[|X_n|\mathbb{1}_{|X_n|\geq\alpha(k)}\Big] \leq 1/k,$$

that goes to zero as $k \to \infty$ (i.e. $\alpha \to \infty$). However, the sequence is undominated. That is because the dominating function Y should assume the value at least $2^n/n$ for $n \in \Omega$ which will lead to

$$\mathbb{E}\Big[Y\Big] = \sum_{n=1}^{\infty}\frac{1}{n} = \infty.$$

■ **Problem 1.56** Let $X, X_1, X_2, \cdots$ be non-negative random variables, defined jointly on some probability triple triple $(\Omega, \mathcal{F}, \mathbb{P})$, each having finite expected value. Assume that $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ for all $\omega \in \Omega$. For $n, K \in \mathbb{N}$, let $Y_{n,k} = \min(X_n, K)$. For each of the following statements either prove it must true, or provide a counter example to show it is sometimes false.

(a) $\lim_{K\to\infty}\lim_{n\to\infty}\mathbb{E}\Big[Y_{n,k}\Big] = \mathbb{E}\Big[X\Big]$.

(b) $\lim_{n\to\infty}\lim_{K\to\infty}\mathbb{E}\Big[Y_{n,k}\Big] = \mathbb{E}\Big[X\Big]$.

Solution (a) This statement is true. To prove this observe that for a fixed K we have $\{Y_{n,k}\} \nearrow \min(X, K)$ as $n \to \infty$. Thus from monotone convergence theorem we have

$$\lim_{n\to\infty}\mathbb{E}\Big[Y_{n,k}\Big] = \mathbb{E}\Big[\min(X, K)\Big].$$

On the other hand $\{\min(X, K)\} \nearrow X$ as $K \to \infty$. So again by monotone convergence theorem we have

$$\lim_{K\to\infty}\lim_{n\to\infty}\mathbb{E}\Big[Y_{n,k}\Big] = \lim_{K\to\infty}\mathbb{E}\Big[\min(X, K)\Big] = \mathbb{E}\Big[X\Big].$$

(b) This statement is not true. Let $\Omega = \mathbb{N}$, $\mathbb{P}(\omega) = 2^{-\omega}$ and define $X_n = 2^n$. We know that $X_n(\omega) \to X(\omega)$ for all $\omega \in \Omega$ as $n \to \infty$. Observe that for a fixed n we have $\{Y_{n,k}\} \nearrow X_n$ as $k \to \infty$. So from the monotone convergence theorem

$$\lim_{K\to\infty}\mathbb{E}\Big[\min(X_n, K)\Big] = \mathbb{E}\Big[X_n\Big].$$

Thus

$$\lim_{n\to\infty}\lim_{K\to\infty}\mathbb{E}\Big[\min(X_n, K)\Big] = \lim_{n\to\infty}\mathbb{E}\Big[X_n\Big] = 1 \neq \mathbb{E}\Big[X\Big] = 0.$$

Observation 1.7.1 In the question above, we worked with the random variable $\min(X_n, K)$. This random variable is very interesting. Because if $X_n \to X$ almost surely, then

$$\{\min(X_n, k)\} \nearrow X_n \quad as\ k \to \infty,$$

and also

$$\{\min(X_n, k)\} \nearrow \min(X, k) \quad as\ n \to \infty.$$

The whole idea of the problem above is precisely around the behaviour of $\min(X_n, K)$ as shown above.

■ **Problem 1.57** Suppose that $\lim_{n\to\infty} X_n(\omega) = 0$ for all $\omega \in \Omega$, but $\lim_{n\to\infty} \mathbb{E}\Big[X_n\Big] \neq 0$. Prove that $\mathbb{E}\Big[\sup_n |X_n|\Big] = \infty$.

Solution Assume otherwise, i.e. $\mathbb{E}\Big[\sup_n |X_n|\Big] < \infty$. Let $Y = \sup_n |X_n|$. The Y is a possible choice for the dominant r.v. in the theorem 9.1.2. Thus by the dominated convergence theorem we will have

$$\lim_{n\to\infty} \mathbb{E}\Big[X_n\Big] = \mathbb{E}\Big[X\Big] = 0.$$

This contradicts the assumption. So the we should have $\mathbb{E}\Big[\sup_n |X_n|\Big] = \infty$.

■ **Problem 1.58** Prove that Theorem 9.1.2 implies Theorem 4.2.2, assuming $\mathbb{E}\Big[|X|\Big] < \infty$. In other words, prove that the Dominated Convergence Theorem, implies the Monotone Convergence Theorem if $\mathbb{E}\Big[|X|\Big] < \infty$.

Solution Let $X, X_1, X_2, \cdots$ be non-negative random variables such that $\{X_n\} \nearrow X$ as $n \to \infty$. By hypothesis we have $\mathbb{E}\Big[|X|\Big] < \infty$. So $\{X_n\}$ is dominated by X, and using dominated convergence theorem we conclude that $\mathbb{E}\Big[X_n\Big] \to \mathbb{E}\Big[X\Big]$ as $n \to \infty$.

■ **Problem 1.59** Let $X_1, X_2, \cdots$ be i.i.d., each with $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = 1/2$.

(a) Compute the moment generating function $M_{X_i}(s)$.

(b) Use Theorem 9.3.4 to obtains an exponentially-decreasing upper bound on $\mathbb{P}(\frac{1}{n}(X_1 + \cdots + X_n) \geq 0.1)$.

Solution (a) From the formula for the moment generating function we have

$$M_{X_i}(s) = \mathbb{E}\Big[e^{sX_i}\Big] = 1/2e^s + 1/2e^{-s} = \cosh(s).$$

(b) Observe that $\mathbb{E}\Big[X_i\Big] = M'_{X_i}(0) = 0$. Let $\epsilon = 0.1$. Then

$$\rho = \inf_s (e^{-s\epsilon} \cosh(s)) \approx 0.995.$$

■ **Problem 1.60** Let $X_1, X_2, \cdots$ be i.i.d., each having the standard normal distribution $\mathcal{N}(0, 1)$. Use Theorem 9.3.4 to obtain an exponentially decreasing upper bound on $\mathbb{P}(1/n(X_1 + \cdots + X_n) \geq 0.1)$. *Hint: You can use the fact that for $X \sim \mathcal{N}(0, 1)$ we have*

$$M_X(s) = e^{s^2/2}.$$

Solution Using the formula for ρ we can write

$$\rho = \inf_{s>0} (e^{-s(m+\epsilon)M_{X_i}(s)}).$$

Observe that $m = 0$, and also $M_{X_i}(s) = e^{s^2/2}$. Thus

$$\rho = \inf_{s>0} (e^{s^2/2 - 0.1s}).$$

Doing a numerical computation will reveal that

$$\rho \approx 0.995012.$$

■ **Problem 1.61** Let $X \sim \text{Exp}(5)$ with density $f_X(x) = 5e^{-5x}$ for $x \geq 0$ and $f_X(x) = 0$ for $x < 0$.

(a) Compute the moment generating function $M_X(s)$.

(b) Use $M_X(s)$ to compute (with explanation) the expected value $\mathbb{E}\big[X\big]$.

Solution (a) Using the formula for the moment generating function we can write

$$\begin{aligned} M_X(s) = \mathbb{E}\Big[E^{sX}\Big] &= \int_{-\infty}^{+\infty} e^s x f_X(x) dx \\ &= \lambda \int_0^{+\infty} e^s x e^{\lambda x} dx \\ &= \frac{\lambda}{s-\lambda} e^{x(s-\lambda)}\Big|_0^{\infty} \\ &= \frac{\lambda}{\lambda - s}, \end{aligned}$$

where we have assumed that $s - \lambda < 0$.

(b) We know that $\mathbb{E}\big[X\big] = M_X'(0)$. So we first calculate M_X'

$$M_X'(s) = \frac{\lambda}{(\lambda - s)^2}.$$

Then it is straightforward to see that $\mathbb{E}\big[X\big] = 1/\lambda$.

■ **Problem 1.62** Let $X \sim \text{Poisson}(a)$ and $Y \sim \text{Poisson}(b)$ be independent. Let $Z = X + Y$. Use the convolution formula to compute $\mathbb{P}(Z = z)$ for all $z \in \mathbb{R}$, and prove that $Z \sim \text{Poisson}(a + b)$.

Solution Recall that for a random variable X with Poisson distribution Poisson(a) we have

$$\mathbb{P}(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}.$$

On the other hand we know that $p_Z = p_X * p_Y$. So

$$\begin{aligned} p_Z(k) &= \sum_n p_X(n) p_Y(k-n) = \sum_n \frac{e^{-a}a^n}{n!} \cdot \frac{e^{-b}b^{k-n}}{(k-n)!} \\ &= \sum_{n=1}^{\infty} e^{-(a+b)} \frac{a^n b^{k-n}}{(k-n)!} \\ &= e^{-(a+b)}/k! \frac{k!}{n!(k-n)!} a^n b^{n-k} \\ &= \frac{e^{-(a+b)}(a+b)^k}{k!}. \end{aligned}$$

Thus $Z \sim \text{Poisson}(a+b)$.

1.8 Weak Convergence

1.8.1 Solved Problems

■ **Problem 1.63** Suppose $\mathcal{L}(X_n) \Rightarrow \delta_c$ for some $c \in \mathbb{R}$. Prove that $\{X_n\}$ convergence to c in probability.

Solution Let $\epsilon > 0$ give. Then

$$\begin{aligned}\mathbb{P}(|X_n - c| \geq \epsilon) &= 1 - \mathbb{P}(|X_n - c| < \epsilon)\\ &= 1 - \mathbb{P}(X_n \in (c-\epsilon, c+\epsilon))\\ &= 1 - \mu_n((c-\epsilon, c+\epsilon)) \to 1 - 1 = 0.\end{aligned}$$

Note that we have used the fact that $\mu_n(A) \to \mu(A)$ for $A \in \mathcal{B}$ such that $\mu(\partial A) = 0$, and the fact that $\mu((c-\epsilon, c+\epsilon)) = 1$. So we conclude that $X_n \to X$ in probability.

■ **Problem 1.64** Let $X, Y_1, Y_2, \cdots$ be independent random variables, with $\mathbb{P}(Y_n = 1) = 1/n$ and $\mathbb{P}(Y_n = 0) = 1 - 1/n$. Let $Z_n = X + Y_n$. Prove that $\mathcal{L}(Z_n) \Rightarrow \mathcal{L}(X)$, i.e. that the law of Z_n converges weakly to the law of X.

Solution Let $\epsilon > 0$ given. Then

$$\mathbb{P}(|Z_n - X| \geq \epsilon) = \mathbb{P}(|Y_n| \geq \epsilon) \to 0 \quad as \quad n \to \infty.$$

So we observe that Z_n converges to X in probability, which implies convergence in distribution.

■ **Problem 1.65** Let $\mu_n = \mathcal{N}(0, 1/n)$ be a normal distribution with mean 0 and variance $1/n$. Doe the sequence $\{\mu_n\}$ converge weakly to some probability measure? If yes, to what measure?

Solution Yes. $\mu_n \Rightarrow \delta_0$ as $n \to \infty$. To see this let f denote the density the $\mathcal{N}(0,1)$ distribution. So

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Similarly, let f_n denote the density of the $\mathcal{N}(0, 1/n)$ distribution. So

$$f_n(x) = \frac{n}{\sqrt{2\pi}} e^{-x^2 n^2/2} = n f(nx).$$

We want to show that F_n (cumulative distribution for $\mathcal{N}(0, 1/n)$) converges to H that is the cumulative distribution of the δ_0 distribution on $\mathbb{R}$ except the origin that is the point of discontinuity of H. Let $t \neq 0$. Then

$$F_n(t) = \int_0^t f_n(t)dt = \int_0^t nf(nt)dt = \int_0^{nt} f(y)dy = F_{\mathcal{N}}(nt),$$

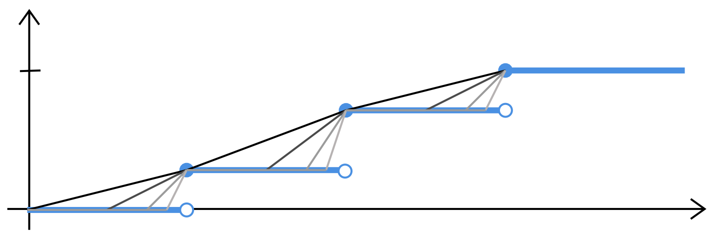
where $F_{\mathcal{N}}$ is the cumulative distribution for $\mathcal{N}(0,1)$. Observe that when $t > 0$ we have $F_n(t) \to 1$ as $n \to \infty$ and when $t < 0$ we have $F_n(t) \to 0$ as $n \to -\infty$. So $F_n(t)$ converges to H when $t \neq 0$. So $\mu_n \Rightarrow \delta_0$ as $n \to \infty$.

■ **Problem 1.66** Let $a_1, a_2, \cdots$ be any sequence of non-negative real numbers with $\sum_i a_i = 1$. Define the discrete measure μ by $\mu(\cdot) = \sum_{i \in \mathbb{N}} a_i \delta_i(\cdot)$, where $\delta_i(\cdot)$ is a point-mass at the positive integer i. Construct a sequence $\{\mu_n\}$ of probability measure, each having a density with respect to Lebesgue measure, such that $\mu_n \Rightarrow \mu$ as $n \to \infty$.

Solution Define μ_n as

$$\mu_n(A) = \sum_{i=1}^{\infty} n a_i \lambda(A \cap (i - 1/n, i])$$

for any $A \in \mathcal{B}$. With this definition of μ_n we are in fact approximating a discontinuous function (F_n) with a piece-wise linear function as shown below. It is easy to show that F_n converges point-wise to F everywhere except at the point of discontinuity.



■ **Remark** In the question above, if there are no limitations that the sequence $\{\mu_n\}$ should ave a density with respect to Lebesgue measure, then one natural choice is

$$\mu_n = \sum_{i}^{n} a_i \delta_i + (\sum_{i=n+1}^{\infty} a_i)\delta_{n+1}.$$

■ **Problem 1.67** Let $\mathcal{L}(Y) = \mu$, where μ has continuous density f. For $n \in \mathbb{N}$, let $Y_n = \lceil nY \rceil / n$, and let $\mu_n = \mathcal{L}(Y_n)$.

(a) Describe μ_n explicitly.

(b) Prove that $\mu_n \Rightarrow \mu$.

(c) Is μ_n discrete, or absolutely continuous, or neither? What about μ?

Solution (a) We start by calculating the cumulative distribution function for Y_n.

$$\begin{aligned}
F_n(y) &= \mu_n((-\infty, x]) = \mathbb{P}(Y_n \in (-\infty, y]) \\
&= \mathbb{P}(\frac{\lfloor nY \rfloor}{n} \in (-\infty, y]) \\
&= \mathbb{P}(\lfloor nY \rfloor \in (-\infty, ny]) \\
&= \mathbb{P}(nY \in (-\infty, \lfloor ny \rfloor + 1)) \\
&= \mathbb{P}(Y \in (-\infty, \frac{\lfloor ny \rfloor}{n} + 1/n)) \\
&= \mathbb{P}(Y \in (-\infty, \frac{\lfloor nY \rfloor}{n}]) + \mu((\frac{\lfloor ny \rfloor}{n}, \frac{\lfloor ny \rfloor + 1}{n})) \\
&= F(\frac{\lfloor ny \rfloor}{n}) + \mu((\frac{\lfloor ny \rfloor}{n}, \frac{\lfloor ny \rfloor + 1}{n}))
\end{aligned}$$

(b) First observe that

$$\frac{\lfloor ny \rfloor}{n} \to y \quad as \quad n \to \infty.$$

That is because $|\frac{\lfloor ny \rfloor}{n} - y| = \frac{|\lfloor ny \rfloor - ny|}{n} \leq \frac{1}{n}$. On the other hand

$$\mu((\frac{\lfloor ny \rfloor}{n}, \frac{\lfloor ny \rfloor + 1}{n})) \to \mu((y, y)) = \mu(\varnothing) = 0 \quad as \quad n \to \infty.$$

So from part (a) we can see that

$$F_n(y) \to F(y) \quad as \quad n \to \infty$$

for all $y \in \mathbb{R}$. Thus $\mu_n \Rightarrow \mu$ as $n \to \infty$.

(c) TODO: TO BE ADDED.

■ **Problem 1.68** Let $0 < M < \infty$, and let $f, f_1, f_2, \cdots : [0,1] \to [0, M]$ be Borel-measurable functions with $\int_0^1 f\, d\lambda = \int_0^1 f_n\, d\lambda = 1$. Suppose $\lim_n f_n(x) = f(x)$ for each fixed $x \in [0,1]$. Define probability measures $\mu, \mu_1, \mu_2, \cdots$ by $\mu(A) = \int_A f\, d\lambda$, and $\mu_n(A) = \int_A f_n\, d\lambda$, for Borel $A \subset [0,1]$. Prove that $\mu_n \Rightarrow \mu$.

Solution Let $g : [0,1] \to \mathbb{R}$ be any bounded continuous function. Then by change of variable formula we have

$$\int g d\mu_n = \int_0^1 g f_n d\lambda, \qquad \int g d\mu = \int_0^1 g f d\lambda.$$

Since $f_n \to f$ pointwise, then $gf_n \to gf$ pointwise as well. Observe that $gf_n \leq gM$ and since g is bounded and continuous then $\int gMd\lambda < \infty$. Thus by monotone convergence theorem we have

$$\int g f_n d\lambda \to \int g f d\lambda \quad \text{as } n \to \infty.$$

In other words $\int g d\mu_n \to \int g d\mu$ as $n \to \infty$. This proves that $\mu_n \Rightarrow \mu$.

■ **Problem 1.69** Let $f : [0,1] \to (0, \infty)$ be a continuous function such that $\int_0^1 f d\lambda = 1$ (where λ is Lebesgue measure on $[0,1]$). Define probability measure μ and $\{\mu_n\}$ by $\mu(A) = \int_0^1 f \mathbb{1}_A d\lambda$ and $\mu_n(A) = \sum_{i=1}^n f(i/n) \mathbb{1}_A(i/n) / \sum_{i=1}^n f(i/n)$.

(a) Prove that $\mu_n \Rightarrow \mu$ as $n \to \infty$.

(b) Explicitly construct random variables Y and $\{Y_n\}$ so that $\mathcal{L}(Y) = \mu$, $\mathcal{L}(Y_n) = \mu_n$, and $Y_n \to Y$ with probability 1.

Solution (a) Let $A \in \mathcal{B}$ such that $\mu(\partial A) = 0$. Then

$$\begin{aligned}
\lim_n \mu_n(A) &= \lim_n \frac{\sum_{i=1}^n f(i/n) \mathbb{1}_A(i/n)}{\sum_{i=1}^n f(i/n)} \\
&= \lim_n \frac{\sum_{i=1}^n f(i/n) \mathbb{1}_A(i/n) 1/n}{\sum_{i=1}^n f(i/n) 1/n} \\
&= \frac{\lim_n \sum_{i=1}^n f(i/n) \mathbb{1}_A(i/n) 1/n}{\lim_n \sum_{i=1}^n f(i/n) 1/n} \\
&= \frac{\int_0^1 f \mathbb{1} d\lambda}{\int_0^1 f d\lambda} \\
&= \int_0^1 f \mathbb{1}_A d\lambda = \mu(A).
\end{aligned}$$

Since A was arbitrary then $\mu_n \Rightarrow \mu$ as $n \to \infty$.

(b) From the definition of μ and μ_n we define F_n and F (the cumulative distribution functions) and then define the random variables $Y, Y_1, Y_2, \cdots$ as

$$Y_n(\omega) = \sup_x \{x : F_n(x) < \omega\}, \qquad Y(\omega) = \sup_x \{x : F(x) < \omega\}.$$

■ **Remark** In Rosenthal, Theorem 7.2.1 defines the "inverse" of F as

$$Y(\omega) = \inf_x \{x : F(x) \geq \omega\}.$$

It is not clear to me if these two definitions are the same or not. Because one of them leads to a left continuous function while the other one leads to a right continuous function.

1.9 Canonical Examples

■ **Example 1.1 — Converging in probability but not almost surely.** Let $X, X_1, X_2, \cdots$ be random variables with $X \equiv 0$, and $X_n = \mathbb{1}_{B_n}$ for some $B_n \in \mathcal{B}$ such that $\mathbb{P}(B) = 1/n$. In other words $\mathbb{P}(X_n = 1) = 1/n$ and $\mathbb{P}(X_n = 0) = 1 - 1/n$. Then

(a) $X_n \to X$ in probability.

(b) $X_n \not\to X$ almost surly. (Because $1/n$ is not summable).

■

■ **Remark** (a) Defining B_n such that $\mathbb{P}(B_n) = 1/n^2$ then we will also have the almost sure convergence.

(b) Defining $X_n = n\mathbb{1}_{B_n}$ with $\mathbb{P}(B_n) = 1/n$, then $X_n \to X$ in probability, and $1 = \mathbb{E}\left[X_n\right] \not\to \mathbb{E}\left[X\right] = 0$.

(c) Defining $X_n = n^2\mathbb{1}_{B_n}$ with $\mathbb{P}(B_n) = 1/n^2$, then $X_n \to X$ almost surely, but $1 = \mathbb{E}\left[X_n\right] \not\to \mathbb{E}\left[X\right] = 0$.

■ **Example 1.2 — Convergence in distribution but not in probability.** Let $X, X_1, X_2, \cdots$ be i.i.d. random variables each equal to ± 1 with probability $1/2$. Then obviously $\mu_n \Rightarrow \mu$ however X_n does not converge to X in probability. For instance, let $\epsilon = 2$. Then $\mathbb{P}(|X_n - X| \geq 2) = 1/2 \not\to 0$. ■

■ **Example 1.3** We give an example of a sequence of random variables which is unbounded, uniformly integrable, and undominated (in the sense of the hypothesis of dominated convergence theorem). Define

$$\Omega = \mathbb{N}, \quad \mathbb{P}(\Omega) = 2^{-\omega}, \quad X_n(\omega) = \frac{2^n}{n}\delta_{\omega,n}$$

■

Observation 1.9.1 Let $x \in \mathbb{R}$. Then $1 - x < e^{-x}$. This is a very important inequality that can be justified by simply looking at the graph of $(1-x)$ and e^{-x}. This inequality is used in the proof of the Borel-Cantelli lemma, as well as in Problem 1.70.

Observation 1.9.2 Recall the definition of lim sup and lim inf of a sequence of events $A_1, A_2, \cdots$

$$\limsup_n A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k, \qquad \liminf_n A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_n.$$

The important observation is that

$$\{\bigcap_{k=n}^{\infty} A_n\} \nearrow \liminf_n A_n, \qquad \{\bigcup_{k=n}^{\infty} A_n\} \searrow \limsup_n A_n.$$

1.10 Interesting Questions and Remarks From Other Books

This section will contain the interesting questions and remarks collected from other books. The content will be cited.

1.10.1 Normal Numbers, Coin Tossing, WLLN, and SLLN

This subsection contains a brief summary of the interesting story in Billingsley section 1.1. This story is interesting because it contains most of the main ideas under one umbrella demonstrates most of the interesting concepts at once.

The story begins by the notion of length for half open intervals as a subset of $[0,1]$. Let $(a,b]$ be such an interval. The notion of length for this interval is

$$\ell((a,b]) = b - a.$$

Now let Ω be the space of all infinitely long binary strings. For instance $\omega = 01110101 \in \Omega$. Each point in Ω can be thought of as an experiment in which we toss a coin infinitely many times. For instance ω given above corresponds to the experiment where on the first toss we got a Tails, on the second toss we got a heads, and etc. Consider the set $A_0 \subset \Omega$ given by

$$A_0 = \{\omega : d_1(\omega) = 0\}.$$

Similarly we can define

$$A_{10} = \{\omega : d_1(\omega) = 1, d_2(\omega) = 0\},$$

and etc. We associate each $\omega \in \Omega$ with a real number as follows

$$\Omega \ni u_1 u_2 u_3 \cdots \quad \longleftrightarrow \quad 0.u_1 u_2 u_3 \cdots \in [0,1]. \tag{♩}$$

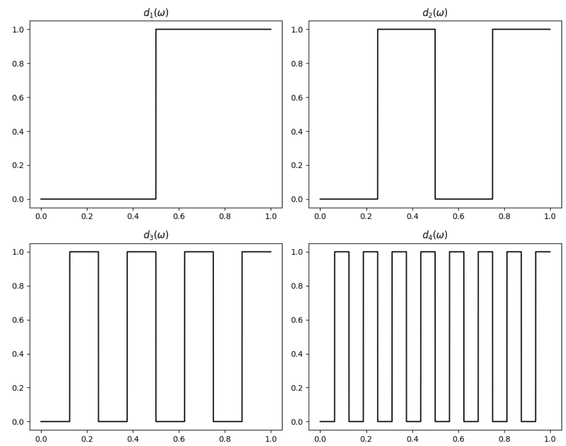
With this association it is easy to check that for instance we have

$$A_{10} \longleftrightarrow (\frac{1}{2}, \frac{3}{4}], \; A_{101} \longleftrightarrow (\frac{5}{8}, \frac{6}{8}], \quad \text{and etc.}$$

And in general

$$A_{u_1 \cdots u_n} \longleftrightarrow (0.u_1 \cdots u_n, 0.u_1 \cdots u_n + \frac{1}{2^n}].$$

With an slightly different point of view, we can get a very clear understanding of the weak law of large numbers, as well as the strong law of large numbers. Let $d_n : [0,1] \to \{0,1\}$ be the "i^{th} component function in base 2". For instance $d_2(0.01101\cdots) = 1$ as the second position in the binary expansion is 1. The graph of d_n is depicted in the following figure for few different values of n.



Now assume that in our coin tossing experiment, we are interested in the total number of Heads. For instance in $\omega = 1010100\dots$, where the sequence of 0s continue in the tail, the number of Heads is 3. This can also be calculated by the function $H_n : [0,1] \to \mathbb{N}$ given as

$$H_n = \sum_{i=1}^{n} d_i,$$

that gives the number of heads in the first n trials. So for the example above, the number of heads are also given as $H_8(0.1010100\dots) = d_1(0.1010100\dots) + d_2(0.1010100\dots) + d_3(0.1010100\dots) + \cdots + 0 = 1 + 0 + 1 + 0 + 1 + 0 + 0 + \cdots + 0 = 3$. So the 1-1 correspondence between Ω and $[0,1]$ in (♩) is a very useful correspondence. That is because most of the events of interests on Ω can be converted back to some intervals on $[0,1]$ for which we can assign probability. For another example, we want to know the behaviour of the fraction of heads as the number of experiments goes to infinity. For $\omega \in \Omega$, the fraction of heads in the n first trials is

$$F_n(\omega) = \frac{1}{n}\sum_{i=1}^{n} d_i(\omega) = \frac{H_n(\omega)}{n}.$$

Consider the following example.

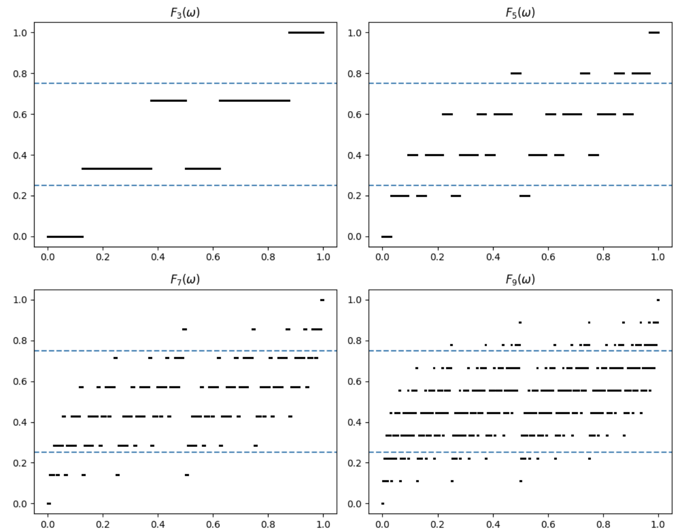
■ **Example 1.4** In an infinite coin tossing experiment what is the probability that the sum of first three tosses is 2 (i.e. 2 heads appear in the first 3 trials?) ■

Solution Consider the pre-image 2 under F_3. I.e.

$$F_3^{-1}(2) = (\frac{3}{8}, \frac{4}{8}] \cup (\frac{5}{8}, \frac{6}{8}] \cup (\frac{6}{8}, \frac{7}{8}].$$

The length of $F_3^{-1}(2)$ is $3/8$. So we can assign the probability $3/8$ to this set.

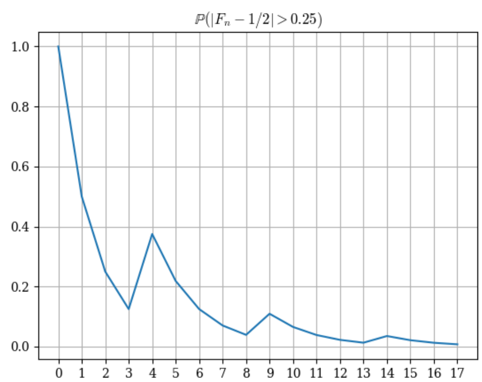
The graph of the set functions F_n is very illustrative as in the following figure



The graph of the functions F_n shows that F_n the "deviation" of the sequence of function from 1/2 gets more and more "negligible". Mathematically, for any $\epsilon > 0$ (in the figure above $\epsilon = 1/4$) we have

$$\mathbb{P}(|F_n - 1/2| \geq \epsilon) \to 0 \quad \text{as } n \to \infty.$$

This is shown in the figure below



This is the Weak Law of Large Numbers! In fact, more is true with our coin tossing experiment. For "almost all" of the points $\omega \in [0, 1]$ we have

$$F_n(\omega) \to 1/2 \quad \text{as } n \to \infty.$$

The meaning of "almost all" is that the set of all points $\eta \in [0, 1]$ such that $F_n(\eta) \not\to 1/2$ is "negligible". This is the Borel's Normal Numbers Theorem.

1.10.2 Solved Problems

■ **Problem 1.70 — From Billingsley.** (a) Show that a discrete probability space (see Example 2.8 for the formal definition) can not contain an infinite sequence $A_1, A_2, \cdots$ of independent events each of probability 1/2. Note that A_n ve be identified with heads on the n-the toss of a coin.

(b) Suppose that $0 \leq p_n \leq 1$. and put $\alpha_n = \min(p_n, 1-p_n)$. Show that if $\sum_n a_n$ diverges, then no discrete probability space can contain independent events $A_1, A_2, \cdots$ such that A_n has probability p_n.

Solution Let $\{x\} \subset \Omega$. Then x belongs to only one of the four disjoint sets below

$$A_1 \cap A_2, \quad A_1^c \cap A_2 \quad A_1 \cap A_2^c, \quad A_1^c \cap A_2^c.$$

Observe that the probability of each of these events are 1/4 (noting that if A, B are independent then A^c, B^c are independent as well). So $\{x\}$ must have probability 1/4 at most. Now consider the events A_1, A_2, A_3. Then x should belong to only one of the disjoint events below

$$\begin{gathered} A_1 \cap A_2 \cap A_3, \quad A_1 \cap A_2^c \cap A_3, \quad A_1 \cap A_2 \cap A_3^c, \quad A_1 \cap A_2^c \cap A_3^c, \\ A_1^c \cap A_2 \cap A_3, \quad A_1^c \cap A_2^c \cap A_3, \quad A_1^c \cap A_2 \cap A_3^c, \quad A_1^c \cap A_2^c \cap A_3^c. \end{gathered}$$

Since A_1, A_2, A_3 are independent events the probability of each of the events above are 1/8. So $\{x\}$ can have probability 1/8 at most. Continuing this we will get $\mathbb{P}(\{x\}) = 0$ for all $x \in \Omega$. On the other hand since Ω is countable then

$$1 = \mathbb{P}(\Omega) = \sum_{x \in \Omega} \mathbb{P}(\{x\}) = 0$$

which is a contradiction.

1. When p_n goes to zero fast enough such that $\sum_n p_n < \infty$, then A_n can be independent events on a discrete probability space. To see this let B_i be A_i or A_i^c for each i. Then

$$\mathbb{P}(B_1 \cap \cdots \cap B_n) \leq \prod_{i=1}^{n} (1 - \alpha_i) \leq e^{-\sum_{i=1}^{n} \alpha_i}.$$

Thus when the sum $\sum_i \alpha_i$ diverges the RHS goes to zero and this leads to conclusion that $\mathbb{P}(\{x\}) = 0$ for all $x \in \Omega$ which leads to a similar kind of contradiction we had in part (a). But when $\sum_i \alpha_i$ is converging, then we don't have a contradiction which leads to the conclusion that such independent events can exist.

■ **Remark** (i) Note that by discrete probability measure we also require it to be countable.

(ii) In part (b) of the question above we use an argument very similar to the one in the proof of Borel-Cantelli. However, it is not yet clear for me if we can somehow directly use Borel-Cantelli Lemma.

■ **Problem 1.71 — From Ross.** Ben can talk a course in compute science or chemistry. If she takes the computer science course, then she will get A grade with probability $\frac{1}{2}$. If she takes the chemistry course, then she will get A grade with probability $\frac{1}{3}$. She decides to base her decision on the flip of a fair coin. What is the probability that she gets an A in chemistry?

Solution We define the following events

A: she will get an A grade.
CO: she will take the computer science course.
CH: she will take the chemistry course.

Then the question is basically asking for $\mathbb{P}(A \cap CH)$. We can compute it by

$$\mathbb{P}(A \cap CH) = \mathbb{P}(A|CH)\mathbb{P}(CH) = \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}.$$

■ **Problem 1.72** And urn contains seven black balls and five white balls. We draw two times from the urn. Given that the each ball has the same probability to be drawn, what is the probability that both balls drawn are black?

Solution This question nicely demonstrates the fact that there are many ways to define the event spaces, and not all of them are very useful in computing the desired probability. Define

E: two drawn balls are black.

The question is in fact asking $\mathbb{P}(E)$. But this even is not very useful in any progress with the solution. Thus we need to define some finer events

E_1: The first drawn ball is black.
E_2: The second drawn ball is black.

It is clear that $E = E_1 \cap E_2$. These two finer events allows us to compute the probability of interest given the data we have in our hand.

$$\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_2|E_1)\mathbb{P}(E_1) = \frac{6}{11} \cdot \frac{7}{12}$$

■ **Problem 1.73 — From Ross.** Three men at a party through their hats into the center of the room, and then, after mixing the hats, each pick a hat randomly. What is the probability if non of them get their own hat back.

Solution There are a million ways to tack a probability problem. We can construct a suitable sample space and then compute the probabilities explicitly, or we can use the properties of the probability function to computer the desired probability without any need to construct the sample space. Here, we will demonstrate two ways.

Solving the problem by utilizing the properties of the probability function. First we need to define some suitable events. There are again many ways to define event sets and each have their own pros and cons. We proceed with the following definition.

E_i: The person i "selects" his own hat.

Also, with this particular construction of the event sets, it is much more easier to compute the complementary probability of the desired probability first and then compute the desired one by simply subtracting it from 1. The complement of the event "no men gets his own hat back" is "at least one man gets his hat back" which is $\mathbb{P}(E_1 \cup E_2 \cup E_3)$. To compute the terms of this we first need to calculate $\mathbb{P}(E_i)$, $\mathbb{P}(E_i \cap E_j)$ where $i \neq j$ and also $\mathbb{P}(E_1 \cap E_2 \cap E_3)$. We know that $\mathbb{P}(E_i) = 1/3$ for $i = 1, 2, 3$. That is because it is equally likely he selects any of the hats at the center. For $\mathbb{P}(E_i \cap E_j)$ we can write

$$\mathbb{P}(E_i \cap E_j) = \mathbb{P}(E_i|E_j)\mathbb{P}(E_j) = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}.$$

In which we used the fact that $\mathbb{P}(E_i|E_j)$ is $\frac{1}{2}$ for distinct i, j. That is because given person j selects his hat correctly, then there are two possibilities for E_i to select his hat (he can pick the correct one or the wrong one). Lastly for $\mathbb{P}(E_1 \cap E_2 \cap E_3)$ we write

$$\mathbb{P}(E_1 \cap E_2 \cap E_3) = \mathbb{P}(E_1|E_2 \cap E_3)\mathbb{P}(E_2 \cap E_3) = \mathbb{P}(E_1|E_2 \cap E_3)\mathbb{P}(E_2|E_3)\mathbb{P}(E_3) = 1 \cdot \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}.$$

Thus

$$\mathbb{P}(E_1 \cup E_2 \cup E_3) = (1) - (1/2) + (1/6) = \frac{4}{6}.$$

Then the probability of interest will be

$$\mathbb{P}(E) = 1 - \frac{4}{6} = \frac{1}{3}.$$

Solving by constructing a sample space. A suitable sample space for this problem can be the set of all permutations on three letters. This set is

$$\Omega = \{\begin{pmatrix} a & b & c \\ \boxed{a} & \boxed{b} & \boxed{c} \end{pmatrix}, \begin{pmatrix} a & b & c \\ \boxed{a} & c & b \end{pmatrix}, \begin{pmatrix} a & b & c \\ b & a & \boxed{c} \end{pmatrix}, \begin{pmatrix} a & b & c \\ b & c & a \end{pmatrix}, \begin{pmatrix} a & b & c \\ c & a & b \end{pmatrix}, \begin{pmatrix} a & b & c \\ c & \boxed{b} & a \end{pmatrix}\}.$$

Note that the elements in the box represents the fixed point of the permutation. The probability of interest is basically the number of permutations that has no fixed point. As it is clear from the set Ω, the probability is

$$\mathbb{P}(E) = \frac{2}{6} = \frac{1}{3}.$$

■ **Problem 1.74 — Conditional probability mass function (from Ross).** Let X, Y be two random variables with the joint probability mass function given as

$$P(1,1) = 0.5 \qquad P(1,2) = 0.1, \qquad P(2,1) = 0.1, \qquad P(2,2) = 0.3.$$

Calculate the conditional probability mass function of X given that $Y = 1$.

Solution We will use the following identity

$$P_{X|Y}(x|y) = \mathbb{P}(X = x|Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}.$$

Observe that

$$\mathbb{P}(Y = y) = \sum_x \mathbb{P}(Y = y, X = x)$$

thus $\mathbb{P}(Y = 1) = 0.5 + 0.1 = 0.6$. So we will have

$$P_{X|Y}(1|1) = \frac{0.5}{0.6} = \frac{5}{6}, \qquad P_{X|Y}(2|1) = \frac{0.1}{0.6} = \frac{1}{6}.$$

■ **Problem 1.75 — Conditional probability mass function for geometric random variables (from Ross).** Let X_1, X_2 be two independent random variables with geometric distributions with parameters (n_1, p) and (n_2, p). Calculate the conditional probability mass function of X_1 given that $X_1 + X_2 = m$.

Solution First, observe that $Y = X_1 + X_2$ is a binomial distribution with parameter $(n_1 + n_2, p)$. Thus we can write

$$P_{X_1|Y}(k|m) = \mathbb{P}(X_1 = k|Y = m) = \frac{\mathbb{P}(X_1 = k, X_1 + X_2 = m)}{\mathbb{P}(X_1 + X_2 = m)} = \frac{\mathbb{P}(X_1 = k, X_2 = m - k)}{\mathbb{P}(Y = m)}$$

Since the random variables X_1 and X_2 are independent, we can write

$$P_{X_1|Y}(k|m) = \frac{\mathbb{P}(X_1 = k)\mathbb{P}(X_2 = m - k)}{\mathbb{P}(Y = m)} = \frac{\binom{n_1}{k}\binom{n_2}{m-k}}{\binom{n_1+n_2}{m}}.$$

■ **Problem 1.76 — Conditional probability mass function for Poisson random variables (from Ross).** Let X, Y be two independent Poisson random variables with parameters λ_1 and λ_2 respectively. Calculate the conditional probability mass function for X given that $X_1 + X_2 = n$.

Solution First observe that $Z = X + Y$ is a Poisson random variable with parameter $\lambda_1 + \lambda_2$. Thus we will have

$$P_{X|X+Y}(m|n) = \frac{\mathbb{P}(X = m|X + Y = n)}{\mathbb{P}(X + Y = n)} = \frac{\mathbb{P}(X = m, Y = n - m)}{\mathbb{P}(X + Y = n)}$$

Given that X, Y are independent random variables then we can write

$$P_{X|X+Y}(m, n) = \frac{\mathbb{P}(X = m)\mathbb{P}(Y = n - m)}{\mathbb{P}(X + Y = n)} = \frac{\lambda_1^n \lambda_2^{n-m} n!}{m!(n-m)!(\lambda_1 + \lambda_2)^n} = \binom{n}{m}(\frac{\lambda_1}{\lambda_1 + \lambda_2})^m(\frac{\lambda_2}{\lambda_1 + \lambda_2})^{n-m}$$

Thus the conditional probability mass function of X given that $X + Y = n$ will be a binomial random variable with parameter $(n, \lambda_1/(\lambda_1 + \lambda_2))$. We can now easily compute the conditional expectation value as

$$\mathbb{E}\Big[X|X + Y = n\Big] = \frac{n\lambda_1}{\lambda_1 + \lambda_2}$$

■ **Problem 1.77** Let X, Y be two discrete random variables. Prove that

$$\mathbb{E}\Big[\mathbb{E}\Big[X|Y\Big]\Big] = \mathbb{E}\Big[X\Big].$$

Solution We start with the definition of the expectation of a discrete random variable.

$$\begin{aligned}
\mathbb{E}\Big[\mathbb{E}\Big[X|Y\Big]\Big] &= \sum_y \mathbb{E}\Big[X|Y = y\Big]\mathbb{P}(Y = y) = \sum_y \sum_x x\mathbb{P}(X = x|Y = y)\mathbb{P}(Y = y) \\
&= \sum_{x,y} x\mathbb{P}(X = x, Y = y) = \sum_x x \sum_y \mathbb{P}(X = x, Y = y) = \sum_x x\mathbb{P}(X = x) = \mathbb{E}\Big[X\Big]
\end{aligned}$$

■ **Problem 1.78 — The expectation of a random number of random variables (from Ross).** Let the expected number of injures in an industrial field be 4 per week. Also, assume that the number of workers injured at each incidence are independent random variables with average 2. Then what is the expected number of injuries in one week?

Solution Let $X_1, X_2, \cdots$ be i.i.d random variables representing the number of workers injured at each incidence. We are interested in

$$\mathbb{E}\Big[X_1 + \cdots + X_N\Big]$$

where N is a random variable representing the number of incidences occurred in a week. By the law of conditional expectation we can write

$$\mathbb{E}\Big[X_1 + \cdots + X_N\Big] = \sum_n \mathbb{E}\Big[X_1 + \cdots + X_n\Big]\mathbb{P}(N = n) = \sum_n n\mathbb{E}\Big[X\Big]\mathbb{P}(N = n) = \mathbb{E}\Big[X\Big]\mathbb{E}\Big[N\Big].$$

Thus the average number of workers injured in a week will be 8.

■ **Problem 1.79 — An alternative way to compute the expectation of a geometric random variable.** Consider a coin with probability p to fall heads. What is the expectation value of the number of tosses required until we get the first head?

Solution Let $X_1, X_2, \cdots$ be Bernoulli random variables with parameter p. Let N be a random variable denoting the number of tosses required until we get the first heads. We can condition the expected value of E to the first outcome.

$$\mathbb{E}\Big[N\Big] = \mathbb{E}\Big[N|X_1 = H\Big]\underbrace{\mathbb{P}(X_1 = H)}_{=p} + \mathbb{E}\Big[N|X_1 = T\Big]\underbrace{\mathbb{P}(X_1 = T)}_{=1-p}$$

Observe that

$$\mathbb{E}\Big[N|X_1 = H\Big] = 1, \qquad \mathbb{E}\Big[N|X_1 = T\Big] = 1 + \mathbb{E}\Big[N\Big].$$

Thus we will have

$$\mathbb{E}\Big[N\Big] = \frac{1}{p}.$$

■ **Problem 1.80 — Trapped miner (from Ross).** A miner is trapped in the mine and has three doors in front of him. He is equally likely to choose any of the three. The first door will take him to safety after 2 hours of walking, the second door will take him to the mine again after 3 hours of walking, and the third door will take him to the mine again after 5 hours of walking. What is the expected time that the miner will arrive to safety?

Solution Let $X_1, X_2, \cdots$ be random variables denoting the doors that the miner choose at each time that he attempts to escape. Furthermore, let T be a random variable showing the the time it takes for the miner to escape. To calculate $\mathbb{E}\Big[T\Big]$ we can condition it on the first door choice. I.e.

$$\mathbb{E}\Big[T\Big] = \mathbb{E}\Big[T|X_1 = 1\Big]\mathbb{P}(X_1 = 1) + \mathbb{E}\Big[T|X_1 = 2\Big]\mathbb{P}(X_1 = 2) + \mathbb{E}\Big[T|X_1 = 3\Big]\mathbb{P}(X_1 = 3)$$

Observe that

$$\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = 2) = \mathbb{P}(X_1 = 3) = 1/3.$$

Also

$$\mathbb{E}\Big[T|X_1 = 1\Big] = 2, \qquad \mathbb{E}\Big[T|X_1 = 2\Big] = 3 + \mathbb{E}\Big[T\Big], \qquad \mathbb{E}\Big[T|X_1 = 3\Big] = 3 + \mathbb{E}\Big[T\Big].$$

Thus we will have

$$\mathbb{E}\Big[T\Big] = 10.$$

So on average it will take the miner to exit the mine in 10 hours. Note that this does not guarantee that the miner will eventually escape. It is possible that we will get in trap by repeatedly choosing the door number 3.

■ **Problem 1.81 — From Rosenthal.** Suppose that $\Omega = \{1, 2\}$, with $\mathbb{P}(\varnothing) = 0$ and $\mathbb{P}(\{1, 2\}) = 1$. Suppose $\mathbb{P}(\{1\}) = \frac{1}{4}$. Prove that $\mathbb{P}$ is countably additive if and only if $\mathbb{P}(\{2\}) = \frac{3}{4}$.

Solution The proof has two parts

$\boxed{\Longrightarrow}$ Since $\mathbb{P}$ is countably additive, then

$$\mathbb{P}(\{1\}\dot\cup\{2\}) = \mathbb{P}(\{1\}) + \mathbb{P}(\{2\}) = 1.$$

This implies $\mathbb{P}(\{2\}) = 3/4$.

$\boxed{\Longleftarrow}$ Assume $\mathbb{P}(\{2\}) = 3/4$. Then it is very straightforward to check that for every disjoint subset $A, B \subset \Omega$, we have

$$\mathbb{P}(A\dot\cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

Thus we conclude that $\mathbb{P}$ is countably additive.

■ **Problem 1.82 — From Rosenthal.** Suppose $\Omega = \{1, 2, 3\}$ and $\mathcal{F}$ is a the collection of all subsets of Ω. Find (with proof) necessary and sufficient conditions on the real numbers x, y, z such that there exists a countably additive probability measure $\mathbb{P}$ on $\mathcal{F}$ such that $x = \mathbb{P}\{1, 2\}, y = \mathbb{P}\{2, 3\}, z = \mathbb{P}\{1, 3\}$.

Solution To find the necessary conditions, we assume that $\mathbb{P}$ is an additive probability measure. Let $a = \mathbb{P}\{1\}, b = \mathbb{P}\{2\}$, and $c = \mathbb{P}\{3\}$. Then the countable additivity implies

$$a + b = x, \qquad b + c = y, \qquad a + c = z.$$

Then due to countable additivity, and the fact that $\mathbb{P}$ is a probability measure (i.e. $\mathbb{P}\{1, 2, 3\} = 1$), we have $a + b + c = 1$, thus we need to have

$$x + y + z = 2. \tag{♪}$$

Further, we solve the a, b, c in terms of x, y, z are require the singleton probabilities to be positive. We have

$$a = \frac{x - y + z}{2}, \qquad b = \frac{x + y - z}{2}, \qquad c = \frac{-x + y + z}{2}.$$

One of the necessary conditions is also to have

$$x - y + z \geq 0, \qquad x + y - z \geq 0, \qquad -x + y + z \geq 0. \tag{♫}$$

The two conditions (♪) and (♫) together are the necessary and sufficient conditions for $\mathbb{P}$ to be a valid probability measure.

■ **Problem 1.83 — From Rosenthal.** Suppose that $\Omega = \mathbb{N}$ is the set of positive integers, and $\mathbb{P}$ is defined for all $A \subseteq \Omega$ by $\mathbb{P}(A) = 0$ if A is finite, and $\mathbb{P}(A) = 1$ if A is infinite. Is $\mathbb{P}$ finitely additive?

Solution Not it is not. Consider the partitioning of the set Ω by the even E and odd O integers.

$$\mathbb{P}(\Omega) = \mathbb{P}(E) + \mathbb{P}(O) \implies 1 = 2,$$

which is not true. Thus $\mathbb{P}$ is not finitely additive.

■ **Problem 1.84 — From Rosenthal.** Suppose that $\Omega = \mathbb{N}$, and $\mathbb{P}$ is defined for all $A \subseteq \Omega$ by $\mathbb{P}(A) = |A|$ if A is finite, and $P(A) = \infty$ if A is infinite. This P is of course not a probability measure (in fact it is counting measure), however we can still ask the following: (be the convention $\infty + \infty = \infty$)

(I) Is $\mathbb{P}$ finitely additive?

(II) Is $\mathbb{P}$ countably additive?

Solution (I) Yes. $\mathbb{P}$ being finitely additive is equivalent being additive for disjoint $A, B \subset \Omega$. There are three cases for these choices

- (i) A, B are both finite. In this case $\mathbb{P}(A \dot\cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ since $0 = 0 + 0$.
- (ii) A, B are both infinite. In this case $\mathbb{P}(A \dot\cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ since $\infty + \infty = \infty$.
- (iii) One of the sets A, B is infinite. Then $\mathbb{P}(A \dot\cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ since $0 + \infty = \infty$.

(II) No. We will show this by counterexample. We can write $\Omega = \dot\bigcup_{i \in \mathbb{N}} \{i\}$. Then the countable additivity implies

$$\mathbb{P}(\Omega) = \mathbb{P}(\dot\bigcup_{i \in \mathbb{N}} \{i\}) \implies 1 = 0.$$

which is not true.

■ **Problem 1.85 — From Rosenthal.** Let $\mathcal{I}$ be the set of all intervals in $[0,1]$ (open, closed, half-open, singleton, empty set). Show that $\mathcal{I}$ is a semi-algebra.

Solution By definition of $\mathcal{I}$ we have $\varnothing \in \mathcal{I}$. Let A_1, A_2 be two intervals in $[0,1]$. If A_1, A_2 are disjoint, then $A_1 \cap A_2 \in \mathcal{I}$. If they are not disjoint, then without loss of generality we can assume that

$$A_1 = \{x \in [0,1] \mid a < x < b\}, A_2 = \{x \in [0,1] \mid c < x < d\},$$

where $a < c < b < d$. Thus $A_1 \cap A_2 = (c, b)$. So $\mathcal{I}$ is closed under finite intersection. The proof is the same for any other choices of A_1, A_2 (i.e. being closed set, etc). To show the third property, again, without the loss of generality, let $A = (a, b)$. Then $A^c = (-\infty, a] \dot\cup [b, \infty)$. For other choices of A (i.e. being closed, etc) we will have a similar argument. Thus we conclude that the collection $\mathcal{I}$ is a semi-algebra of the subsets of $[0,1]$.

■ **Problem 1.86 — From Rosenthal.** Let $\mathcal{I}$ be the semi-algebra consisting of all intervals in $[0,1]$. Define

$$\mathcal{B}_0 = \{\text{all finite unions of elements of } \mathcal{I}.\}$$

Show that $\mathcal{B}_0$ is not a σ-algebra.

Solution Along with many other sets, $\mathcal{I}$ contains all of the singletons, so does $\mathcal{B}_0$. Consider the following collection

$$\mathcal{A} = \{\{x\} \ : \ x \in \mathbb{Q} \cap [0,1]\}.$$

By definition, all of the sets in the collection $\mathcal{A}$ belongs to $\mathcal{B}_0$. However, the following countable union

$$\dot\bigcup_{A \in \mathcal{A}} A = [0,1] \cap \mathbb{Q}$$

does not belong to $\mathcal{B}_0$ (as it is not possible to generate with only finite unions of the elements of singletons in $\mathcal{I}$).

■ **Problem 1.87 — From Rosenthal.** Prove that the outer measure $\mathbb{P}^*$ is countably sub-additive, i.e.

$$\mathbb{P}^*(\bigcup_{n=1}^{\infty} B_n) \leq \sum_{n=1}^{\infty} \mathbb{P}^*(B_n) \qquad \text{for any } B_1, B_2, \cdots \in \Omega.$$

Solution This problem is the proof of Lemma 2.3.6 in Rosenthal. See the text for more context. A very quick review on the context is that we have a semi-algebra $\mathcal{I}$ of the subsets of Ω, and we have the function $\mathbb{P} : \mathcal{I} \to [0,1]$ that satisfies the properties required for the extension theorem, hence there exist a valid probability space $(\Omega, \mathcal{M}, \mathbb{P}^*)$, where $\mathcal{M}$ is a σ-algebra and $\mathbb{P}^*$ is a probability measure (which is also the outer measure). The proof of this question is as follows.

Fix $\epsilon > 0$. Since the outer measure is the infimum of the sum of the probabilities on all $\mathcal{I}$ covers, then for each B_n we can find a collection $\{C_{nk}\}$ where $C_{nk} \in \mathcal{I}$ such that

$$\sum_k \mathbb{P}(C_{nk}) \leq \mathbb{P}^*(B_n) + \epsilon 2^{-n}.$$

On the other hand, since $\{C_{nk}\}_{nk}$ covers $\bigcup_n B_n$, then again from the properties of inf we have

$$\mathbb{P}^*(\bigcup_n B_n) \leq \sum_{nk} \mathbb{P}(C_{nk}).$$

combining these two we will get

$$\mathbb{P}^*(\bigcup_n B_n) \leq \sum_n \mathbb{P}^*(B_n) + \epsilon.$$

Since this is true for all $\epsilon > 0$, then it implies that

$$\mathbb{P}^*(\bigcup_n B_n) = \sum_n \mathbb{P}^*(B_n).$$

■ **Problem 1.88 — From Rosenthal.** If $A_1, A_2, \cdots \in \mathcal{M}$ are disjoint, then prove that

$$\mathbb{P}^*(\dot{\bigcup}_n A_n) = \sum_n \mathbb{P}^*(A_n).$$

Solution First, we start to show the finite additivity, and then using the properties of monotnonicity and sub-additivity, we will prove the countable additivity as well. Let $A_1, A_2 \in \mathcal{M}$ disjoint. In particular, since $A_1 \in \mathcal{M}$, from the definition of $\mathcal{M}$ (see page 12 Rosenthal), then

$$\mathbb{P}^*(A_1 \cup A_2) = \mathbb{P}^*(A_1^c \cap (A_1 \cup A_2)) + \mathbb{P}^*(A_1 \cap (A_1 \cup A_2)) = \mathbb{P}^*(A_2) + \mathbb{P}^*(A_1).$$

This implies that for any finite disjoint collection of A_i we have the additivity property (by induction). Now for any $m \in \mathbb{N}$ we have

$$\sum_{n<m} \mathbb{P}^*(A_n) = \mathbb{P}^*(\bigcup_{n\leq m} A_n) \leq \mathbb{P}^*(\bigcup_n A_n)$$

where the last inequality follows form the monotnonicity property of $\mathbb{P}^*$. Since this is true for all $m \in \mathbb{N}$, then we conclude that

$$\mathbb{P}^*(\bigcup_n A_n) \geq \sum_n \mathbb{P}^*(A_n).$$

On the other hand, from the sub-additivity property we have

$$\mathbb{P}^*(\bigcup_n A_n) \leq \sum_n \mathbb{P}^*(A_n).$$

These two implies that

$$\mathbb{P}^*(\bigcup_n A_n) = \sum_n \mathbb{P}^*(A_n).$$

■ **Problem 1.89 — From Rosenthal.** Let $\mathcal{M}$ be the σ-algebra we get from the extension theorem, where by definition it contains all of the sets like $A \in \Omega$ for which the outer measure is additive on the union of $A \cap E$ and $A^c \cap E$ for $\forall E \subset \Omega$. In other words

$$\mathcal{M} = \{A \subseteq \Omega \ : \ \mathbb{P}^*(A \cap E) + \mathbb{P}^*(A^c \cap E) = \mathbb{P}^*(E) \quad \text{for all } E \subset \Omega\}.$$

Prove that $\mathcal{M}$ is an algebra.

Solution Let $A = \Omega$. Then

$$\mathbb{P}^*(A \cap E) + \mathbb{P}^*(A^c \cap E) = \mathbb{P}^*(E) + \mathbb{P}^*(\varnothing) = \mathbb{P}^*(E).$$

So we conclude that $\Omega \in \mathcal{M}$. Also, it follows immediately from the definition of $\mathcal{M}$ that if $A \in \mathcal{M}$ then $A^c \in \mathcal{M}$. Now it remains to show if $A_1, A_2 \in \mathcal{M}$ then $A_1 \cap A_2 \in \mathcal{M}$. Let $E \subset \Omega$. Then

$$\begin{aligned}
&\mathbb{P}^*(E \cap (A_1 \cap A_2)) + \mathbb{P}^*(E \cap (A_1 \cap A_2)^c) \\
&= \mathbb{P}^*(E \cap (A_1 \cap A_2)) + \mathbb{P}^*(E \cap A_1^c \cap A_2) + \mathbb{P}^*(E \cap A_1 \cap A_2^c) + \mathbb{P}^*(E \cap A_1^c \cap A_2^c) \\
&\leq \mathbb{P}^*(E \cap A_1 \cap A_2) + \mathbb{P}^*(E \cap A_1^c \cap A_2) + \mathbb{P}^*(E \cap A_1 \cap A_2^c) + \mathbb{P}^*(E \cap A_1^c \cap A_2^c) \\
&= \mathbb{P}^*(E \cap A_2) + \mathbb{P}^*(E \cap A_2^c) \qquad (\text{because } A_1 \in \mathcal{M}) \\
&= \mathbb{P}^*(E) \qquad (\text{because } A_2 \in \mathcal{M}).
\end{aligned}$$

On the other hand, from the sub-additivity property we know that

$$\mathbb{P}^*(E \cap (A_1 \cap A_2)) + \mathbb{P}^*(E \cap (A_1 \cap A_2)^c) \geq \mathbb{P}^*(E).$$

Thus we conclude that

$$\mathbb{P}^*(E \cap (A_1 \cap A_2)) + \mathbb{P}^*(E \cap (A_1 \cap A_2)^c) = \mathbb{P}^*(E).$$

This implies that $A_1 \cap A_2 \in \mathcal{M}$ and this finishes the proof.

■ **Problem 1.90 — From Rosenthal.** Let $A_1, A_2, \cdots \in \mathcal{M}$ be disjoint. For each $m \in \mathbb{N}$, let $B_m = \bigcup_{n \leq m} A_n$. Prove that for all $m \in \mathbb{N}$, and for all $E \subseteq \Omega$ we have

$$\mathbb{P}^*(E \cap B_m) = \sum_{n \leq m} \mathbb{P}^*(E \cap A_n).$$

Solution First, observe that this statement is true for $m = 1$ in a trivial way. For $m = 2$, since $A_2 \in \mathcal{M}$, then we can expand $E \cap B_2$ according to A_2, i.e.

$$\mathbb{P}^*(E \cap B_2) = \mathbb{P}^*((E \cap B_2) \cap A_2) + \mathbb{P}^*((E \cap B_2) \cap A_2^c)$$

On the other hand $(E \cap B_2) \cap A_2 = E \cap A_2$ and $(A \cap B_2) \cap A_2^c = E \cap B_1 = E \cap A_1$. Thus we can write

$$\mathbb{P}^*(E \cap B_2) = \mathbb{P}^*(E \cap A_1) + \mathbb{P}^*(E \cap A_2).$$

In general, for $m \in \mathbb{N}$ we can write

$$\mathbb{P}^*(E \cap B_m) = \mathbb{P}^*(E \cap A_m) + \mathbb{P}^*(E \cap B_{m-1}).$$

Thus using induction we can write

$$\mathbb{P}^*(E \cap B_m) = \sum_{n \leq m} \mathbb{P}^*(E \cap A_n).$$

■ **Problem 1.91 — From Rosenthal.** In this question covers some of the proves for constructing a uniform probability measure on $\Omega = [0, 1]$. Let $\mathcal{I}$ be the set of all intervals in Ω, and let $\mathcal{P} : \mathcal{I} \to [0, 1]$ be a function that assigns the length of an interval to that interval. We want to prove that $\mathcal{P}$ satisfies the property (2.3.3) of extension theorem (2.3.1) in Rosenthal. I.e. we want to prove that for $A, A_1, \cdots \in \mathcal{I}$ such that $A \subset \cup_i A_i$ satisfies

$$\mathbb{P}(A) \leq \sum_i \mathbb{P}(A_i).$$

Solution We will do this in three parts.

(i) Step 1. First, we prove that for any finite $A, A_1, \cdots, A_n \in \mathcal{I}$ collection where $A \subset \cup_{i=1}^n A_i$ we have

$$\mathbb{P}(A) \leq \sum_{i=1}^n \mathbb{P}(A_i).$$

To see this let $A_1, A_2, A \in \mathcal{I}$ such that $A \subset A_1 \cup A_2$. We also assume that $A_1 \cap A \neq \varnothing$ as well as $A_2 \cap A \neq \varnothing$. This is to ensure that we do not have redundant interval in our collection that does not cover A. Note that from any collection of intervals we can put aside the redundant intervals and do our reasoning here and then finally at the last step consider the redundant intervals as well. So our assumption above does not lose the generality of the proof. Let a_i, b_i represent the left (right) endpoints of the intervals A_i and a_0, b_0 represent the left (right) endpoints of the interval A. In order to the intervals A_1, A_2 to cover A white neither of them are redundant we need to have

$$\min\{a_1, a_2\} \leq a_0 \leq a_2 \leq b_1 \leq b_0 \leq \max\{b_1, b_2\}.$$

Then it follows that

$$b_0 - a_0 \leq (b_1 - a_1) + (b_2 - a_2).$$

Thus this implies that

$$\mathbb{P}(A) \leq \mathbb{P}(A_1) + \mathbb{P}(A_2).$$

We can generalize this by induction to any finite number of intervals.

(ii) Step 2. We now want to prove that for any countable *open* intervals $A_1, A_2, \cdots \in \mathcal{I}$ such that $A \subset \cup_n A_n$ for $A \in \mathcal{I}$ closed, we have

$$\mathbb{P}(A) \leq \sum_n \mathbb{P}(A_n).$$

To see this, we will use the Heine-Borel theorem. Since the collection $\{A_1, A_2, \cdots\}$ is an open cover for the closed set A. if A is the whole space Ω, then the inequality that we want to show follows immediately (LHS is 1 while RHS is infinite). However, if A is not the whole space, then it is bounded. Thus A is closed and bounded, hence compact (by Heine-Borel). So the open cover has a finite sub-cover and this completes the proof by reducing this case to case (i) above.

(iii) Step 3. We now want to show that if $A_1, A_2, \cdots \in \mathcal{I}$ is any countable collection of intervals, and if $A \subset \cup_n A_n$ for any $A \in \mathcal{I}$ then

$$\mathbb{P}(A) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

TODO: TO BE COMPLETED

■ **Problem 1.92 — From Rosenthal.** Let $\mathcal{A} = \{(-\infty, x] \ : \ x \in \mathbb{R}\}$. Prove that $\sigma(\mathcal{A}) = \mathcal{B}$, i.e. that the smallest σ-algebra of subsets of $\mathbb{R}$ which contains $\mathcal{A}$ is equal to the Borel σ-algebra of subsets of $\mathbb{R}$.

Solution By definition, we know that $\mathcal{B}$ is the smallest σ-algebra that contains all of the intervals. However, we claim that the σ-algebra $\sigma(\mathcal{A})$ also contains all of intervals. To see this, let I be an interval which will have different cases $(-\infty, a), (-\infty, a], (a, b), [a, b], (a, b], [a, b), (a, \infty), [a, \infty)$. Each of these sets can be constructing by using the sets in $s\sigma(\mathcal{A})$ and using its sigma-algebra properties. Thus we showed that $\sigma(\mathcal{A})$ contains $\mathcal{I}$ the set of all intervals. However, by definition $\mathcal{B}$ was the smallest σ-algebra containing all of intervals. Thus the σ-algebra generated by $\mathcal{A}$ (i.e. the smallest σ-algebra by definition) is equal to $\mathcal{B}$.

■ **Problem 1.93 — From Rosenthal.** Prove the following statements.

(a) Prove that the Cantor set K and its complement K^c is in $\mathcal{B}$, the Borel set of the subset of $[0, 1]$.

(b) Prove that $K, K^c \in \mathcal{M}$, where $\mathcal{M}$ is the σ-algebra by the extension applied for the uniform distribution on $[0, 1]$ (see theorem 2.4.4 Rosenthal).

(c) Prove that $K^c \in \mathcal{B}_1$ where $\mathcal{B}_1$ is defined by (2.2.6) Rosenthal (i.e. the set of all finite or countable unions of intervals $\mathcal{I}$).

(d) Prove that $\mathcal{B}_1$ is not a σ-algebra.

Solution (a) In the construction of the cantor set, at each step we remove the middle 1/3 of the intervals. So starting with $I_0 = [0, 1]$ at the first step we will get $I_1 = [0, 1/3] \cup [2/3, 1]$, etc. The Cantor set is $I_0 \cap I_1 \cap I_2 \cap \cdots$. Since $I_i \in \mathcal{B}$ for all $i \in \mathbb{N}$, and $\mathcal{B}$ is a σ-algebra and is closed under countable intersection, then $K \in \mathcal{B}$ as well. It follows immediately that $K^c \in \mathcal{B}$ as well, as $\mathcal{B}$ is closed under complement.

(b) Since $\mathcal{M} \supset \mathcal{B}$, and as we showed above that $K, K^c \in \mathcal{B}$, then $K, K^c \in \mathcal{M}$ as well.

(c) From the construction given in part (a), we have

$$K = I_0 \cap I_1 \cap I_2 \cap \cdots .$$

From the De Morgan's law we will have

$$K^c = I_0^c \cup I_1^c \cup I_2^c \cup \cdots .$$

Thus by definition of $\mathcal{B}_1$ we have $K^c \in \mathcal{B}_1$.

(d) First, observe that (see Rosenthal page 17) that the Cantor set is uncountable. So there is no way to construct it by finite or countable union of singletons (which are in $\mathcal{I}$). We can not construct it with the finite or countable union of any intervals as K is a nowhere dense set. I.e. for every interval $(a, b) \in \mathcal{I}$ containing $x \in K$ there exists, $y \in \mathbb{R}$ such that $y \notin K$. To put this precisely, let $I_1, I_2, I_3, \cdots$ is collection of intervals such that $I = \cup_n I_n$. Since K is nowhere dense, for any I_i in the collection that contains $x \in K$ we can find some $y \in \mathbb{R}$ that $y \notin K$. This is a contradiction and we have $I \subset \cup_n I_n$.

(e) As we saw above, $K^c \in \mathcal{B}_1$ but $K \notin \mathcal{B}_1$. Thus $\mathcal{B}_1$ is not closed under complement, thus it is not a σ-algebra.

■ **Problem 1.94 — An extension of the extension theorem (from Rosenthal).** Let $\mathcal{I}$ be a semialgebra of subsets of Ω. Let $\mathbb{P} : \mathcal{I} \to [0, 1]$ with $\mathbb{P}(\varnothing) = 0$ and $\mathbb{P}(\Omega) = 1$, satisfying

$$\mathbb{P}(\bigcup_n A_n) \geq \sum_n \mathbb{P}(A_n) \qquad \text{for } A_1, A_2, \cdots \in \mathcal{I} \textit{ disjoint}, \text{ and } \bigcup_n A_n \in \mathcal{I},$$

as well as

$$\mathbb{P}(A) \leq \mathbb{P}(B), \qquad A \subseteq B,$$

and

$$\mathbb{P}(\bigcup_n B_n) \leq \sum_n \mathbb{P}(B_n) \qquad \text{for } B_1, B_2, \cdots \in \mathcal{I}, \text{ and } \bigcup_n B_n \in \mathcal{I}.$$

Then there exist a valid probability space $(\Omega, \mathcal{M}, \mathbb{P}^*)$ such that $\mathbb{P}$ and $\mathbb{P}^*$ agree on the elements of $\mathcal{I}$

Solution According to the extension theorem 2.3.1 we need to prove that these alternative statements implies 2.3.3. Let $A, A_1, A_2, \cdots \in \mathcal{I}$ (note that the union of A_is do not belong to $\mathcal{I}$ necessarily). Define

$$B_n = A_n \cap A.$$

Then we will have $A = \cup_n B_n$, thus $\cup_n B_n \in \mathcal{I}$. So

$$\mathbb{P}(A) = \mathbb{P}(\bigcup_n B_n) \leq \sum_n \mathbb{P}(B_n) \leq \sum_n \mathbb{P}(A_n).$$

Observation 1.10.1 In the prove above, one might attempt

$$\mathbb{P}(A) \le \mathbb{P}(\bigcup_i A_i) \le \sum_i \mathbb{P}(A_i)$$

where for the first inequality we use the monotnonicity (as $A \subset \bigcup_i A_i$) and for the second inequality we use the countable sub additivity given in the statement of the theorem. However, this prove is *wrong!*. Because we are not allowed to use the second inequality, as it does not satisfies the requirements for the sub-additivity statement to work. That is because $\bigcup_n A_n$ may not be in $\mathcal{I}$ necessarily.

■ **Problem 1.95 — Uniqueness property of the extension theorem (from Rosenthal).** Let $\mathcal{I}, \mathbb{P}, \mathbb{P}^*$ be the same as in Theorem 2.3.1 (Rosenthal). Let $\mathcal{F}$ be any σ-algebra with $\mathcal{I} \subseteq \mathcal{F} \subseteq \Omega$. Let $\mathbb{Q}$ be any probability measure on $\mathcal{F}$, such that $\mathbb{Q}(A) = \mathbb{P}(A)$ for all $A \in \mathcal{I}$. Then prove that $\mathbb{Q}(A) = \mathbb{P}^*(A)$ for all $A \in \mathcal{F}$.

Solution Let $A \in \mathcal{F}$, and $A_i \in \mathcal{I}$ for $i = 1, 2, \dots$ such that $A \subseteq \bigcup_i A_i$. Since $\mathbb{Q}$ is a probability measure, then

$$\mathbb{Q}(A) \le \mathbb{Q}(\bigcup_i A_i) \le \sum_i \mathbb{Q}(A_i).$$

Then we can write

$$\begin{aligned}
\mathbb{P}^*(A) &= \inf_{\substack{A_1, A_2, \dots \in \mathcal{I} \\ A \subseteq \bigcup_i A_i}} \sum_i \mathbb{P}(A_i) \\
&= \inf_{\substack{A_1, A_2, \dots \in \mathcal{I} \\ A \subseteq \bigcup_i A_i}} \sum_i \mathbb{Q}(A_i) \\
&\ge \inf_{\substack{A_1, A_2, \dots \in \mathcal{I} \\ A \subseteq \bigcup_i A_i}} \mathbb{Q}(\bigcup_i A_i) \\
&\ge \inf_{\substack{A_1, A_2, \dots \in \mathcal{I} \\ A \subseteq \bigcup_i A_i}} \mathbb{Q}(A) \\
&= \mathbb{Q}(A).
\end{aligned}$$

However, we could do the same with $A^c \in \mathcal{F}$, which would lead to

$$\mathbb{P}^*(A^c) \ge \mathbb{Q}(A^c).$$

Since $\mathbb{P}^*(A^c) = 1 - \mathbb{P}^*(A)$ and $\mathbb{Q}(A^c) = 1 - \mathbb{Q}(A)$, then this implies $\mathbb{P}^*(A) \ge \mathbb{Q}(A)$. Thus $\mathbb{P}^*(A) = \mathbb{Q}(A)$.

■ **Problem 1.96 — Properties of Random Variables (from Rosenthal).** Prove the followings.

(i) If X, Y are random variables and $c \in \mathbb{R}$, then $X + c, cX, X^2, X + Y$, and XY are random variables.

(ii) If $Z_1, Z_2, \cdots$ are random variables such that $\lim_{n\to\infty} Z_n(\omega)$ exists for all $\omega \in \Omega$,, and $Z(w) = \lim_{n\to\infty} Z_n(\Omega)$, then Z is also a random variable.

Solution the proves are as follows.

(i) (a) $(X + c)^{-1}((-\infty, x]) = \{w : X(w) + c \le x\} = \{w : X(w) \le x - c\} = X^{-1}((0\text{-}\infty, x - c]) \in \mathcal{F}$.

(b) Assume $c \neq 0$. Then $(cX)^{-1}((-\infty, x]) = \{w : cX(w) \leq x\} = \{w : X(w) \leq x/c\} = X^{-1}((-\infty, x/c]) \in \mathcal{F}$. For the case were $c = 0$, then $cX \equiv 0$ on all Ω, and this is a random variable as $(cX^{-1})((0, x]) = \varnothing$ if $x < 0$ and $(cX^{-1})((0, x]) = \Omega$ if $x \geq 0$.

(c) $X^2((-\infty, a]) = \{w : X^2 \leq a\} = \{w : X \in [-\sqrt{a}, \sqrt{a}]\} \in \mathcal{F}$.

(d) $(X+Y)^{-1}((\infty, x]) = \{w : X(w) + Y(w) < x\} = \bigcup_{r \in \mathbb{Q}}(\{X < r\} \cap \{Y < x - r\})$. Since this is a countable union, thus in $\mathcal{F}$.

(e) I have the following prove but I am not sure if it is a correct one or not. I feel that this is a correct proof as there seems to be nothing that can make it not to work.

$$(XY)^{-1}((-\infty, x]) = \{XY < x\} = \bigcup_{n \in \mathbb{N}} (\{X < n\} \cap \{Y < x/n\}).$$

The following prove is the idea by Rosenthal. Once we know that X^2, $X + Y$, and cX are random variables, then we can deduce XY is also a random variable as

$$XY = \frac{1}{2}((X+Y)^2 - X^2 - Y^2))$$

(ii) In a nutshell this statements claims that the point-wise convergence of a sequence of random variables is a random variable. We need to show that the event $\{Z \leq r\} \in \mathcal{F}$. To see this, let $w \in \{Z \leq r\}$. This means that $Z(w) \leq r$. Since $Z_n(\omega) \to Z(\omega)$ as $n \to \infty$, then this implies that $\forall m \in \mathbb{N}$ there exists $N \in \mathbb{N}$ such that $\forall n > N$ we have $Z_n(\omega) \leq r + \frac{1}{m}$. Thus we can write

$$\{Z(w) \leq r\} = \bigcap_{m=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \{Z_n(\omega) \leq r + \frac{1}{m}\}$$

■ **Problem 1.97** Let $f : \mathbb{R} \to \mathbb{R}$ be continouse or piece-wise continuous function. Then f is a Borel function.

Solution We first prove the statement for a continuous function. To show that f is a Borel function we need to show $f^{-1}(B) \in \mathcal{B}$ for $B \in \mathcal{B}$. Since the pre-image preserves the union, intersection, and complements, and by definition for a continuous function the pre-image of an open set is an open set, and using the fact that we can write any Borel set as a countable intersection, union, or complements of open sets then we conclude that $f^{-1}(B) \in \mathcal{B}$. A second way to show this is observe that $f^{-1}((x, \infty)) \in \mathcal{B}$ as (x, ∞) is open and f is continuous, thus its pre-image is also an open set thus a Borel set. Then its complement is also a Borel set, i.e.

$$(f^{-1}((x, \infty)))^c = f^{-1}((-\infty, x]) \in \mathcal{B}.$$

Thus shows that f is a Borel function (since the pre-image of $(0, x]$ is a Borel set.

For the case where f is piece-wise continuous, by the definition of the piece-wise continuoity, f has at most countably many discontinuities. The we can write f as

$$f(x) = f_1(x)\mathbb{1}_{I_1}(x) + f_2(x)\mathbb{1}_{I_2}(x) + f_3(x)\mathbb{1}_{I_3}(x) + \cdots + f_n(x)\mathbb{1}_{I_n}(x),$$

where $I_1, I_2, I_3 \cdots, I_n$ are disjoint intervals on which $f_1, f_2, f_3, \cdots, f_n$ are continuous respectively. By the statement for the first part of the proof, we know that f_i is a Borel function (since it is continuous) as well as the indicator function $\mathbb{1}_{I_i}$. Their multiplication is also a Borel function and the sum of these Borel functions is a also a Borel function, thus $f(x)$ is a Borel function.

■ **Problem 1.98** Prove that if A, B are two independent events, then (A^c, B), (A, B^c), and (A^c, B^c) are pairwise independent.

Solution Since A, B are independent, then $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. From the identity $B = (B \cap A)\dot{\cup}(B \cap A^c)$. From the properties of the probability measure we have

$$\mathbb{P}(B) = \underbrace{\mathbb{P}(A \cap B)}_{\mathbb{P}(A)\mathbb{P}(B)} + \mathbb{P}(A^c \cap B).$$

Then we can write

$$\mathbb{P}(A^c \cap B) = \mathbb{P}(B) - \mathbb{P}(A \cap B) = \mathbb{P}(B)(1 - \mathbb{P}(A)) = \mathbb{P}(B)\mathbb{P}(A^c).$$

Thus A^c, B are also independent events. We use a similar argument for A, B^c. To show that the events A^c, B^c are also independent, we use the inclusion-exclusion principle.

$$\mathbb{P}(A^c \cap B^c) = 1 - \mathbb{P}(A \cup B) = 1 - (\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B)) = (1 - \mathbb{P}(A))(1 - \mathbb{P}(B)) = \mathbb{P}(A^c)\mathbb{P}(B^c).$$

Another way of showing this without the inclusion-exclusion principle is to use the identity

$$A^c = (A^c \cap B)\dot{\cup}(A^c \cap B^c).$$

Then

$$\mathbb{P}(A^c) = \underbrace{\mathbb{P}(A^c \cap B)}_{\mathbb{P}(A^c)\mathbb{P}(B)} + \mathbb{P}(A^c \cap B^c).$$

We can write

$$\mathbb{P}(A^c \cap B^c) = \mathbb{P}(A^c)(1 - \mathbb{P}(B)) = \mathbb{P}(A^c)\mathbb{P}(B^c).$$

■ **Problem 1.99** Let X, Y be random variables, and $f, g : \mathbb{R} \to \mathbb{R}$ be Borel functions. Then if X, Y are independent, $f(X)$ and $g(X)$ are also independent.

Solution Since X, Y are independent, thus for any $S_1, S_2 \in \mathcal{B}$ we have

$$\mathbb{P}(\{X \in S_1\} \cap \{Y \in S_2\}) = \mathbb{P}(\{X \in S_1\})\mathbb{P}(\{Y \in S_2\}).$$

Consider

$$\mathbb{P}(\{f(X) \in S_1\} \cap \{g(Y) \in S_2\}) = \mathbb{P}(\{X \in f^{-1}(S_1)\} \cap \{Y \in g^{-1}(S_2)\}) = \mathbb{P}(\{f(X) \in S_1\})\mathbb{P}(\{g(Y) \in S_2\}).$$

The equality above hods because for any $S_1, S_2 \in \mathcal{B}$ we have $f^{-1}(S_1), g^{-1}(S_2) \in \mathcal{B}$, that is because f, g are Borel functions.

■ **Problem 1.100 — Continuity of probabilities.** Prove that the probability measure function is continuous from bellow and above. I.e. for the continuity from below, let $A, A_1, A_2, \cdots \in \mathcal{F}$ such that $\{A_n\} \nearrow A$, i.e. $A_1 \subseteq A_2 \subseteq \cdots$ and $A = \bigcup_n A_n$. Then $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$. For the continuity from above, let $A, A_1, A_2, \cdots \in \mathcal{F}$ such that $\{A_n\} \searrow A$, i.e. $A_1 \subseteq A_2 \subseteq \cdots$ and $A = \bigcap_n A_n$, then $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$.

Solution First, we prove the probability from below. In general, one can observe that if $\{A_n\} \nearrow A$, then $\{\mathbb{P}(A_n)\}$ indeed converges as this is a bounded monotone sequence in $\mathbb{R}$. However, to show that this sequence converges to $\mathbb{P}(A)$, we do as following. Consider the following sets

$$B_1 = A_1, \qquad B_2 = A_2 \backslash A_1, \qquad B_3 = A_3 \backslash A_2, \cdots.$$

Then $A = \dot{\cup} B_n$. Thus $\mathbb{P}(A) = \sum_{n=1}^{\infty} \mathbb{P}(B_n)$. Thus the series on the rand hand side converges. This implies that the corresponding partial sums also converges. However, by the construction we have

$$\mathbb{P}(A_1) = \mathbb{P}(B_1), \quad \mathbb{P}(A_2) = \mathbb{P}(B_1) + \mathbb{P}(B_2) + \cdots.$$

Thus the converges of the partial sums implies the convergence of $\{\mathbb{P}(A_n)\}$.

For the proof for the continuity from above, first observe that if for a collection $A, A_1, A_2, \cdots \in \mathcal{F}$ we have $\{A_n\} \searrow A$, then this is equivalent to $\{A_n^c\} \nearrow A^c$. This follows from the De Morgan's law as well as the change of the direction of the inclusion $\subseteq$ under taking complements. By hypothesis we have $\{A_n\} \searrow A$, which is equivalent to$\{A_n^c\} \nearrow A^c$. From the first part of the proof, it follows that $\lim_{n\to\infty} \mathbb{P}(A_n^c) = \mathbb{P}(A^c)$. Thus

$$\lim_{n\to\infty} (1 - \mathbb{P}(A_n)) = 1 - \mathbb{P}(A).$$

Note that $\mathbb{P}(A_n)$ converges to some real number as $n \to \infty$, because it is a bounded decreasing sequence. Thus by the laws of the limit we can write

$$1 - \lim_{n\to\infty} \mathbb{P}(A_n) = 1 - \mathbb{P}(A).$$

Then it follows that

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A).$$

■ **Problem 1.101 — From Rosenthal.** Let $A_1, A_2, \cdots, A_n \in \mathcal{F}$. Generalize the principle of inclusion-exclusion to:

$$\mathbb{P}(A_1 \cup \cdots \cup A_n) = \sum_{i=1}^n A_i - \sum_{1\le i<j\le n} \mathbb{P}(A_i \cap A_j) + \sum_{1\le i<j<k\le n} \mathbb{P}(A_i \cap A_j \cap A_k) - \cdots \pm \mathbb{P}(A_1 \cap \cdots \cap A_n).$$

Hint: Expand $1 - \prod_{i=1}^n (1 - \mathbb{1}_{A_i})$, *and take expectation of both sides.*

Solution First, observe that $\mathbb{1}_{A^c} = 1 - \mathbb{1}_A$. So

$$1 - \prod_{i=1}^n (1 - \mathbb{1}_{A_i}) = 1 - \prod_{i=1}^n \mathbb{1}_{A_i^c} = 1 - \mathbb{1}_{A_1^c \cap \cdots \cap A_n^c} = \mathbb{1}_{A_1 \cup \cdots \cup A_n}.$$

On the other hand,

$$1 - \prod_{i=1}^n (1 - \mathbb{1}_{A_i}) = \sum_{i=1}^n \mathbb{1}_{A_i} - \sum_{1\le i<j\le n} \mathbb{1}_{A_i \cap A_j} + \cdots \pm \mathbb{1}_{A_1 \cap \cdots \cap A_n}.$$

So we have

$$\mathbb{1}_{A_1 \cup \cdots \cup A_n} = \sum_{i=1}^n \mathbb{1}_{A_i} - \sum_{1\le i<j\le n} \mathbb{1}_{A_i \cap A_j} + \cdots \pm \mathbb{1}_{A_1 \cap \cdots \cap A_n}.$$

By applying the expectation to both sides we will get

$$\mathbb{P}(A_1 \cup \cdots \cup A_n) = \sum_{i=1}^n A_i - \sum_{1\le i<j\le n} \mathbb{P}(A_i \cap A_j) + \sum_{1\le i<j<k\le n} \mathbb{P}(A_i \cap A_j \cap A_k) - \cdots \pm \mathbb{P}(A_1 \cap \cdots \cap A_n).$$

■ **Problem 1.102 — From Rosenthal.** Let $f(x) = ax^2 + bx + c$ be a second-degree polynomial function where $a, b, c \in \mathbb{R}$.

(a) Find necessary and sufficient condition on a, b and c such that the equation $\mathbb{E}\Big[f(\alpha x)\Big] = \alpha^2 \mathbb{E}\Big[f(X)\Big]$ holds for all $\alpha \in \mathbb{R}$ and all random variable X.

(b) Find necessary and sufficient condition on a, b and c such that the equation $\mathbb{E}\Big[f(x - \beta)\Big] = \mathbb{E}\Big[f(x)\Big]$ holds of all $\beta \in \mathbb{R}$ and all random variable X.

(c) Do parts (a) and (b) account for the properties of the variance function? Why or why not?

Solution (a) For the LHS we have

$$\mathbb{E}\Big[\alpha^2 a x^2 + \alpha b x + c\Big] = a\alpha^2\mathbb{E}\Big[x^2\Big] + b\alpha\mathbb{E}\Big[x\Big] + c.$$

And for the RHS we have

$$\alpha^2\mathbb{E}\Big[f(x)\Big] = a\alpha^2\mathbb{E}\Big[x^2\Big] + b\alpha^2\mathbb{E}\Big[x\Big] + c\alpha^2.$$

Thus the necessary and sufficient condition for the equality to hold for every $\alpha \in \mathbb{R}$ is to have

$$b = 0, \qquad c = 0.$$

(b) For the LHS we have

$$\mathbb{E}\Big[f(x-\beta)\Big] = a\mathbb{E}\Big[x^2\Big] + (b - 2a\beta)\mathbb{E}\Big[x\Big] + a\beta^2 - b\beta + c.$$

For the RHS we have

$$\mathbb{E}\Big[f(x)\Big] = a\mathbb{E}\Big[x^2\Big] + b\mathbb{E}\Big[x\Big] + c.$$

This implies that we need to have

$$a = 0, \qquad b = 0.$$

I.e. the polynomial should be a constant polynomial.

(c) For the property $\mathrm{Var}\,(\alpha C) = \alpha^2\,\mathrm{Var}\,(C)$, it follows from the fact that in part (b) we found that for the polynomial we need to have $b = 0$ and $c = 0$. However, part (b) does not account for the property of variance that $\mathrm{Var}(X + b) = \mathrm{Var}(X)$. Because in the case of Var the constants of the polynomial depends on the random variable under consideration. I.e. we have

$$b = -2\mathbb{E}\Big[X\Big], \qquad c = \mathbb{E}\Big[X\Big]^2.$$

■ **Problem 1.103 — From Rosenthal.** Let $X_1, X_2, \cdots$ be independent, each with mean μ and variance σ^2, and let N be an integer-valued random variable with mean m and variance v, with N independent of all the X_i. Let $S = X_1 + X_2 + \cdots + X_n = \sum_{i=1}^{\infty} X_i \mathbb{1}_{N\geq i}$. Compute $\mathrm{Var}(S)$ and $\mathbb{E}\Big[S\Big]$.

Solution — Using the conditional expectation. For this problem we can either use the conditional expectation or use the first principles.

$$\mathbb{E}\Big[S\Big] = \mathbb{E}\Big[X_1 + \cdots + X_N\Big] = \sum_n \mathbb{E}\Big[X_1 + \cdots + X_n\Big]\mathbb{P}(N = n) = \mu\sum_n n\mathbb{P}(N = n) = \mu m.$$

Similarly we have

$$\begin{aligned}
\mathbb{E}\Big[S^2\Big] &= \sum_n \mathbb{E}\Big[S^2 \mid N = n\Big]\mathbb{P}(N = n) \\
&= \sum_n \mathbb{E}\Big[(X_1 + \cdots + X_n)^2\Big]\mathbb{P}(N = n) \\
&= \sum_n \mathbb{E}\Big[\sum_i X_i^2 + \sum_{i<j} X_i X_j\Big]\mathbb{P}(N = n) \\
&= \sum_n (\sum_i \mathbb{E}\Big[X_i^2\Big] + 2\sum_{i<j}\mathbb{E}\Big[X_i X_j\Big])\mathbb{P}(N = n) \\
&= \sum_n ((\sigma^2 + \mu^2)n + n(n-1)\mu^2)\mathbb{P}(N = n) \\
&= (\sigma^2 + \mu^2)m + \mu^2(v + m^2 - m).
\end{aligned}$$

Where we have used $\mathbb{E}\Big[X_i^2\Big] = \mathrm{Var}(X_i) + \mathbb{E}\Big[X_i\Big]^2 = \mu^2 + \sigma^2$, and also used the fact that the sum $\sum_{i<j}$ has $\binom{n}{2}$ terms. Now we can compute the variance

$$\mathrm{Var}(S) = \mathbb{E}\Big[S^2\Big] - \mathbb{E}\Big[S\Big]^2 = (\sigma^2 + \mu^2)m + \mu^2(v + m^2 - m) - \mu^2 m^2 = \sigma^2 m + \mu^2 v.$$

■ **Problem 1.104 — from Rosenthal.** Let X, Z be independent random variables each with standard normal distribution. Let $a, b \in \mathbb{R}$ (not both 0), and let $Y = aX + bZ$.

(a) Compute $\mathrm{Corr}(X, Y)$.

(b) Show that $|\,\mathrm{Corr}(X, Y)| \le 1$.

(c) Given necessary and sufficient conditions on the values of a and b such that $\mathrm{Corr}(X, Y) = 1$.

(d) Given necessary and sufficient conditions on the values of a and b such that $\mathrm{Corr}(X, Y) = -1$.

Solution (a) First, observe that

$$\mathbb{E}\Big[X\Big] = \mathbb{E}\Big[Z\Big] = 0, \quad \mathrm{Var}(X) = \mathrm{Var}(Y) = 1, \quad \mathbb{E}\Big[Y\Big] = a + b, \quad \mathrm{Var}(Y) = a^2 + b^2$$

where we have used the properties of variance to compute $\mathrm{Var}(Y)$. By the definition of correlation we have

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} = \frac{\mathbb{E}\Big[XY\Big] - \mathbb{E}\Big[X\Big]\mathbb{E}\Big[Y\Big]}{\sqrt{a^2 + b^2}} = \frac{a}{\sqrt{a^2 + b^2}}.$$

(b) It follows immediately by

$$|\,\mathrm{Corr}(X, Y)| = |\frac{1}{\sqrt{1 + b^2/a^2}}| \le 1.$$

(c) From our solution in part (a), the necessary and sufficient condition for $\mathrm{Corr}(X, Y) = 1$ is that $b = 0$ and $a > 0$.

(d) From the solution in part (a). the necessary and sufficient condition for $\mathrm{Corr}(X, Y) = -1$ is that $b = 0$ and $a < 0$.

■ **Problem 1.105 — From Rosenthal.** Let X, Y be independent general non-negative random variables, and let $X_n = \Psi_n(X)$, where $\Psi_n(x) = \min(n, 2^{-n}\lfloor 2^n x \rfloor)$.

(a) Give an example of a sequence of functions $\Phi_n : [0, \infty) \to [0, \infty)$ other than $\Phi_n = \Psi_n$, such that for all x we have $0 \le \Phi_n(x) \le x$ and $\{\Phi_n(x)\} \nearrow x$ as $n \to \infty$.

(b) Suppose that $Y_n = \Phi_n(Y)$ with Φ_n as in part (a). Must X_n and Y_n be independent.

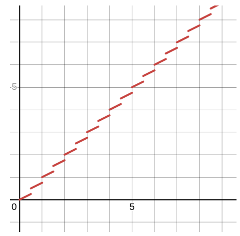
(c) Suppose $\{Y_n\}$ is an arbitrary collection of non-negative random variables such that $\{Y_n\} \nearrow Y$. Must X_n and Y_n be independent?

(d) Under the assumption of part (c), determine which quantities in 4.2.7 are necessarily equal?

Solution (a) Define

$$f_n(x) = \min\{n, \frac{1}{2^n}(\frac{1}{2}(2^n x + \lfloor 2^n x \rfloor))\}.$$

The graph of this function will be as follows.



(b) First observe that Φ_n is a Borel measurable function as the pre-image of any open interval is a union of half open intervals or the empty set. Since X, Y are independent, then X_n, Y_n are also independent.

(c) No. We demonstrate a counterexample. Let

$$Y_n = \max\{\Psi_n(Y) - \frac{1}{n^2}\Psi_n(X), 0\}.$$

(d) Since $\{X_n\} \nearrow X$ and $\{Y_n\} \nearrow Y$, then $\{X_nY_n\} \nearrow XY$. Thus by the monotone convergence theorem we have

$$\mathbb{E}\Big[XY\Big] = \lim_n \mathbb{E}\Big[X_nY_n\Big].$$

Also since (by monotone convergence theorem) $\mathbb{E}\Big[X\Big] = \lim_n \mathbb{E}\Big[X_n\Big]$ and $\mathbb{E}\Big[Y\Big] = \lim_n \mathbb{E}\Big[Y_n\Big]$ then by the limit laws we have

$$\lim_n \mathbb{E}\Big[X_n\Big]\mathbb{E}\Big[Y_n\Big] = \mathbb{E}\Big[X\Big]\mathbb{E}\Big[Y\Big].$$

■ **Problem 1.106 — From Rosenthal.** Give examples of a random variable X defined on Lebesgue measure on $[0, 1]$, such that

(a) $\mathbb{E}\Big[X^+\Big] = \infty$ and $0 < \mathbb{E}\Big[X^-\Big] < \infty$.

(b) $\mathbb{E}\Big[X^-\Big] = \infty$ and $0 < \mathbb{E}\Big[X^+\Big] < \infty$.

(c) $\mathbb{E}\Big[X^+\Big] = \mathbb{E}\Big[X^-\Big] = \infty$.

(d) $0 < \mathbb{E}\Big[X\Big] < \infty$ but $\mathbb{E}\Big[X^2\Big] = \infty$.

Solution (a) Define

$$X^+ = 2 \cdot \mathbb{1}_{(1/2,3/4)} + \sum_{n=2}^{\infty} 2^n \cdot \mathbb{1}_{(2^{-n},2^{-n+1})}, \qquad X^- = -2 \cdot \mathbb{1}_{(3/4,1)}.$$

Then $\mathbb{E}\Big[X^+\Big] = 1/2 + 1 + 1 + \cdots = \infty, \qquad \mathbb{E}\Big[X^-\Big] = 1/2.$

(b) Similar to part (a) but exchange X^- and X^+.

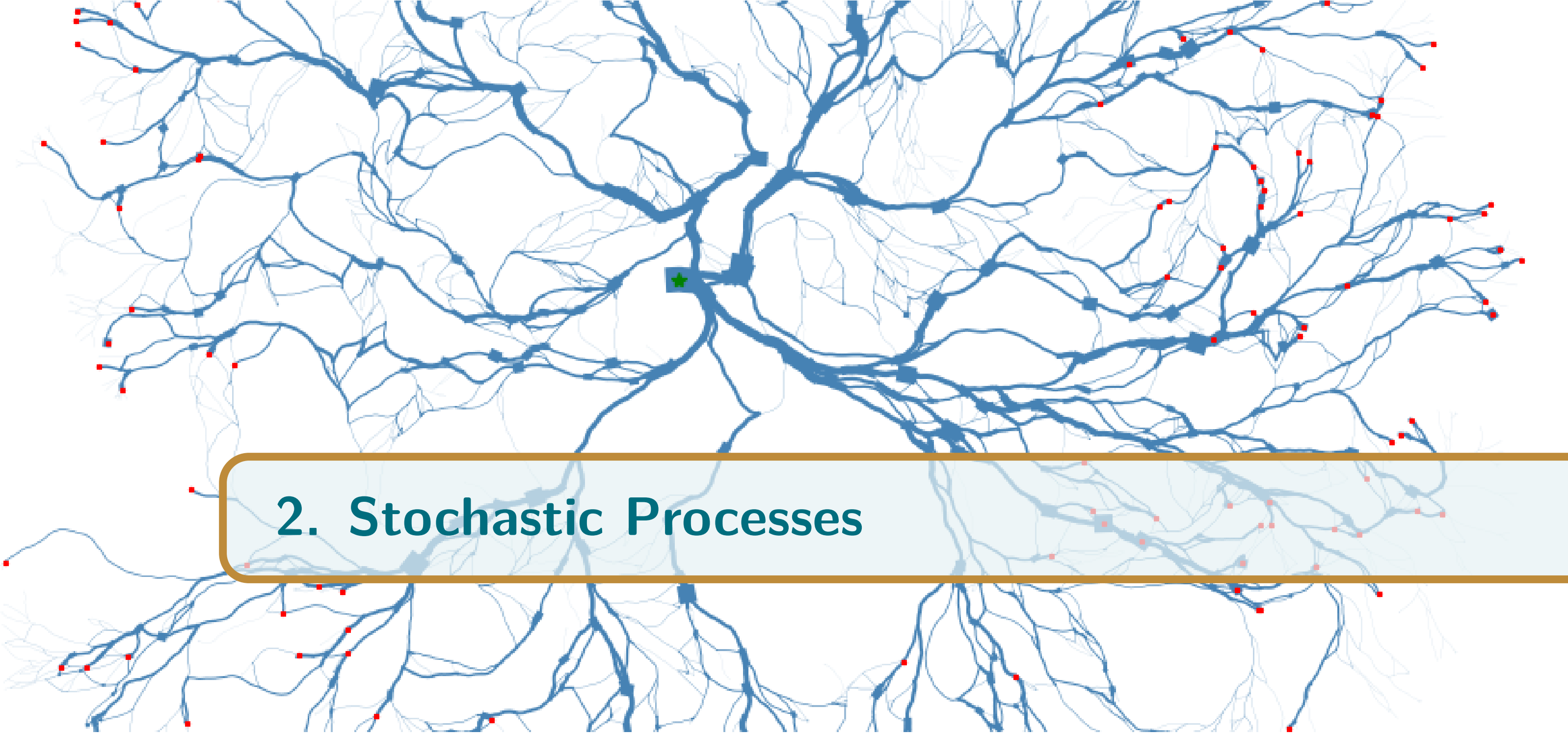
(c) Define

$$X^+ = \sum_{n \text{ even}} 2^n \cdot \mathbb{1}_{(2^{-n},2^{-n+1})}, \qquad X^+ = \sum_{n \text{ odd}} 2^n \cdot \mathbb{1}_{(2^{-n},2^{-n+1})}.$$

Clearly

$$\mathbb{E}\Big[X^+\Big] = \mathbb{E}\Big[X^-\Big] = \infty.$$

(d) TODO: TO BE ADDED.



2. Stochastic Processes

The question in this chapter are mainly from the book by Ross.

■ **Problem 2.1** An urn always contains 2 balls. Ball colors are red and blue. At each stage a ball is randomly chosen and then replaced by a new ball, which with probability 0.8 is the same color, and with probability 0.2 is the opposite color, as the ball it replaces. If initially both balls are red, find the probability that the fifth ball selected is red. [This question is from Ross]

Solution First, we need to translate this problem to a suitable Markov chain. There are many ways we can do so, each with its own pros and cons. The difference between all of these formulations come down to our choice for the state space (i.e. the co-domain of the random variable). For instance, we can assume that the state space is $S = \{RR, RB, BB\}$ that is the content of the Urn, or we can simply say that the state space is $S = \{0, 1, 2\}$ that is the number of red ball inside the Urn. Since these two sets are isomorphic (as there is a bijection between these two sets), but the actual choice depends on personal preference. Let's proceed with $S = \{0, 1, 2\}$. Then, we need to determine the transition matrix. We can do so by doing the first step argument. We start with $P(0, 0)$.

$$P(0,0) = \mathbb{P}(X_1 = 0|X_0 = 0) = \mathbb{P}(X_1 = 0|X_0 = 0, E_R)\underbrace{\mathbb{P}(E_R|X_0 = 0)}_{0} + \underbrace{\mathbb{P}(X_1 = 0|X_0 = 0, E_B)}_{0.8}\underbrace{\mathbb{P}(E_B|X_0 = 0)}_{1},$$

where E_R is the event at which a red balls is drawn from the Urn, while E_B is the event where a blue ball is drawn. The reason behind the values for the term above are very straight forward. For instance $\mathbb{P}(E_R|X_0 = 0) = 0$ because given the fact that number of red balls in the Urn is zero ($X_0 = 0$), then the probability that we draw a red ball is zero (as there is not red balls in the Urn). For the term $\mathbb{P}(X_1 = 0|X_0 = 0, E_B) = 0.8$, because given there is no red balls inside the urn, and also given the fact that the drawn ball is blue, the probability of ending up at the state $X_1 = 0$ (i.e. still no red balls) is that probability is that we replaced the drawn ball with a blue ball (same color) which has the probability 0.8. Similarly, we can calculate the first step transition probabilities.

$$P(0,1) = \mathbb{P}(X_1 = 1|X_0 = 0) = \mathbb{P}_0(X_1 = 1) = \mathbb{P}_0(X_1 = 1|E_R)\underbrace{\mathbb{P}_0(E_R)}_{0} + \underbrace{\mathbb{P}_0(X_1 = 1|E_B)}_{0.2}\underbrace{\mathbb{P}_0(E_B)}_{1} = 0.2,$$

$$P(0,2) = \mathbb{P}(X_1 = 2|X_0 = 0) = \mathbb{P}_0(X_1 = 2) = \mathbb{P}_0(X_1 = 2|E_R)\underbrace{\mathbb{P}_0(E_R)}_{0} + \underbrace{\mathbb{P}_0(X_1 = 2|E_B)}_{0}\underbrace{\mathbb{P}_0(E_B)}_{1} = 0,$$

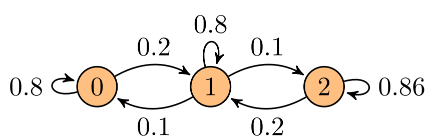
$$P(1,0) = \mathbb{P}(X_1 = 0|X_0 = 1) = \mathbb{P}_1(X_1 = 0) = \underbrace{\mathbb{P}_1(X_1 = 0|E_R)}_{0.2}\underbrace{\mathbb{P}_1(E_R)}_{0.5} + \underbrace{\mathbb{P}_1(X_1 = 0|E_B)}_{0}\underbrace{\mathbb{P}_1(E_B)}_{0.5} = 0.1.$$

$$P(1,1) = \mathbb{P}(X_1 = 1|X_0 = 1) = \mathbb{P}_1(X_1 = 1) = \underbrace{\mathbb{P}_1(X_1 = 1|E_R)}_{0.8}\underbrace{\mathbb{P}_1(E_R)}_{0.5} + \underbrace{\mathbb{P}_1(X_1 = 1|E_B)}_{0.8}\underbrace{\mathbb{P}_1(E_B)}_{0.5} = 0.8.$$

and so on. Then we will have the following transition matrix for this problem.

$$M = \begin{pmatrix} 0.8 & 0.2 & 0 \\ 0.1 & 0.8 & 0.1 \\ 0 & 0.2 & 0.8 \end{pmatrix}$$

with the following graph



Now, we need to compute the probability that the fifth ball drawn is red. This means that we have already drawn four balls, and now we want to draw the fifth one. So, we need to consider the 4 step transition matrix, i.e. M^4. Then

$$M^4 = \begin{pmatrix} 0.4872 & 0.4352 & 0.0776 \\ 0.2176 & 0.5648 & 0.2176 \\ 0.0776 & 0.4352 & 0.4872 \end{pmatrix}$$

Given that we have started with 2 red balls, then the probability of finding the Urn with 0 red balls is 0.0776, with 1 red ball is 0.4352, and with 2 red balls is 0.4872. So the probability that the next drawn balls is red is

$$\mathbb{P}(E_R) = \underbrace{\mathbb{P}(E_R|X_4 = 0)}_{0}\underbrace{\mathbb{P}(X_4 = 0)}_{0.0776} + \underbrace{\mathbb{P}(E_R|X_4 = 1)}_{0.5}\underbrace{\mathbb{P}(X_4 = 1)}_{0.4352} + \underbrace{\mathbb{P}(E_R|X_4 = 2)}_{1}\underbrace{\mathbb{P}(X_4 = 2)}_{0.4872} = 0.7048.$$

■ **Problem 2.2 — Turning non-Markov processes to Markov-chain.** Suppose that whether or not it rains today depends on previous weather conditions through the last two days. Specifically, suppose that if it has rained for the past two days, then it will rain tomorrow with probability 0.7; if it rained today but not yesterday, then it will rain tomorrow with probability 0.5; if it rained yesterday but not today, then it will rain tomorrow with probability 0.4; if it has not rained in the past two days, then it will rain tomorrow with probability 0.2. [This question is from Ross]. Given that it rained on Monday and Tuesday, what is the probability that it will rain on Thursday?

Solution This random process is not a Markov chain, the value of the random variable at the next state, depends on two previous states. However, we can turn this into a Markov chain. Define the following states

RR: Rained yesterday and today.
$R\overline{R}$: Rained yesterday, but not today.
$\overline{R}R$: Not rained yesterday, but rained today.
$\overline{RR}$: Not rained yesterday and today.

Suppose that we are at state RR. Suppose that it rained yesterday and also today. Thus we are at state RR. If it rains tomorrow, then we will be still at state RR. That is because, That is because the yesterday of tomorrow is today! So if it rains tomorrow, since today (yesterday of tomorrow) was also rainy, thus if it rains tomorrow then we will stay at state RR. If it does not rain tomorrow, then we will get to state $\overline{R}R$. The following matrix is the transition matrix for this Markov chain

$$M = \begin{pmatrix} 0,7 & 0.3 & 0 & 0 \\ 0 & 0 & 0.4 & 0.6 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.2 & 0.8 \end{pmatrix}$$

Now, to calculate the probability of raining on Thursday, given it rained on Monday and Tuesday, we first need to calculate the two step transition probability.

$$M^2 = \begin{pmatrix} \boxed{0.49} & 0.21 & \boxed{0.12} & 0.18 \\ 0.2 & 0.2 & 0.12 & 0.48 \\ 0.35 & 0.15 & 0.2 & 0.3 \\ 0.1 & 0.1 & 0.16 & 0.64 \end{pmatrix}$$

The probability to rain on Thursday is the sum of the boxed elements in the matrix above. So the desired probability is

$$p = 0.61.$$

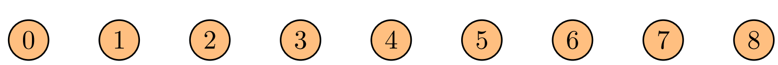
■ **Problem 2.3** Suppose that balls are successively distributed among 8 urns, with each ball being equally likely to be put in any of these urns. What is the probability that there will be exactly 3 nonempty urns after 9 balls have been distributed? [Question from Ross]

Solution Before going through the solution, it might be more informative to explicitly write down what is the sample space Ω. At each time step, we basically throwing a 8 sided dice, and then put a ball at the urn number i if the output of the dice is i. So, each time we repeat the experiment, we will get a sequence of number each of which is one of $1, 2, \cdots, 8$. So the sample space will be the set of all sequences consisting of number $1, \cdots, 8$.

$$\Omega = \{21342\ldots, 44513\ldots, 11234\ldots, 88432\ldots, \cdots\}.$$

So, the outputs of the throwing dice at different steps are independent and identically distributed random variables. I.e. for a fixed $\omega \in \Omega$, The t-th element of the sequence is a random variable Y_t and all of the random variables $\{Y_t\}_t$ are independent and identically distributed. Note that the sample space associated with these random variables is $\{1, 2, 3, 4, 5, 6, 7, 8\}$ (i.e. the sample space of a 8 sided dice experiment).w

Let the random variable X_n be the number of filled (non-empty) urns at step n. So the state space will be $S = \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$, which is represented in the following graph.



This picture is not yet complete and we need to include the transition probabilities. We will do so by the first step argument. First, observe that $P(0, 0) = 0$, because if we start with all of the urns empty, then after one step, we have put a ball somewhere, thus it is impossible to end up with zero filled urn. Similarly, $P(8,8)=1$, that is because if all of the urns are filled, then adding any new ball somewhere to any of the urns will keep the number of filled urns at 8. Then for $X_0=n$, i.e. starting with n filled urns, we have

$$\mathbb{P}_n(X_1=n-1)=0.$$

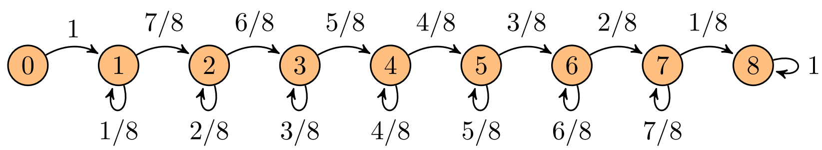
That is because starting with n filled urns, after doing one step, it is not possible to have less urns filled. I.e. after each step, we can either end up with more filled urns or the same number of filled urns. For $P(n,n)$, define the event E be the event of putting the ball in any of the filled urns. Thus E^c will be the probability of putting the ball at one of the empty urns.

$$\mathbb{P}_n(X_1=n)=\underbrace{\mathbb{P}_n(X_1=n|E)}_{1}\underbrace{\mathbb{P}_n(E)}_{n/8}+\underbrace{\mathbb{P}_n(X_1=n|E^c)}_{0}\underbrace{\mathbb{P}_n(E^c)}_{(8-n)/8}=\frac{n}{8}.$$

Now for $\mathbb{P}(n,n+1)$ we can write

$$\mathbb{P}_n(X_1=n+1)=\underbrace{\mathbb{P}_n(X_1=n+1|E)}_{0}\underbrace{\mathbb{P}_n(E)}_{n/8}+\underbrace{\mathbb{P}_n(X_1=n+1|E^c)}_{1}\underbrace{\mathbb{P}_n(E^c)}_{(8-n)/8}=1-\frac{n}{8}.$$

Thus the completed graph will be



The corresponding transition matrix will be

$$M=\begin{pmatrix}0&1&0&0&0&0&0&0&0\\0&1/8&7/8&0&0&0&0&0&0\\0&0&1/8&6/8&0&0&0&0&0\\0&0&0&3/8&5/8&0&0&0&0\\0&0&0&0&4/8&4/8&0&0&0\\0&0&0&0&0&5/8&3/8&0&0\\0&0&0&0&0&0&6/8&2/8&0\\0&0&0&0&0&0&0&7/8&1/8\\0&0&0&0&0&0&0&0&1\end{pmatrix}$$

The probability that after 9 steps, there are exactly three empty urns is $(M^9)_{(0,3)}$, which is

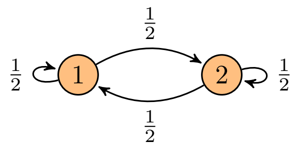
$$p=(M^9)_{(0,3)}\approx 0.007572.$$

■ **Problem 2.4** It is a good practice to derive the value of the transition probability of a simple Markov chain using the first principles. Consider the Markov chain representing a lamp that turns on with probability 1/2 and turns off with probability 1/2, and stays at the old state with probability 1/2. Thus we will have the following diagram for this Markov chain.
In this example, the state space is $S=\{0,1\}$, and the sample space is

$$\Omega=\{(x_1,x_2,\cdots):x_i\in S\}$$

which is basically the set of all sequences of one's and zero's. Given this, the random variables $(X_n)_n$ defined t be

$$X_n(\omega)=x_n,$$



where $\omega \in \Omega$ and x_n is the n-th letter in ω. Intuitively speaking, we know that

$$P(1,0) = \mathbb{P}(X_{n+1} = 1 | X_n = 0) = \frac{1}{2}.$$

However, here we want to derive that number more explicitly by working directly with the elements of the probability space. First, we need to determine the event associated with $X_{n+1} = 1$. This is the event that has elements where the $n+1$-th position is 1. I.e.

$$E = \{(x_1, x_2, \cdots, x_n, 1, x_{n+2}, \cdots) : x_i \in S\}.$$

Similarly, we have

$$F = \{(x_1, x_2, \cdots, x_{n-1}, 0, x_{n+1}, \cdots) : x_i \in S\}.$$

So we have

$$\mathbb{P}(X_{n+1} = 1 | X_n = 0) = \mathbb{P}(E|F) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)} = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F \cap E) + \mathbb{P}(F \cap E^c)} = \frac{\frac{1}{|\Omega|}}{\frac{1}{|\Omega|} + \frac{1}{|\Omega|}} = \frac{1}{2}.$$

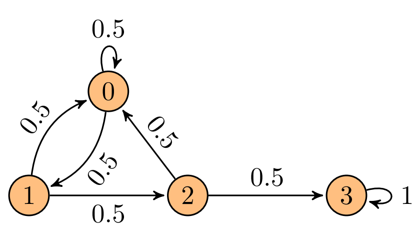
Note that $\mathbb{P}(E \cap F) = \frac{1}{|\Omega|}$, since out of many combinations of the sequence of zeros and ones, there is one one sequence whose n-th place is 0 and $n+1$-th place is 1. Furthermore, $\mathbb{P}(F \cap E^c) = \frac{1}{|\Omega|}$ as there is only one string where its n-th and $(n+1)$-th string are both zero.

■ **Problem 2.5** In a sequence of independent flips of a fair coin, let N denote the number of flips until there is a run of three consecutive heads. Find

(a) $\mathbb{P}(N \leq 8)$,

(b) $\mathbb{P}(N = 8)$.

Solution Let X_n denote the number of consecutive heads at step n. For instance for the outcome $\omega \in \Omega$ where $\omega = HTTHTTHHHTTHT\ldots$, $X_2(\omega) = 0$ since the second symbol is T thus there is no consecutive heads. But $X_4(\omega) = 1$, as there is one consecutive heads at step 4. Lastly $X_9(\omega) = 3$, since there is three consecutive heads at step 9. This Markov chain will have the following transition diagram.



The transition probabilities are simply computed by the first step argument. For instance, for $P(0,1)$ we have

$$\mathbb{P}_0(X_1 = 1) = \underbrace{\mathbb{P}_0(X_1 = 1|H)}_{1}\underbrace{\mathbb{P}_0(H)}_{1/2} + \underbrace{\mathbb{P}_0(X_1 = 1|T)}_{0}\underbrace{\mathbb{P}_0(T)}_{1/2},$$

where H is the event that the flipped coin is heads and $H^c = T$. The transition matrix for this Markov chain will be

$$M = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0.5 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

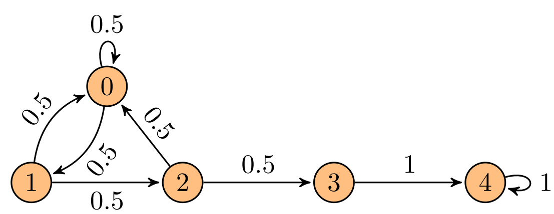
Since the state 3 is an absorbing state, then if we get there we will be there for the rest of our life! Thus the probability that the random walker has got there for $N \leq 8$ is simply $(M^8)_{(0,3)}$. Then

$$\mathbb{P}(N \leq 8) = 0.4180.$$

Now for part (b), the probability that the random walker has arrived at the state 3 right at the step 8, is

$$\mathbb{P}(N = 8) = \mathbb{P}(N \leq 8) - \mathbb{P}(N \leq 7) = 0.0508.$$

There is yet another approach that we can compute the probability $\mathbb{P}(N = 8)$. To do this, we need to consider 4 states $S = \{0, 1, 2, 3, 4\}$ where the state 4 is of when 3 consecutive heads has occurred at the past. So when the random walker enters the state 3 at some time, it moves to the state 4 at the next time and remains there forever. The state diagram will be



Then the transition matrix will be

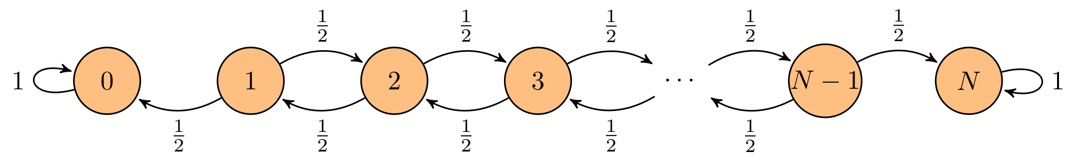
$$M = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0.5 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Then the probability $\mathbb{P}(N = 8) = (M^8)_{(0,3)} = 0.05080$.

■ **Problem 2.6 — Gambler's Ruin.** Suppose Alice and Bob have in total of N coins. Alice and Bob play a game with a fair coin. When Alice wins, gets a coin from Bop, and vise versa. What is the probability that Alice wins if she starts with $0 \leq a \leq N$ coins.

Solution There are many ways to tackle a probability problem like this and the solution presented here is not the only way to find the solution to this problem. We want to model this with Markov chain whose state space is $\{0, 1, 2, \cdots, N\}$. Thus X_n represents the fortune of Alice after playing the games for n times.
Let p_a be the probability of Alice wining if she starts with a coins. First, observe that $p_0 = 0$ and $p_N = 1$. Let E denote that event of Alice wining the whole game. Also, let F_1 be the event in which she looses the first game and F_2 the event in which she wins the first game. Then

$$p_a = \mathbb{P}_a(E) = \underbrace{\mathbb{P}_a(E|F_1)}_{\mathbb{P}(E|F_1, X_0=a)} \mathbb{P}(F_1) + \underbrace{\mathbb{P}_a(E|F_1^c)}_{\mathbb{P}(E|F_1^c, X_0=a)} \mathbb{P}(F_1^c)$$



(note that this identity is actually true for any set F_1, but here F_1 is the specific event explained above). The probability that she looses or wins the first game is $\frac{1}{2}$. Also, observe that $\mathbb{P}_a(E|F_1) = p_{a+1}$ (since if she wins the first game she will have one more coin) and $\mathbb{P}_a(E|F_1^c) = p_{a-1}$. Thus

$$p_a = \frac{1}{2}p_{a+1} + \frac{1}{2}p_{a-1}.$$

Now we can solve this recurrent equation with the characterization polynomial which is $2 = X + 1/X$ or $X^2 - 2X + 1 = (X-1)^2 = 0$. Thus the characteristic polynomial has a double root. Thus

$$p_a = (Aa + B)(1)^a = Aa + B.$$

Since $p_0 = 0,\ p_N = 1$, then it turns out that

$$p_a = \frac{a}{N}.$$

■ **Problem 2.7 — Gambler's Ruin with Draw.** Let Alice and Bob play Rock-Paper-Scissors. If Alice and Bob has a total of N coins, and at each play, the winner gets one coin from the loser, what is the probability that Alice will win the game if he starts with a coins. When they draw, then they repeat the game (or equivalently, they play another game without any coins exchange).

Solution We need to do a first step analysis similar to what we did before. Let E be the event that Alice wins the whole game, and the event $F = F_{-1} \cup F_0 \cup F_1$ where

F_{-1}: Alice loses the first game,
F_0: Alice draws the first game,
F_1: Alice wins the first game.

It is clear that $\mathbb{P}(F) = 1$, since the components are mutually disjoint. Thus $E \cap F_{-1},\ E \cap F_0,\ E \cap F_1$ are also mutually disjoints where. Thus we can write

$$\mathbb{P}_a(E) = \mathbb{P}_a(E \cap F_{-1}) + \mathbb{P}_a(E \cap F_0) + \mathbb{P}_a(E \cap F_1) = \mathbb{P}_a(E|F_{-1})\mathbb{P}_a(F_{-1}) + \mathbb{P}_a(E|F_0)\mathbb{P}_a(F_0) + \mathbb{P}_a(E|F_1)\mathbb{P}_a(F_1).$$

Since the game is fair we know

$$\mathbb{P}_a(F_{-1}) = \mathbb{P}_a(F_0) = \mathbb{P}_a(F_1) = \frac{1}{3}.$$

Furthermore, we know

$$\mathbb{P}_a(E|F_{-1}) = p_{a-1}, \qquad \mathbb{P}_a(E|F_0) = p_a, \qquad \mathbb{P}_a(E|F_1) = p_{a+1}.$$

Thus the first step analysis will lead to the following identity.

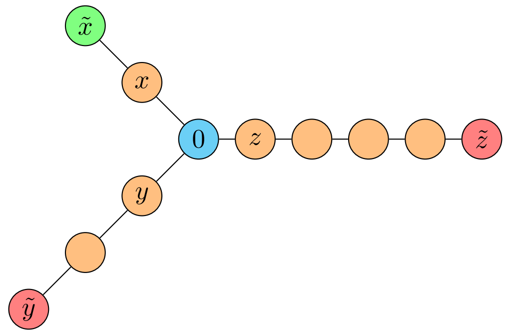
$$\mathbb{P}_a(E) = p_a = \frac{1}{3}(p_{a-1} + p_a + p_{a+1}),$$

which after simplification becomes

$$2p_a = p_{a-1} + p_{a+1},$$

which is the same recursive formula we got in the previous example. So the possibility of the draw, will not change the behaviour of the system.

■ **Problem 2.8** Consider the a simple random walker on the following graph. Let $B = \{T_{\tilde{x}} < T_{\{\tilde{z},\tilde{y}\}}\}$. Compute the probability $\mathbb{P}_0(B)$.



Solution This problem is simply asking what is the probability that we hit $\tilde{x}$ state before hitting any of $\tilde{y}$ or $\tilde{z}$ states, given the fact that the random walker starts from the state 0. To keep unnecessary details out of the way, we have only labeled the vertices that we will use in our analysis. We will have the following notation to simplify the solution

$$p_v = \mathbb{P}_v(B),$$

where v is any vertex in the graph. Note that starting at 0, i.e. $X_0 = 0$, then going to any of the states x, y, or z, are mutually disjoint events, and the probability of the union of these events is one. With our first time step analysis (see Proposition 5.2) we can write

$$\mathbb{P}_0(B) = \frac{1}{3}(p_x + p_y + p_z).$$

Now we need to analyze each of terms in the RHS. Let's start with p_z. Consider two events $\{T_0 < T_{\tilde{z}}\}$ and $\{T_0 > T_{\tilde{z}}\}$, where the first time is the event where the random walker hits the 0 state before hitting the $\tilde{z}$ step first, and the second one is the vice versa. These two events are disjoint and the probability of the union is 1. Thus we write the conditional expansion of p_z based on these events

$$p_z = \mathbb{P}_z(B) = \mathbb{P}_z(B|T_0 < T_{\tilde{z}})\mathbb{P}_z(T_0 < T_{\tilde{z}}) + \mathbb{P}_z(B|T_0 > T_{\tilde{z}})\mathbb{P}_z(T_0 > T_{\tilde{z}}).$$

We know that $\mathbb{P}_z(B|T_0 > T_{\tilde{z}}) = \mathbb{P}(B|X_0 = z, X_i = \tilde{z})$ for some $i > 0$. From Markov property it follows that

$$\mathbb{P}(B|X_0 = z, X_i = \tilde{z}) = \mathbb{P}(B|X_i = \tilde{z}) = \mathbb{P}(B|X_0 = \tilde{z}) = p_{\tilde{z}}.$$

Also $\mathbb{P}_z(B|T_0 < T_{\tilde{z}}) = \mathbb{P}_0(B) = p_0$ by the Markov property. Lastly, $\mathbb{P}_z(T_0 < T_{\tilde{z}})$ is determined by the Gambler's ruin method we say before, which is basically

$$\mathbb{P}_z(T_0 < T_{\tilde{z}}) = \frac{5}{4}, \qquad \mathbb{P}_z(T_0 > T_{\tilde{z}}) = \frac{1}{5}.$$

By doing the same kind of analysis for p_x as well as p_y we will get

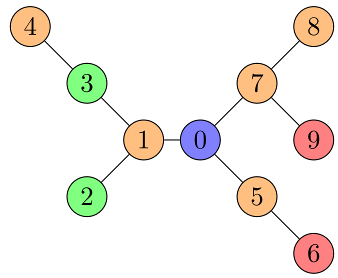
$$p_z = \frac{4}{5}p_0, \qquad p_y = \frac{2}{3}p_0, \qquad p_x = \frac{1}{2}p_0 + \frac{1}{2}.$$

Now by substituting in the identity we got from the first time step argument, we can fine that

$$p_0 = \frac{15}{31},$$

And this completes our solution for the problem.

■ **Problem 2.9** Consider the graph $\gamma = (V, E)$ drawn below. Set $Z = \{2, 3\}$, and $W = \{6, 9\}$. Compute $\mathbb{P}_0(T_Z < T_W)$. In colors: we start at blue, win if we reach green, and lose of we reach red.



Solution As always, we start with our powerful tool in hand, which is the first step argument (which is basically a special form of the more general conditional expansion). We start with first step argument at state 0. We will get

$$\mathbb{P}_0(B) = \frac{1}{3}(\mathbb{P}_1(B) + \mathbb{P}_7(B) + \mathbb{P}_5(B)),$$

and now we need to analyze each of the terms in the right hand side. We start with $\mathbb{P}_5(B)$ which is the most straight forward one. As we saw in the last example, we can analyze this state with a conditional expansion on the two disjoint events, whose union probability is 1. Let those two events be $\{T_6 < T_0\}$ (where the random walker hits the state 6 before hitting the state 0), and $\{T_6 > T_0\}$, where the random walker hits the state 0 before hitting the state 6. Thus the expansion will be

$$\mathbb{P}_5(B) = \mathbb{P}_5(B|T_6 < T_0)\mathbb{P}_5(T_6 < T_0) + \mathbb{P}_5(B|T_6 > T_0)\mathbb{P}_5(T_6 > T_0).$$

We know that if we hit the state 6 before 0, we have no chance to hit any of the green states (we will lose). Thus

$$\mathbb{P}_5(B|T_6 < T_0) = 0.$$

And from the Gambler's ruin we know that $\mathbb{P}_5(T_6 > T_0) = 1/2$, and from the Markov property we know that $\mathbb{P}_5(B|T_6 > T_0) = \mathbb{P}_0(B)$, because the conditional probability $\mathbb{P}_5(B|T_6 > T_0)$ is basically stating what is the probability of B happening, if we start from 5 and $X_i = 0$ for some i in the future. Thus

$$\mathbb{P}_5(B) = \frac{\mathbb{P}_0(B)}{2}.$$

Now, we need to analyze the term $\mathbb{P}_1(B)$. Again, at this step, we do another first step analysis.

$$\mathbb{P}_1(B) = \frac{1}{3}(\underbrace{\mathbb{P}_3(B)}_{=1} + \underbrace{\mathbb{P}_2(B)}_{=1} + \mathbb{P}_0(B)) = \frac{2 + \mathbb{P}_0(B)}{3}.$$

Note that from the assumption, we know that if we reach any of green states, then we are declared winner, that is why we have $\mathbb{P}_3(B) = \mathbb{P}_2(B) = 1$. Now it only remains to analyze the term $\mathbb{P}_7(B)$. Again, similar to the case above, we do a first time step argument

$$\mathbb{P}_7(B) = \frac{1}{3}(\mathbb{P}_0(B) + \underbrace{\mathbb{P}_8(B)}_{=\mathbb{P}_7(B)} + \underbrace{\mathbb{P}_9(B)}_{=0}) \implies \mathbb{P}_7(B) = \frac{\mathbb{P}_0(B)}{2}.$$

Note that $\mathbb{P}_8(B) = \mathbb{P}_7(B)$ by a first stem analysis when starting at the state 8. Putting all of these terms back to the original identity we derived the first, we can conclude that

$$p_0 = \mathbb{P}_0(B) = \frac{2}{5}.$$

■ **Problem 2.10** The French roulette game has slots numbered from 0 to 36. The slot 0 is green, Among the slots from 1 to 36, 18 are black and 18 are red. Alex goes to a casino to play roulette. Their strategy is to always bet "red". They start with 50 coins, play 1 coin each turn, and stop when reaching 100 or getting broke.

(a) What is the probability that Alex reaches 100?

(b) How many coins should Alex start with to have about 50% chance to reach 100?

Solution (a) Let B be the event $B = \{T_{100} < T_0\}$ and we are looking for $\mathbb{P}_a(B)$ where $0 \leq a \leq 100$ and indicates the number of coins we are starting with. First observe that

- $p_0 = 0$: Since if we start with zero coins we are already broken and the game is over.
- $p_{100} = 1$: Since if we start with 100 coins then we won the game and the game is finished.

To compute the probability for intermediate values of a, we do the first step argument. Let WF be the event where Alex wins the first bet, and LF the event where Alex loses the first bet. Then we can write

$$p_a = \mathbb{P}_a(B) = \mathbb{P}_a(B|WF)\mathbb{P}_a(WF) + \mathbb{P}_a(B|LF)\mathbb{P}_a(LF).$$

Since there are 18 red spots, then the chance to win the first bet is

$$\mathbb{P}_a(WF) = \frac{18}{37}.$$

and since there are 19 non-red spots in total, then the chance to win is

$$\mathbb{P}_a(LF) = \frac{19}{37}.$$

Also, from Markov property, we know that

$$\mathbb{P}_a(B|WF) = p_{a+1}, \qquad \mathbb{P}_a(B|LF) = p_a.$$

Thus the first step argument formula will be

$$p_a = \frac{18}{37}p_{a+1} + \frac{19}{37}p_{a-1} \implies \boxed{37p_a = 18p_{a+1} + 19p_{a-1}}.$$

The characteristic equation for the recursive equation is

$$37 = 18x + \frac{19}{x} \implies \boxed{18x^2 - 37x + 19 = 0}.$$

We can write it as $(x-1)(18x-19) = 0$. Thus the roots will be

$$r_1 = 1, \qquad r_2 = \frac{19}{18}.$$

So

$$p_a = A + Br_2^a.$$

To fine A and B we use the fact $p_0 = 0$, and $p_{100} = 1$. Then $A = -B$, and $A = 1/(1 - r_2^{100})$. Thus

$$\boxed{p_a = \frac{1 - r_2^a}{1 - r_2^{100}}}.$$

(b) We basically need to compute find a for which $p_a = 1/2$. Thus we need to solve for a

$$\frac{1 - r_2^a}{1 - r_2^{100}} = \frac{1}{2}.$$

After some algebra we will find

$$\boxed{a = \frac{\ln\left(\frac{1+r_2^{100}}{2}\right)}{\ln(r_2)} \approx 87.26}.$$

Thus we need to start with at least 88 coins to have a 50% chance of winning.

□

■ **Problem 2.11** There are 6 coins on a table, each showing heads (H) or tails (T). In each step we

- Select uniformly one of the coins.
- If it is heads, toss it and replace on the table (with random side).
- If it sis tails, toss it. If it comes up heads, leave it at that. If it comes up tails, toss it a second time, and leave the result as it is. Let X_n be the number of heads showing after n such steps. Answer the following questions
  (a) Determine the transition probabilities for this Markov chain.
  (b) Draw the transition diagram and write the transition matrix.
  (c) What is $\mathbb{P}(X_2 = 4 | X_0 = 5)$?

Solution (a) To compute the transition probabilities, we need to perform the first step analysis. Let the events

$$I = \{X_1 = a + 1\}, \qquad S = \{X_1 = a\}, \qquad D = \{X_1 = a - 1\},$$

where $0 \le a \le 6$ is the number of heads. So to compute the transition probabilities, we need to compute

$$P(a, a+1) = \mathbb{P}_a(I), \qquad P(a, a) = \mathbb{P}_a(S), \qquad P(a, a-1) = \mathbb{P}_a(D).$$

We start with $\mathbb{P}_a(I)$. Let ST be the event where the selected coin is tails, and SH be the event where the selected coin is heads. These two events are disjoint and the probability of their union is 1, thus

$$\mathbb{P}_a(I) = \underbrace{\mathbb{P}_a(I|SH)}_{\text{see Eq (2.I.1)}} \underbrace{\mathbb{P}_a(SH)}_{\frac{a}{6}} + \underbrace{\mathbb{P}_a(I|ST)}_{\text{see Eq (2.I.2)}} \underbrace{\mathbb{P}_a(ST)}_{\frac{6-a}{6}}. \tag{2.I}$$

Note that if we start with a coins heads, then the chance we choose a random coin from the table and find it heads is $\frac{a}{6}$, hence $\mathbb{P}_a(SH) = \frac{a}{6}$, and $\mathbb{P}_a(ST) = \frac{6-a}{6}$. Now we need to expand the remaining terms with appropriate conditioning. Let TT be the event where we toss a coin and find it tails and TH be the event where we toss a coin and find it heads. Thus we can write

$$\mathbb{P}_a(I|SH) = \underbrace{\mathbb{P}_a(I|SH, TH)}_{0} \underbrace{\mathbb{P}_a(TS)}_{\frac{1}{2}} + \underbrace{\mathbb{P}_a(I|SH, TT)}_{0} \underbrace{\mathbb{P}_a(TT)}_{\frac{1}{2}}. \tag{2.I.1}$$

Note that $\mathbb{P}_a(TT) = \mathbb{P}_a(TH) = \frac{1}{2}$, since the coin tossing is fair. Also, note that $\mathbb{P}_a(I|SH, TH) = \mathbb{P}_a(I|SH, TT) = 0$ since if we select a heads, and then toss it, finding it either heads or tails will not increase the total number of heads on the table. Similarly, for the other term in (2.1) we have

$$\mathbb{P}_a(I|ST) = \underbrace{\mathbb{P}_a(I|ST,TH)}_{1}\underbrace{\mathbb{P}_a(TH)}_{\frac{1}{2}} + \underbrace{\mathbb{P}_a(I|ST,TT)}_{\text{see Eq (2.I.3)}}\underbrace{\mathbb{P}_a(TT)}_{\frac{1}{2}}. \tag{2.I.2}$$

Now we need to expand the remaining terms in the equation above.

$$\mathbb{P}_a(I|ST,TT) = \underbrace{\mathbb{P}_a(I|ST,TT,TH)}_{1}\mathbb{P}_a(TH) + \underbrace{\mathbb{P}_a(I|ST,TT,TT)}_{0}\mathbb{P}_a(TT) = \frac{1}{2}. \tag{2.I.3}$$

Putting all together we can write

$$\boxed{P(a,a+1) = \mathbb{P}_a(I) = \frac{6-a}{8}}.$$

Similarly, we can compute other transition probabilities. For instance for $\mathbb{P}_a(S)$ we can write

$$\mathbb{P}_a(S) = \underbrace{\mathbb{P}_a(S|SH)}_{\text{see Eq (2.S.1)}}\underbrace{\mathbb{P}_a(SH)}_{\frac{a}{6}} + \underbrace{\mathbb{P}_a(S|ST)}_{\text{see Eq (2.S.2)}}\underbrace{\mathbb{P}_a(ST)}_{\frac{6-a}{6}}. \tag{2.S}$$

and for the remaining terms we can write

$$\mathbb{P}_a(S|SH) = \underbrace{\mathbb{P}_a(S|SH,TH)}_{1}\underbrace{\mathbb{P}_a(TS)}_{\frac{1}{2}} + \underbrace{\mathbb{P}_a(S|SH,TT)}_{0}\underbrace{\mathbb{P}_a(TT)}_{\frac{1}{2}}, \tag{2.S.1}$$

and

$$\mathbb{P}_a(S|ST) = \underbrace{\mathbb{P}_a(S|ST,TH)}_{0}\underbrace{\mathbb{P}_a(TH)}_{\frac{1}{2}} + \underbrace{\mathbb{P}_a(S|ST,TT)}_{\text{see Eq (2.S.3)}}\underbrace{\mathbb{P}_a(TT)}_{\frac{1}{2}}. \tag{2.S.2}$$

And for the remaining term above

$$\mathbb{P}_a(S|ST,TT) = \underbrace{\mathbb{P}_a(S|ST,TT,TH)}_{0}\mathbb{P}_a(TH) + \underbrace{\mathbb{P}_a(S|ST,TT,TT)}_{1}\mathbb{P}_a(TT) = \frac{1}{2}. \tag{2.S.3}$$

and by putting all together we will get

$$\boxed{P(a,a) = \mathbb{P}_a(S) = \frac{6+a}{24}}.$$

Finally, since $\mathbb{P}_a(I \cup S \cup D) = 1$, and I, S, D are mutually disjoint, we can write

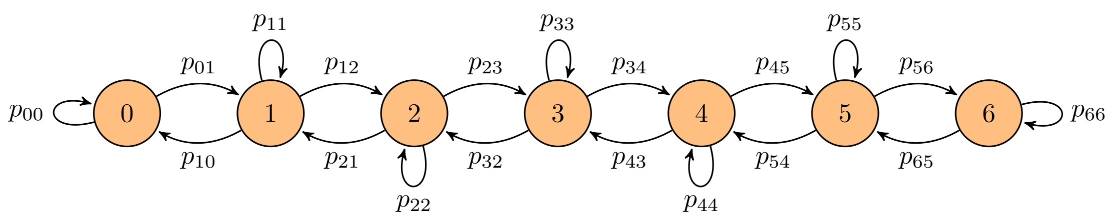
$$\mathbb{P}_a(D) = 1 - (\mathbb{P}_a(I) + \mathbb{P}_a(S)),$$

hence

$$\boxed{P(a,a-1) = \mathbb{P}_a(D) = \frac{a}{12}}.$$

so the transition probabilities are as calculated.

(b) The transition diagram is plotted below.



And the transition matrix is

$$M = \begin{pmatrix} 1/4 & 3/4 & 0 & 0 & 0 & 0 & 0 \\ 1/12 & 7/24 & 5/8 & 0 & 0 & 0 & 0 \\ 0 & 1/6 & 1/3 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/4 & 3/8 & 3/8 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 5/12 & 1/4 & 0 \\ 0 & 0 & 0 & 0 & 5/12 & 11/24 & 1/8 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix}$$

(c) $\mathbb{P}(X_2 = 4|X_0 = 5)$ is the second transition probability $P_2(5,4)$. To compute this, we need to fine the element in the 6-th row and 5-th column in the M^2 matrix, which is basically the inner product between the vectors formed by the 6-th row and the 5-th column.

$$P_2(5,4) = (\frac{5}{12})^2 + \frac{11}{24} \cdot \frac{5}{12} = \frac{35}{96}$$

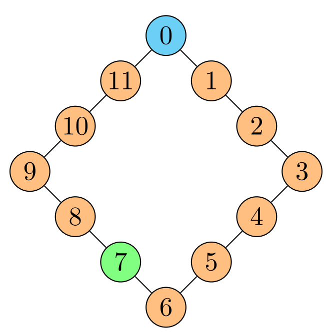
which after simplification becomes

$$\boxed{P_2(5,4) = \frac{35}{96}}.$$

□

■ **Problem 2.12** A clock is broken. It has only one hand which moves every hour either clockwise with probability 1/2 or counter-clockwise with probability 1/2 (the numbers are from 0 to 11 and the hand moves by one full hour when it moves). Assume it starts at 0. What is the probability that it reaches 7 before coming back to 0 for the first time?

Solution First, let's draw the graph representing the state space of the random variable of interest.



Define the event B be $B = \{T_0^+ > T_7\}$. We are interested in finding $\mathbb{P}_0(B)$. Now we can perform the first step argument as follows

$$p_0 = \frac{1}{2}(p_1 + p_{11}). \tag{3.1}$$

Then we analyze each term in the right hand side of the equation above. For p_1 we have

$$\mathbb{P}_1(B) = \underbrace{\mathbb{P}_1(B|T_0 > T_7)}_{1}\underbrace{\mathbb{P}_1(T_0 > T_7)}_{1/5} + \underbrace{\mathbb{P}_1(B|T_0 < T_7)}_{0}\underbrace{\mathbb{P}_1(T_0 < T_7)}_{6/7} = \frac{1}{5}.$$

Note that $\mathbb{P}_1(B|T_0 > T_7) = 1$ since it literally means the random walker reaches 7 before 0. Also $\mathbb{P}_1(B|T_0 < T_7) = 0$ since the event B is conditioned on reaching 0 before 7, which is clearly 0. The term $\mathbb{P}_1(T_0 > T_7)$ is computed using the Gambler's ruin analysis. Similarly, for the p_{11} term we have

$$\mathbb{P}_{11}(B) = \underbrace{\mathbb{P}_{11}(B|T_0 > T_7)}_{1}\underbrace{\mathbb{P}_{11}(T_0 > T_7)}_{1/7} + \underbrace{\mathbb{P}_{11}(B|T_0 < T_7)}_{0}\mathbb{P}_{11}(T_0 < T_7) = \frac{1}{7}.$$

The rationale behind the values of the terms are the same as the ones discussed above. Now we can substitute everything in (3.1)

$$\boxed{p_0 = \frac{1}{2}(\frac{12}{35}) = \frac{6}{35}}.$$

■ **Problem 2.13** The Fibonacci sequence is the sequence $(F_n)_{n\geq 0}$ defined by $F_0 = 0, F_1 = 1$ and

$$F_{n+2} = F_{n+1} + F_n \quad \text{for } n \geq 0.$$

Find a general formula for F_n

Solution First, we construct the characteristic polynomial of the sequence. From the recursive formula we can write

$$X^2 = X + 1 \implies \boxed{X^2 - X - 1 = 0}.$$

The roots of the equation is

$$r_1, r_2 = \frac{1 \pm \sqrt{5}}{2}.$$

Now the general formula will be

$$F_n = Ar_1^n + Br_2^n.$$

To find the coefficients, we utilize the first two terms

$$0 = A + B, \qquad 1 = \frac{1}{2}(A + B) + \frac{\sqrt{5}}{2}(A - B).$$

This system of equations implies that

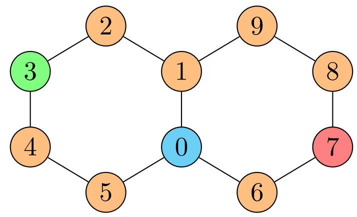
$$A = \frac{1}{\sqrt{5}}, \qquad B = \frac{-1}{\sqrt{5}}.$$

Thus the general formula will be

$$\boxed{F_n = \frac{1}{\sqrt{5}}((\frac{1+\sqrt{5}}{2})^n - \frac{1-\sqrt{5}}{2})^n)}.$$

□

■ **Problem 2.14** Let (X_n) be the simple random walk on the following graph. Compute $\mathbb{P}_0(T_3 < T_7)$.



Solution For a much more simpler solution, let's define the two following notations

$$B = \{T_3 < T_7\}, \qquad p_v = \mathbb{P}_v(B).$$

Then, by first step argument at state 0, we can write

$$p_0 = \frac{1}{3}(p_5 + p_6 + p_1). \tag{5.1}$$

Now we need to evaluate each of the terms in the right hand side. We start with p_5.

$$p_5 = \mathbb{P}_5(B) = \underbrace{\mathbb{P}_5(B|T_3 < T_0)}_{1}\underbrace{\mathbb{P}_5(T_3 < T_0)}_{1/3} + \underbrace{\mathbb{P}_5(B|T_3 > T_0)}_{p_0}\underbrace{\mathbb{P}_5(T_3 > T_0)}_{2/3} = \frac{1}{3} + \frac{2}{3}p_0.$$

note that $\mathbb{P}_5(B|T_3 < T_0) = 1$, since if we get to state 3, before getting to state 0, then it means that we have reached the state 3 before reaching the state 7, thus the event B occurs with probability 1. Also $\mathbb{P}_5(T_3 < T_0) = 1/3$ from the Gambler's ruin. Furthermore $\mathbb{P}_5(B|T_3 > T_0) = p_0$ by using the Markov property, and finally $\mathbb{P}_5(T_3 > T_0) = 2/3$ by the Gambler's ruin.
Now, we need to evaluate the term p_6. To analyze this term, we will do a first step argument starting at this point

$$p_6 = \mathbb{P}_6(B) = \frac{1}{2}(\underbrace{p_7}_{0} + p_0) = \frac{p_0}{2}.$$

Note that $p_7 = 0$, since then the event B has not occurred.
Finally, we need to analyze the term p_1. Again, by first step argument on this state we have

$$p_1 = \frac{1}{3}(p_0 + p_9 + p_2).$$

By doing a analysis on p_9 similar to the one we did for 5, we can write

$$p_9 = \mathbb{P}_9(B) = \underbrace{\mathbb{P}_9(B|T_7 < T_1)}_{0}\mathbb{P}_9(T_7 < T_1) + \underbrace{\mathbb{P}_9(B|T_7 > T_1)}_{p_1}\underbrace{\mathbb{P}_9(T_7 > T_1)}_{2/3} = \frac{2}{3}p_1.$$

The rationale behind the values for each term in the equation above, is exactly the same as in analyzing the terms of p_5.
Now, we analyze the term p_2 by performing another first step analysis, similar to the one we did for state 6.

$$p_2 = \frac{1}{2}(\underbrace{p_3}_{1} + p_1) = \frac{1}{2}(1 + p_1).$$

Now we can calculate p_1 in terms of p_0 which turns out to be

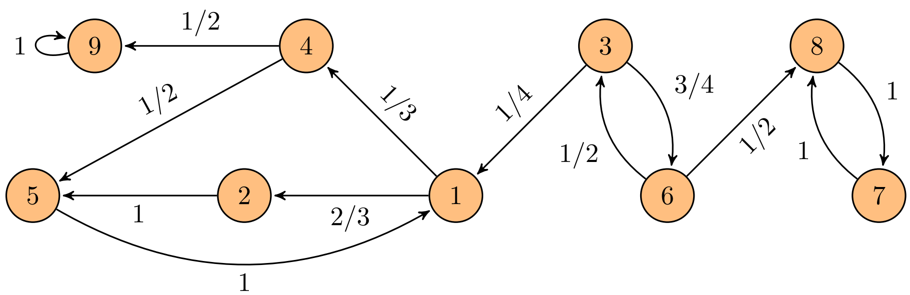
$$p_1 = \frac{6}{11}p_0 + \frac{3}{11}.$$

Now we insert all of the terms in the equation (5.1) to get

$$
\begin{aligned}
3p_0 &= \frac{1}{3} + \frac{2}{3}p_0 + \frac{1}{2}p_0 + \frac{6}{11}p_0 + \frac{3}{11} \\
\implies 3p_0 - \frac{113}{66}p_0 &= \frac{40}{33} \\
\implies p_0 &= \frac{66}{85} \cdot \frac{40}{33} = \frac{16}{17} \\
\implies & \boxed{p_0 = \frac{16}{17}}.
\end{aligned}
$$

□

■ **Problem 2.15** Consider the Markov chain on the state space $S = \{1, 2, \cdots, 9\}$ which has the following transition diagram.



(a) Which states are recurrent? (Justify it only for the states 1 and 7).

(b) What are the periods of all states? (Justify it only for the state 2).

(c) What are the communicating classes of the chain?

(d) Compute f_3.

(e) Compute f_1.

Solution (a) The transient states are $T = \{4, 5, 2, 1, 3, 6\}$, and the recurrent states are $R = \{9, 8, 7\}$.

Justification for state 1. The state 1 is transient. Because there is a positive probability that a random walk starting from this state will never come back to this state. For instance, if the random walker takes the path $1 \to 4$ with probability $1/3$ and then take the path $4 \to 9$ with probability $1/2$, then there is a $1/6$ chance that the random walker starting from state 1 will end up at 9 and will never return to the state 1 again.

Justification for state 7. The state 7 is indeed recurrent. That is because there is no chance for a random walker starting from the state 7 do not come back to 7. To be more specific, if $X_0 = 7$, then $\mathbb{P}_7(X_1 = 8) = 1$ and $\mathbb{P}_7(X_2 = 7) = \mathbb{P}_8(X_1 = 7) = 1$. Thus the random walker will return to the state 7 every even step.

(b) To calculate the period, we first need to determine $\mathcal{T}(x) = \{n \geq 1 : P_n(x, x) > 0\}$. Then $\text{per}(x) = \gcd(\mathcal{T}(x))$.

- **State 1**. $\mathcal{T}(1) = \{3, 6, 9, \cdots\}$. Thus $\text{per}(1) = 3$.
- **State 2**. $\mathcal{T}(2) = \{3, 6, 9, \cdots\}$. Thus $\text{per}(2) = 3$.

- **State 3**. $\mathcal{T}(3) = \{2, 4, 6, 8, \cdots\}$. Thus $\text{per}(3) = 2$.
- **State 4**. $\mathcal{T}(4) = \{3, 6, 9, \cdots\}$. Thus $\text{per}(4) = 3$.
- **State 5**. $\mathcal{T}(5) = \{3, 6, 9, \cdots\}$. Thus $\text{per}(5) = 3$.
- **State 6**. $\mathcal{T}(6) = \{2, 4, 6, \cdots\}$. Thus $\text{per}(6) = 2$.
- **State 7**. $\mathcal{T}(7) = \{2, 4, 6, \cdots\}$. Thus $\text{per}(7) = 2$.
- **State 8**. $\mathcal{T}(8) = \{2, 4, 6, \cdots\}$. Thus $\text{per}(8) = 2$.
- **State 9**. $\mathcal{T}(9) = \{1, 2, 3, \cdots\}$. Thus $\text{per}(9) = 1$.

Justification for state 2. The set of all different paths that we can start from the state 2 and return to this state is $\mathcal{P} = \{\overrightarrow{2512}, \overrightarrow{2514512}\}$ or any concatenation of these two paths. Thus the first return times will be $\{3, 6, 9, \cdots\}$, whose gcd is 3. Thus the period of the state 2 is 3.

(c) The communicating classes of the graph are $\{9\}, \{1, 2, 5, 4\}, \{3, 6\}$, and $\{8, 7\}$.

(d) f_3 is the probability that the random walker will return to the state 3 given $X_0 = 3$. We do a first step analysis. Let the event $E = \{T_3^+ < \infty | X_0 = 3\}$, and $E' = \{T_3 < \infty | X_0 \neq 3\}$.

$$\mathbb{P}_3(E) = P(3,1)\underbrace{\mathbb{P}_1(E')}_{0} + \underbrace{P(3,6)}_{3/4}\mathbb{P}_6(E').$$

$\mathbb{P}_1(E) = 0$ because there the state 3 is not accessible from the state 1. Now we need to determine $\mathbb{P}_6(E)$ as follows

$$\mathbb{P}_6(E') = P(6,8)\underbrace{\mathbb{P}_8(E')}_{0} + \mathbb{P}(6,3)\underbrace{\mathbb{P}_3(E')}_{1}.$$

$\mathbb{P}_8(E) = 0$ because the state 3 is not accessible from the state 8. Also note that we put a $'$ on E in $\mathbb{P}_3(E')$, that is because it is not the same is $\mathbb{P}_3(E)$. The reason is that $\mathbb{P}_3(E')$ is basically 1 because it is when the first recurrence to the state E has happened. Putting all of these together we will get

$$\mathbb{P}_3(E) = \frac{3}{8}.$$

Thus we can conclude

$$\boxed{f_3 = \frac{3}{8}}.$$

(e) We want to compute $f_1 = \mathbb{P}_1(T_1^+ < \infty)$. For convenience in notation, let $E = \{T_1^+ < \infty | X_0 = 1\}$ and $E' = \{T_1 < \infty | X_0 \neq 1\}$. Note that E is the first recurrent time given we are started at 1 while E' is simply the event where we will visit the state 1 in finite time if we start elsewhere. So $f_1 = \mathbb{P}_1(E)$. We can do the first step analysis

$$\mathbb{P}_1(E) = P(1,4)\mathbb{P}_4(E') + P(1,2)\underbrace{\mathbb{P}_2(E')}_{1}.$$

Note that $\mathbb{P}_2(E') = 1$ because starting at the state 2, there are no any other possibilities but reaching the state 1 just after 2 steps. But we need to determine $\mathbb{P}_4(E')$ more carefully. We again do a first step analysis

$$\mathbb{P}_4(E') = P(4,5)\underbrace{\mathbb{P}_5(E')}_{1} + P(4,9)\underbrace{\mathbb{P}_9(E')}_{0}$$

Note that $\mathbb{P}_9(E') = 0$ since the state 1 is not accessible from state 9. However, $\mathbb{P}_5(E') = 1$ as starting at the state 5, we will touch the state 1 just after 1 step. In summary

$$\boxed{\mathbb{P}_1(E) = \frac{1}{6} + \frac{2}{3} = \frac{5}{6}}.$$

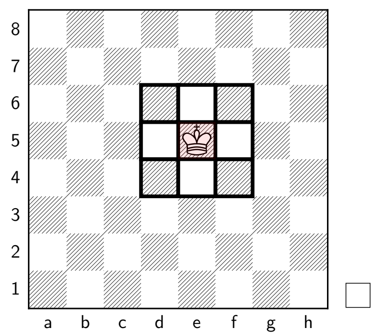
■ **Problem 2.16** Consider the following Markov chains on a chess board (the state space consists of the 64 squares of a chess board). You can solve the problem on a 4 times 4 chess board as it will be easier to explain the ideas on the drawings.

(a) A chess king moves on it uniformally to an allowed position (there are no other pieces on the board). Is this irreducible? Are the states periodic? recurrent?

(b) Same questions if the king replaced by a bishop?

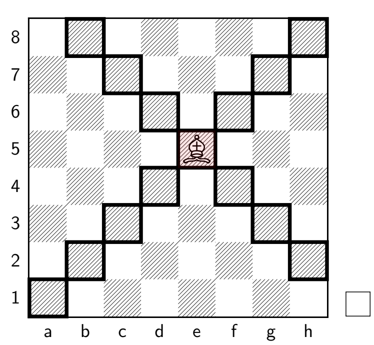
(c) Same questions if the king is replaced by a knight?

Solution (a) Consider the following diagram



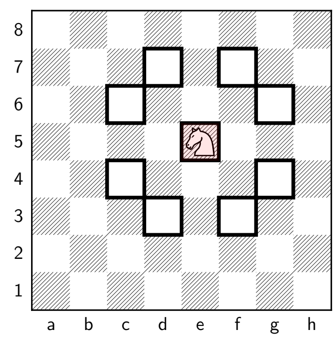
As it is represented in the board game, the King can move to any position. This means that starting from any position, the king can reach other positions, thus implying that every two state commutes. Thus we can say that this system is irreducible.
The states are *not* periodic. In short, that is because of the availability of the diagonal move of the King. In other words, $\mathcal{T}(x) = (2, 3, 4, 5, \cdots)$. It is now clear that $\text{per}(x) = 1$.
Since the Markov chain is irreducible, then it means that all of the states communicate. On the other hand, since the number of states is finite, thus there should be at least one recurrent state. But since this recurrent state is at the same communication class of any other state, then it means that all of the states are recurrent.

(b) The following shows the possible *next* moves for the bishop initially located at e5.



This Markov chain is *not* irreducible. The reason is that we have two communicating classes: white squares and black squares. I bishop starting at a white square can only go to the white square at its life time and a bishop starting at a black square can only go to the black square. To determine the periodicity, we need to calculate the set $\mathcal{T}(x)$. This set is $\mathcal{T}(x) = \{2, 3, 4, \cdots\}$. Thus $\text{per}(x) = 1$ which indicates that the states are not periodic.
If a bishop starts from a white square, then all of the white squares will be communicating with the initial square, and any black square will be no accessible. Since the number of states are finite, amount the white squares there will be at least one state that is recurrent. And since the recurrency is class property, then it means that all of the white squares will be recurrent, while all of the black squares will be transient. Similarly, if a bishop starts from a black square, then all of the black squares will be recurrent, while all of the white squares will be transient.

(c) Consider the following chess board



□

First, notice that if the knight starts are a black square, then it moves to a white square at the next step, and if starts at a white square, it moves to a black square at the next step. Also, due to the "knight's tour" property, the knight can reach every state of the board. Thus we can say that any two states on the board communicates. Thus the Markov chain is irreducible.
To determine the periodicity, we determine $\mathcal{T}(x) = 2, 4, 6, \cdots$. Note that the knight can not come back to its original position after an odd number of moves. That is because every time a knight moves, the destination state has the opposite color of the initial state. Thus it is impossible to get back to the same color as the starting point after odd number of moves. This $\text{pre}(x) = 2$.
Since the Markov chain is irreducible (i.e. all of the states communicated), and the number of states is finite, then there is at least one recurrent state. Then all of the states are recurrent.

■ **Problem 2.17** Consider the fair Gambler's ruin problem on $\{0, 1, \cdots, N\}$. Let $u_a = \mathbb{E}_a[\min(T_0, T_N)]$ be the expected duration (=average number of steps) of the game if Alice starts with $0 \leq a \leq N$ coins (so in particular we have $u_0 = u_N = 0$).

(a) Show that if $1 \geq a \geq N - 1$, we have

$$u_a = \frac{1}{2}(1 + u_{a-1}) + \frac{1}{2}(1 + u_{a+1}).$$

(b) Define $v_a = u_a - u_{a-1}$ for $1 \leq a \leq N$. Show that $v_a = v_{a+1} + 2$ and $\sum_{a=1}^{N} v_a = 0$.d

Solution (a) We are basically dealing with a conditional probability

$$\mathbb{E}_a[\min(T_0, T_N)] = \mathbb{E}[\min(T_0, T_N)|X_0 = a].$$

Now we do a first step analysis

$$\begin{aligned}\mathbb{E}[\min(T_0, T_N)|X_0 = a] =& \overbrace{P(a, a+1)}^{1/2}\overbrace{\mathbb{E}[\min(T_0, T_N)|X_1 = a+1]}^{1+u_{a+1}} \\ &+ \underbrace{P(a, a-1)}_{1/2}\underbrace{\mathbb{E}[\min(T_0, T_N)|X_1 = a-1]}_{1+u_{a-1}} \\ =&\frac{1}{2}(1+u_{a+1}) + \frac{1}{2}(1+u_{a-1}).\end{aligned}$$

Note that in the equation above we have used the fact that $\mathbb{E}[\min(T_0, T_N)|X_0 = a+1]$ as by moving from the state a to the state $a+1$ (and similarly to $a-1$) one step is already passed.

(b) First, we define $v_a = u_a - u_{a-1}$ for $1 \leq a \leq N$. Using the recursive relation from part (a), we have:

$$u_a = \frac{1}{2}(1+u_{a-1}) + \frac{1}{2}(1+u_{a+1})$$

Now, by substituting the definition of v_a into the equation, we get:

$$u_{a-1} = u_a - v_a$$

$$u_{a+1} = u_a + v_{a+1}$$

Now, substituting these into the equation for u_a yields:

$$\begin{aligned}u_a &= \frac{1}{2}(1+u_a - v_a) + \frac{1}{2}(1+u_a+v_{a+1}) \\ 2u_a &= 2 + 2u_a - v_a + v_{a+1} \\ v_a &= v_{a+1} + 2\end{aligned}$$

This shows that $v_a = v_{a+1} + 2$. For the second part, we consider the sum:

$$\sum_{a=1}^{N} v_a = \sum_{a=1}^{N}(u_a - u_{a-1})$$

This is a telescoping series, and thus we have:

$$\sum_{a=1}^{N} v_a = (u_1 - u_0) + (u_2 - u_1) + \ldots + (u_N - u_{N-1})$$

$$\sum_{a=1}^{N} v_a = -u_0 + u_N$$

Since $u_0 = u_N = 0$ (as given in the exercise), it follows that:

$$\sum_{a=1}^{N} v_a = 0$$

This completes the answer for part (b).

(c) We have from part (b) that $v_a = v_{a+1} + 2$. Iterating this, we find a linear relation:

$$v_a = v_1 - 2(a-1)$$

We also know from the sum $\sum_{a=1}^{N} v_a = 0$ that:

$$Nv_1 - 2\sum_{a=1}^{N}(a-1) = 0$$

$$Nv_1 - 2\left(\frac{N(N-1)}{2}\right) = 0$$

$$v_1 = N - 1$$

Substituting back into v_a gives us:

$$v_a = N + 1 - 2a$$

From the recursive definition $u_a = u_{a-1} + v_a$, we can express u_a as:

$$u_a = \sum_{k=1}^{a}(N + 1 - 2k)$$

$$u_a = a(N+1) - 2\left(\frac{a(a+1)}{2}\right)$$

$$u_a = a(N-a)$$

This completes the proof for part (c).

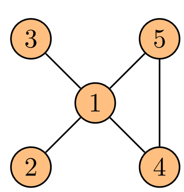
(e) From part we see that the average time of the game is

$$t = \frac{a(N-a)}{1-\delta}.$$

By simply substituting $N = 5$, $a = 2$, and $\delta = 1/3$, we will get

$$t = \frac{2 \cdot 3}{2/3} = 9.$$

■ **Problem 2.18** Let $(X_n)_n$ be the simple random walk on the following graph.



(a) What is $\mathbb{P}_3(T_5 < \infty)$.

(b) What is $\mathbb{E}_2[N_4]$?

(c) Let $\mu = (0 \quad 0.1 \quad 0.2 \quad 0.3 \quad 0.4)$. Compute $\mathbb{P}_\mu(T_5 < T_2)$.

(d) Compute $\mathbb{E}_\mu[T_5]$.

Solution (a) We can do a first step argument, but instead, we will use the fact that the Markov chain is irreducible (all of the states communicate with each other). Since the state space is finite, then there should be a recurrent state. On the other hand, since all of the nodes are at the same communication class, then all of the states are recurrent. So every node will be visited infinitely many times. Thus we can write

$$\mathbb{P}_3(T_5 < \infty) = 1.$$

(b) As we argued in section (a), all of the states are recurrent, thus every state will be visited infinitely many times. Thus

$$\mathbb{E}_2[N_4] = \infty.$$

(c) First we need to calculate $\mathbb{P}_v(T_5 < T_2)$ for $v \in \{1, \dots, 5\}$. Let $E = \{T_5 < T_2\}$. Also, observe that $\mathbb{P}_5(E) = 1, \mathbb{P}_2(E) = 0$.

By doing the first step argument for long enough steps we can write

$$\begin{aligned}
\mathbb{P}_1(E) &= \frac{1}{4}(\underbrace{\mathbb{P}_3(E)}_{\mathbb{P}_1(E)} + \underbrace{\mathbb{P}_2(E)}_{0} + \underbrace{\mathbb{P}_4(E)}_{1/2(\mathbb{P}_1(E)+\mathbb{P}_5(E))} + \mathbb{P}_5(E)) = \frac{3}{5}.\\
\mathbb{P}_3(E) &= \mathbb{P}_1(E) = \frac{3}{5}.\\
\mathbb{P}_4(E) &= \frac{1}{2}(\mathbb{P}_1(E) + \mathbb{P}_5(E)) = \frac{1}{2}(\frac{3}{5} + 1) = \frac{4}{5}.
\end{aligned}$$

Now we can calculate $\mathbb{P}_\mu(E)$ using

$$\mathbb{P}_\mu(E) = \mathbb{P}(E|X_0 \sim \mu) = \sum_{i=1}^{5} \underbrace{\mathbb{P}(E|X_0 = i)}_{\mathbb{P}_i(E)} \mu(i) = 0.2 \cdot \frac{3}{5} + 0.3 \cdot \frac{4}{5} + 0.4 = 0.76.$$

(d) Observe that

$$\mathbb{E}_\mu[T_5] = \sum_{x \in S} \mathbb{E}[T_5|X_0 = x]\mathbb{P}(X_0 = x) = \sum_{x \in S} \mathbb{E}_x[T_5]\mu(x).$$

So, first we need to calculate $\mathbb{E}_x[T_5]$. To simplify the notation we denote $u_x = \mathbb{E}_x[T_5]$. Also, observe that $u_5 = 0$, since starting at 5, we are already at 5, thus the hitting time will be zero, thus the expected value of the hitting time will be zero. Then we can write

$$u_1 = \frac{1}{4}(\underbrace{u_2}_{u_1+1} + \underbrace{u_3}_{u_1+1} + \underbrace{u_4}_{1/2(u_1+u_5)+1} + \underbrace{u_5}_{0}) + 1 \implies \boxed{u_1 = \frac{14}{3}}.$$

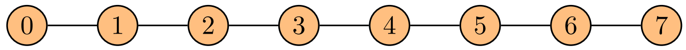
Similarly, for the other states we can write

$$\begin{aligned}
u_2 &= u_1 + 1 \implies \boxed{u_2 = \frac{17}{3}},\\
u_3 &= u_1 + 1 \implies \boxed{u_3 = \frac{17}{3}},\\
u_4 &= \frac{1}{2}(u_1 + u_5) + 1 \implies \boxed{u_4 = \frac{10}{3}}.
\end{aligned}$$

Putting all of these results together we will get

$$\mathbb{E}_\mu[T_5] = (0 \times \frac{14}{3}) + (0.1 \times \frac{17}{3}) + (0.2 \times \frac{17}{3}) + (0.3 \times \frac{10}{3}) + (0.4 \times 0) = 2.7.$$

■ **Problem 2.19** A Pokemon trainer is walking randomly in a one dimensional field which we represent as follows



The trainer just started their journey and only has a Charmander. Each time the trainer visists the left end of the field (the vertex 0), a wild Caterpie attacks and Charmander beats it. If the trainer is at the right end of the field (at vertex 7), their rival attacks them with a Squirtle. To beat Squirtle, Charmander needs to train on Caterpies: it beats squirtle if and only of it has beaten ≥ 3 Caterpies before. What is the probability that Charmander beats Squirtle if the trainer starts at 2?

Solution After abstracting away all of the details, we are basically aiming at finding the probability that the trainer hits the state 0 at least three times before hitting the state 7. To put it more formally, Let $E_1 = \{T_0^{(1)} < T_7^{(1)}\}$ which is the event at which we touch the state 0 before the state 7 for the first time. Similarly, define $E_2 = \{T_0^{(2)} < T_7^{(2)}\}$ and $E_3 = \{T_0^{(3)} < T_7^{(3)}\}$. Also define the event W where the Charmander beats the Squirtle. Then we can write

$$\mathbb{P}_2(W) = \mathbb{P}_2(W|E_1)\underbrace{\mathbb{P}_2(E_1)}_{6/8} + \underbrace{\mathbb{P}_2(W|E_1^c)}_{0}\underbrace{\mathbb{P}_2(E_1^c)}_{2/8}.$$

Note that $\mathbb{P}_2(E_1) = 6/8$ from the Gambler's ruin problem. Now we need to determine $\mathbb{P}_2(W|E_1)$:

$$\mathbb{P}_2(W|E_1) = \mathbb{P}_1(W) = \mathbb{P}_1(W|E_2)\underbrace{\mathbb{P}_1(E_2)}_{7/8} + \underbrace{\mathbb{P}_1(W|E_2^c)}_{0}\mathbb{P}_1(E_2^c).$$

and finally

$$\mathbb{P}_1(W|E_2) = \mathbb{P}_1(W) = \underbrace{\mathbb{P}_1(W|E_3)}_{1}\underbrace{\mathbb{P}_1(E_3)}_{7/8} + \underbrace{\mathbb{P}_1(W|E_3^c)}_{0}\mathbb{P}_1(E_3^c)$$

So the final answer will be

$$\boxed{\mathbb{P}_2(W) = \frac{7}{8} \times \frac{7}{8} \times \frac{6}{8}}.$$

■ **Problem 2.20** Are the following sums finite of infinite? Justify it only for (d).

(a) $\sum_{n=1}^{\infty} \frac{1}{n^6}$,

(b) $\sum_{n=1}^{\infty} \frac{\ln(n)}{n}$,

(c) $\sum_{n=1}^{\infty} \frac{(\ln(n))^{549816}}{n^{1.00017}}$,

(d) $\sum_{n=2}^{\infty} \frac{1}{\sqrt{n}\ln(n)}$.

Solution (a) Converges.

(b) Diverges.

(c) Converges.

(d) Diverges. We can tackle this question with different approaches, among which I will present two. One is to use the Comparison test. The ln function grows slower than $x^{1/4}$. I.e. $\exists C \in \mathbb{R}$ such that $\ln(n) < Cn^{1/4}$ for n large enough. Thus for n large enough

$$\ln(n) < Cn^{1/4} \implies \frac{1}{\ln(n)} > \frac{1}{Cn^{1/4}} \implies \frac{1}{\sqrt{n}\ln(n)} > \frac{1}{\sqrt{n}Cn^{1/4}} = \frac{C_0}{n^{3/4}},$$

for some $C_1 \in \mathbb{R}$. Observe that

$$\sum_{n=N}^{\infty} \frac{C_0}{n^{3/4}} = \infty.$$

Thus by comparison test we can conclude that the series of interest diverges as well.

A second way to show this is to use the Cauchy's condensation test. To show the series $\sum f(n)$ converges or diverges (for monotone decreasing function n), it is enough to show $\sum 2^n f(2^n)$ converges or diverges. So we can transform the series of interest to

$$\sum_{n=2}^{\infty} \frac{2^n}{\sqrt{2^n}\ln(2^n)} = \sum_{n=2}^{\infty} \frac{2^{n/2}}{n\ln(2)} = \infty.$$

■ **Problem 2.21** If $\Gamma = (V, E)$ is a finite connected graph, we call cover time of Γ the first time at which the simple random walk on Γ has visited al vertices at least once.

$$T_{cov} := \max_{x \in V} T_x$$

The aim of this exercise is to study the cover time of the segment of length $N \geq 1$, which is the graph $\text{Seg}_N = (V, E)$ where $V = \{0, 1, \cdots, N\}$ and $E = \{\{n, n+1\} : 0 \leq n \leq N-1\}$.

(a) What is the stationary distribution π for the simple random walk on Seg_N?

(b) Let $a \in V$. Compute $\mathbb{E}_a[T_{cov}]$.

Solution (a) Intuitively speaking the stationary distribution is the fraction of time that the random walker spends on each node. To be more specific, for simple random walk on a graph, the value of the stationary distribution for vertex $v \in V$ is $\deg(v)/(2|E|)$. So the stationary distribution π will be

$$\pi = (\frac{1}{2N} \quad \frac{1}{N} \quad \frac{1}{N} \quad \cdots \quad \frac{1}{N} \quad \frac{1}{2N}).$$

(b) To calculate $E_a(T_{\text{cov}})$, we employ a method that involves establishing recurrence relations for the expected hitting times and then determining the expected cover time using the boundary conditions of the segment graph.

Let H_a denote the expected time to hit either end of the segment starting from vertex a. The recurrence relations for H_a are given by:

$$H_a = \frac{1}{2}H_{a-1} + \frac{1}{2}H_{a+1} + 1, \quad 1 \leq a \leq N-1. \tag{2.0.1}$$

with boundary conditions $H_0 = H_N = 0$.

The solution to the recurrence relations is obtained iteratively starting from the boundary conditions, leading to the following explicit formulas for the hitting times:

$$H_a = a(N-a), \quad 0 \leq a \leq N.$$

The expected cover time starting from vertex a is given by the maximum expected hitting time to either end of the segment:

$$E_a(T_{\text{cov}}) = \max(H_a, H_{N-a}).$$

Due to the symmetry of Seg_N, it suffices to calculate H_a for $a \leq \frac{N}{2}$ and reflect the results for $a > \frac{N}{2}$.

For a given vertex a, the expected cover time $E_a(T_{\text{cov}})$ is formally expressed as:

$$E_a(T_{\text{cov}}) = a(N-a), \quad \text{for } a \leq \frac{N}{2}.$$

For $a > \frac{N}{2}$, we have:

$$E_a(T_{\text{cov}}) = (N-a)a.$$

Small N example

we calculate the expected cover time for a random walk on the segment graph Seg_3, which is a path graph with vertices $V = \{0, 1, 2, 3\}$ and edges $E = \{\{n, n+1\}\}$ for $n = 0, 1, 2$.

For vertex $a = 1$, the expected cover time $E_1(T_{\text{cov}})$ is determined by the maximum expected time to visit all other vertices. Due to the small size of the graph, we can calculate this explicitly.

We denote the expected time to hit vertex 0 starting from vertex 1 as $H_{1,0}$ and the expected time to hit vertex 3 starting from vertex 1 as $H_{1,3}$. Due to the linearity of the graph, we have the following:

$$\begin{aligned} H_{1,0} &= 1 + H_{0,0} = 1 \\ H_{1,3} &= 1 + H_{2,3} = 1 + 1 + H_{3,3} = 2. \end{aligned}$$

The cover time T_{cov} from vertex 1 is the maximum of the hitting times:

$$T_{\text{cov}} = \max(H_{1,0}, H_{1,3}) = \max(1, 2) = 2.$$

Since the random walk must visit both vertices 0 and 3 to cover the graph, and the times to visit each vertex are independent, we have:

$$E_1(T_{\text{cov}}) = H_{1,0} + H_{1,3} = 1 + 2 = 3.$$

■ **Problem 2.22** If $\Gamma = (V, E)$ is a finite connected graph, we call cover time of Γ the first time at which the simple random walk on Γ has visited al vertices at least once.

$$T_{cov} := \max_{x \in V} T_x$$

The aim of this exercise is to study the cover time of the segment of length $N \geq 1$, which is the graph $\text{Seg}_N = (V, E)$ where $V = \{0, 1, \cdots, N\}$ and $E = \{\{n, n+1\} : 0 \leq n \leq N-1\}$.

(a) What is the stationary distribution π for the simple random walk on Seg_N?

(b) Let $a \in V$. Compute $\mathbb{E}_a[T_{cov}]$.

Solution (a) Intuitively speaking the stationary distribution is the fraction of time that the random walker spends on each node. To be more specific, for simple random walk on a graph, the value of the stationary distribution for vertex $v \in V$ is $\deg(v)/(2|E|)$. So the stationary distribution π will be

$$\pi = (\frac{1}{2N} \quad \frac{1}{N} \quad \frac{1}{N} \quad \cdots \quad \frac{1}{N} \quad \frac{1}{2N}).$$

(b) To calculate $E_a(T_{\text{cov}})$, we employ a method that involves establishing recurrence relations for the expected hitting times and then determining the expected cover time using the boundary conditions of the segment graph.

Let H_a denote the expected time to hit either end of the segment starting from vertex a. The recurrence relations for H_a are given by:

$$H_a = \frac{1}{2}H_{a-1} + \frac{1}{2}H_{a+1} + 1, \quad 1 \leq a \leq N-1. \tag{2.0.2}$$

with boundary conditions $H_0 = H_N = 0$.

The solution to the recurrence relations is obtained iteratively starting from the boundary conditions, leading to the following explicit formulas for the hitting times:

$$H_a = a(N-a), \quad 0 \leq a \leq N.$$

The expected cover time starting from vertex a is given by the maximum expected hitting time to either end of the segment:

$$E_a(T_{\text{cov}}) = \max(H_a, H_{N-a}).$$

Due to the symmetry of Seg_N, it suffices to calculate H_a for $a \leq \frac{N}{2}$ and reflect the results for $a > \frac{N}{2}$.

For a given vertex a, the expected cover time $E_a(T_{\text{cov}})$ is formally expressed as:

$$E_a(T_{\text{cov}}) = a(N-a), \quad \text{for } a \leq \frac{N}{2}.$$

For $a > \frac{N}{2}$, we have:

$$E_a(T_{\text{cov}}) = (N-a)a.$$

Small N example

we calculate the expected cover time for a random walk on the segment graph Seg_3, which is a path graph with vertices $V = \{0, 1, 2, 3\}$ and edges $E = \{\{n, n+1\}\}$ for $n = 0, 1, 2$.

For vertex $a = 1$, the expected cover time $E_1(T_{\text{cov}})$ is determined by the maximum expected time to visit all other vertices. Due to the small size of the graph, we can calculate this explicitly.

We denote the expected time to hit vertex 0 starting from vertex 1 as $H_{1,0}$ and the expected time to hit vertex 3 starting from vertex 1 as $H_{1,3}$. Due to the linearity of the graph, we have the following:

$$H_{1,0} = 1 + H_{0,0} = 1$$
$$H_{1,3} = 1 + H_{2,3} = 1 + 1 + H_{3,3} = 2.$$

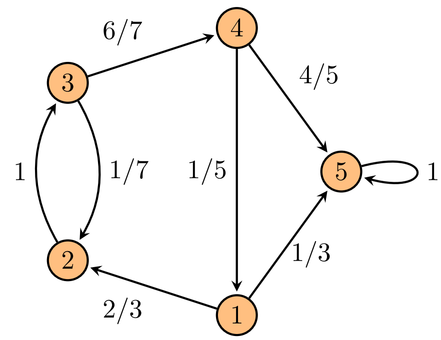
The cover time T_{cov} from vertex 1 is the maximum of the hitting times:

$$T_{\text{cov}} = \max(H_{1,0}, H_{1,3}) = \max(1, 2) = 2.$$

Since the random walk must visit both vertices 0 and 3 to cover the graph, and the times to visit each vertex are independent, we have:

$$E_1(T_{\text{cov}}) = H_{1,0} + H_{1,3} = 1 + 2 = 3.$$

■ **Problem 2.23** Consider the Markov chain on the state space $S = \{1, 2, \ldots, 5\}$ which has the following transition diagram.



(a) What is $\mathbb{E}_2[T_1]$?

(b) Let $\mu = (0.2 \quad 0.2 \quad 0.2 \quad 0.2 \quad 0.2)$. What is $\mathbb{P}_\mu(X_2 = 5)$?

(c) Compute f_3.

Solution (a) First, observe that the state 5 is recurrent, and does not communicate with the state 1. Starting from the state 2, there is a positive probability that we end up in state 5, where we will remain forever. Thus the conditional expectation value of hitting time of state 1, given we start at state 2 is infinite. I.e.

$$\mathbb{E}_2[T_1] = \infty.$$

(b) We have

$$\mathbb{P}_\mu(X_2 = 5) = \mathbb{P}(X_2 = 5 | X_0 \sim \mu) = \sum_{x \in S} \mathbb{P}_x(X_2 = 5)\mu(x).$$

If we denote the transition matrix to be M, then

$$\mathbb{P}_x(X_2 = 5) = (M^2)_{(x,5)}.$$

The transition matrix and its second power will be

$$M = \begin{pmatrix} 0 & 2/3 & 0 & 0 & 1/3 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1/7 & 0 & 6/7 & 0 \\ 1/5 & 0 & 0 & 0 & 4/5 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \qquad M^2 = \begin{pmatrix} 0 & 0 & 2/3 & 0 & 1/3 \\ 0 & 1/7 & 0 & 6/7 & 0 \\ 6/35 & 0 & 1/7 & 0 & 24/35 \\ 0 & 2/15 & 0 & 0 & 13/15 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Now, we can easily calculate the desired probability

$$\mathbb{P}_\mu[X_2 = 5] = 0.2(1/3 + 24/35 + 13/15 + 1) = \frac{101}{175}.$$

(c) To compute $f_3 = \mathbb{P}_3(T_3^+ < \infty)$, define $E = \{T_3^+ < \infty\}$ and $E' = \{T_3 < \infty\}$. Then we can do the first time step argument as follows

$$\mathbb{P}_3(E) = \frac{1}{7}\mathbb{P}_2(E') + \frac{6}{7}\mathbb{P}_4(E').$$

Note that $\mathbb{P}_2(E') = 1$ since if we end up at state 2, then we will certainly be at state 3 after one step. Now we need to determine $\mathbb{P}_4(E')$. We need to do the first step argument again

$$\mathbb{P}_4(E') = \frac{1}{5}\mathbb{P}_1(E') + \frac{4}{5}\mathbb{P}_5(E').$$

Note that since $\mathbb{P}_5(E') = 0$ since the state 5 is absorbing state. Now we need to calculate the term $\mathbb{P}_1(E')$ with another first step analysis

$$\mathbb{P}_1(E') = \frac{2}{3}\mathbb{P}_2(E') + \frac{1}{3}\mathbb{P}_5(E').$$

Putting all of pieces together, we will get

$$\boxed{f_3 = \frac{9}{35}}.$$

■ **Problem 2.24** Find the (probability) generating function $s \mapsto G_X(s) = \mathbb{E}[s^X]$ if X has the following distributions (you have to compute the infinite sums):

1. (3p) Poisson(λ) (i.e. $\mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$ for $k = 0, 1, 2, \ldots$),
2. (3p) Geom(p) (i.e. $\mathbb{P}(X = k) = p(1-p)^{k-1}$ for $k = 1, 2, \ldots$).

Solution **1. Poisson Distribution:** $X \sim \text{Poisson}(\lambda)$

Given the probability mass function (PMF) for the Poisson distribution as $\mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$ for $k = 0, 1, 2, \ldots$, the probability generating function (PGF) is calculated as follows:

$$\begin{aligned} G_X(s) = \mathbb{E}[s^X] &= \sum_{k=0}^{\infty} s^k \cdot \mathbb{P}(X = k) \\ &= \sum_{k=0}^{\infty} s^k \cdot e^{-\lambda}\frac{\lambda^k}{k!} \\ &= \exp\big(\lambda(s-1)\big). \end{aligned}$$

2. Geometric Distribution: $X \sim \text{Geom}(p)$

Given the PMF for the geometric distribution as $\mathbb{P}(X = k) = p(1-p)^{k-1}$ for $k = 1, 2, \ldots$, the probability generating function (PGF) is calculated as follows:

$$\begin{aligned} G_X(s) = \mathbb{E}[s^X] &= \sum_{k=1}^{\infty} s^k \cdot \mathbb{P}(X = k) \\ &= \sum_{k=1}^{\infty} s^k \cdot p(1-p)^{k-1} \\ &= \frac{ps}{ps - s + 1}, \quad \text{for } |s(1-p)| < 1. \end{aligned}$$

■ **Problem 2.25** Let $(X_n)_{n\geq 0}$ be an irreducible Markov chain on a finite state space S, and let π be its transition matrix. We recall that $\pi(x) = \frac{1}{\mathbb{E}_x[T_x^+]}$ for $x \in S$.

(a) (5p) Consider a knight initially at a corner of an 8×8 chessboard. How long does the knight need on average to come back to its starting corner?

(b) (12p) Show that the drunk man in 1 dimension needs on average an infinite number of steps to come back to 0 (i.e. show that on Grid_1 we have $\mathbb{E}_0[T_0^+] = \infty$). *Hint. Start by explaning why $\mathbb{E}_0[T_0^+] \geq 1 + \mathbb{E}_1[\min(T_0, T_N)]$, for all $N \geq 1$ (For this part, rereading the solution of H4E5 can be helpful). Then compute $\mathbb{E}_1[\min(T_0, T_N)]$ (you can do it since it is the same as on a segment).*

(c) (7p) Deduce that on Grid_2 we also have $\mathbb{E}_0[T_0^+] = \infty$.

Solution (a) TODO: TO BE ADDED.

(b) We start by considering the base case, where the drunk man moves one step away from the origin. Without loss of generality, assume he moves to position 1. The expected time to return to 0, $\mathbb{E}_0[T_0^+]$, can then be expressed as:

$$\mathbb{E}_0[T_0^+] \geq 1 + \mathbb{E}_1[\min(T_0, T_N)], \quad \forall N \geq 1.$$

This relation is based on the fact that the drunk man needs at least one step to move away from the origin and then additional steps to return.

To prove that $\mathbb{E}_0[T_0^+] = \infty$, consider the symmetric random walk's properties. The expected time to reach either 0 or N from 1 increases without bound as $N \to \infty$. This infinite expectation implies that the drunk man, on average, never returns to the origin, hence $\mathbb{E}_0[T_0^+] = \infty$.

(c) Given that a two-dimensional random walk (or drunk man's walk) allows for movement in four directions, the space for movement is significantly larger. However, despite the increased freedom, the probability of never returning to the origin remains non-zero. By analogy to the one-dimensional case and considering the nature of two-dimensional random walks, it is known to be recurrent but with an infinite expected return time. Thus, we conclude $\mathbb{E}_0[T_0^+] = \infty$ for a two-dimensional grid.

■ **Problem 2.26** Consider a BGW process $(Z_n)_{n\geq 0}$ and let μ and σ be respectively the mean and the standard deviation of its offspring distribution.

1. (10p) Let $n \geq 1$. Find an expression of $\text{Var}(Z_n)$ which depends only on μ, σ, and $\text{Var}(Z_{n-1})$. *Hint: Differentiate $G_{n+1} = G_1 \circ G_n$ twice.*

2. (5p) Conclude that for $n \geq 1$, $\text{Var}(Z_n) = \sigma^2 \mu^{n-1}(1 + \mu + ... + \mu^{n-1})$. *Hint: prove it by induction.*

Solution 1. Consider a Bienaymé-Galton-Watson process $(Z_n)_{n\geq 0}$ with mean μ and standard deviation σ of its offspring distribution. Let $n \geq 1$. We aim to find an expression for $\text{Var}(Z_n)$ that depends only on μ, σ, and $\text{Var}(Z_{n-1})$.

Given the hint to differentiate $G_{n+1} = G_1 \circ G_n$ twice, we proceed as follows. The first derivative of the generating function $G_{n+1}(s)$ with respect to s is:

$$G'_{n+1}(s) = G'_1(G_n(s)) \cdot G'_n(s)$$

And the second derivative is:

$$G''_{n+1}(s) = G''_1(G_n(s)) \cdot [G'_n(s)]^2 + G'_1(G_n(s)) \cdot G''_n(s)$$

At $s = 1$, the variance of Z_n can be expressed as:

$$\mathrm{Var}(Z_n) = G''_{n+1}(1) + G'_{n+1}(1) - [G'_{n+1}(1)]^2$$

Simplifying, and considering $G'_1(1) = \mu$ and $G''_1(1) = \sigma^2 + \mu - \mu^2$, we find:

$$\mathrm{Var}(Z_n) = \mu^2 \cdot \mathrm{Var}(Z_{n-1}) + \sigma^2 + \mu - \mu^2$$

This formula links the variance of the process at step n to the variance at step $n-1$, and the parameters of the offspring distribution, providing a recursive method to compute the variance at any step given the initial conditions.

2.

For $n = 1$, we have:

$$\mathrm{Var}(Z_1) = \sigma^2$$

which matches the formula $\mathrm{Var}(Z_n) = \sigma^2 \mu^{n-1}(1 + \mu + \cdots + \mu^{n-1})$ for $n = 1$.

Assume for some $n = k$ that:

$$\mathrm{Var}(Z_k) = \sigma^2 \mu^{k-1}(1 + \mu + \cdots + \mu^{k-1})$$

We need to show that:

$$\mathrm{Var}(Z_{k+1}) = \sigma^2 \mu^k (1 + \mu + \cdots + \mu^k)$$

Given:

$$\mathrm{Var}(Z_n) = \mu^2 \cdot \mathrm{Var}(Z_{n-1}) + \sigma^2$$

Substitute $n = k + 1$ and use the inductive hypothesis:

$$\mathrm{Var}(Z_{k+1}) = \mu^2 \cdot \sigma^2 \mu^{k-1}(1 + \mu + \cdots + \mu^{k-1}) + \sigma^2$$

Simplifying, we find:

$$\mathrm{Var}(Z_{k+1}) = \sigma^2 \mu^k (1 + \mu + \cdots + \mu^k)$$

thereby confirming our inductive step.

■ **Problem 2.27** Consider a BGW process with offspring distribution $(p_j)_{j \in \mathbb{Z}_+} = \left(\frac{2}{3} \frac{1}{3^j}\right)_{j \in \mathbb{Z}_+}$. What are μ, σ, η for this process?

Solution Given the offspring distribution $(p_j)_{j \in \mathbb{Z}_+} = \left(\frac{2}{3} \frac{1}{3^j}\right)_{j \in \mathbb{Z}_+}$, we calculate the mean (μ), variance (σ^2), and extinction probability (η) of the process.

The mean of the offspring distribution is calculated as:

$$\mu = 0.5$$

The variance of the offspring distribution is calculated as:

$$\sigma^2 = 0.75$$

Because $\mu < 1$, then we are at sub-critical phase, thus $\eta = 1$.

■ **Problem 2.28** Let P be finite and irreducible. Let π be the stationary distribution of P, and let (λ, f) be a right eigenpair of P (i.e. $Pf = \lambda f$) such that $\lambda \neq 1$. Show that $\pi f = 0$. *Hint: try to plug P between π and f.*

Solution The given conditions include:

- P is the transition matrix of the Markov chain, which is finite and irreducible.
- π is the stationary distribution of P, implying $\pi P = \pi$.
- (λ, f) is a right eigenpair of P, where $Pf = \lambda f$ and $\lambda \neq 1$.

Given $Pf = \lambda f$, we can express:

$$\pi(Pf) = \pi(\lambda f).$$

Since $\pi P = \pi$, it follows that:

$$\pi(Pf) = (\pi P)f = \pi f.$$

Thus, we have:

$$\pi(\lambda f) = \pi f.$$

Expanding the equation gives us:

$$\lambda(\pi f) = \pi f.$$

Rearranging, we find:

$$(\lambda - 1)(\pi f) = 0.$$

Given that $\lambda \neq 1$, $\lambda - 1 \neq 0$. Therefore, the only solution for $(\lambda - 1)(\pi f) = 0$ is if $\pi f = 0$.

■ **Problem 2.29** In this exercise we consider the BGW processes (Z_n) with $p_0 = \frac{7}{10}, p_3 = \frac{2}{10}, p_10 = \frac{0}{100}, p_50 = \frac{1}{100}$, where p_i is the probability that each spices will have i offspring. For example Covid was roughly similar to this. People has a high probability p_0 to contaminate no one, but also a not-so-small probability to contaminate many people (i.e. a small number of people contributed to a significant proportion of the contamination).

(a) What is G_1 in this case?

(b) Compute μ and σ.

(c) Make a drawing representing the graph of G_1 and the first diagonal $y = x$. Indicating the drawing of the find η and use a calculator to find an approximation of η

Solution (a) Given the probabilities for the branching process, we can write the generating function $G_1(s)$ as follows:

$$G_1(s) = \frac{7}{10} + \frac{2}{10}s^3 + \frac{1}{100}s^{10} + \frac{1}{100}s^{50}$$

This function generates the probabilities of the number of offspring an individual can have, based on the provided distribution.

(b) The mean μ of the offspring distribution can be found by taking the first derivative of $G_1(s)$ and evaluating it at $s = 1$:

$$\mu = G_1'(1) = \frac{d}{ds}\left(\frac{7}{10} + \frac{2}{10}s^3 + \frac{1}{100}s^{10} + \frac{1}{100}s^{50}\right)\bigg|_{s=1}$$

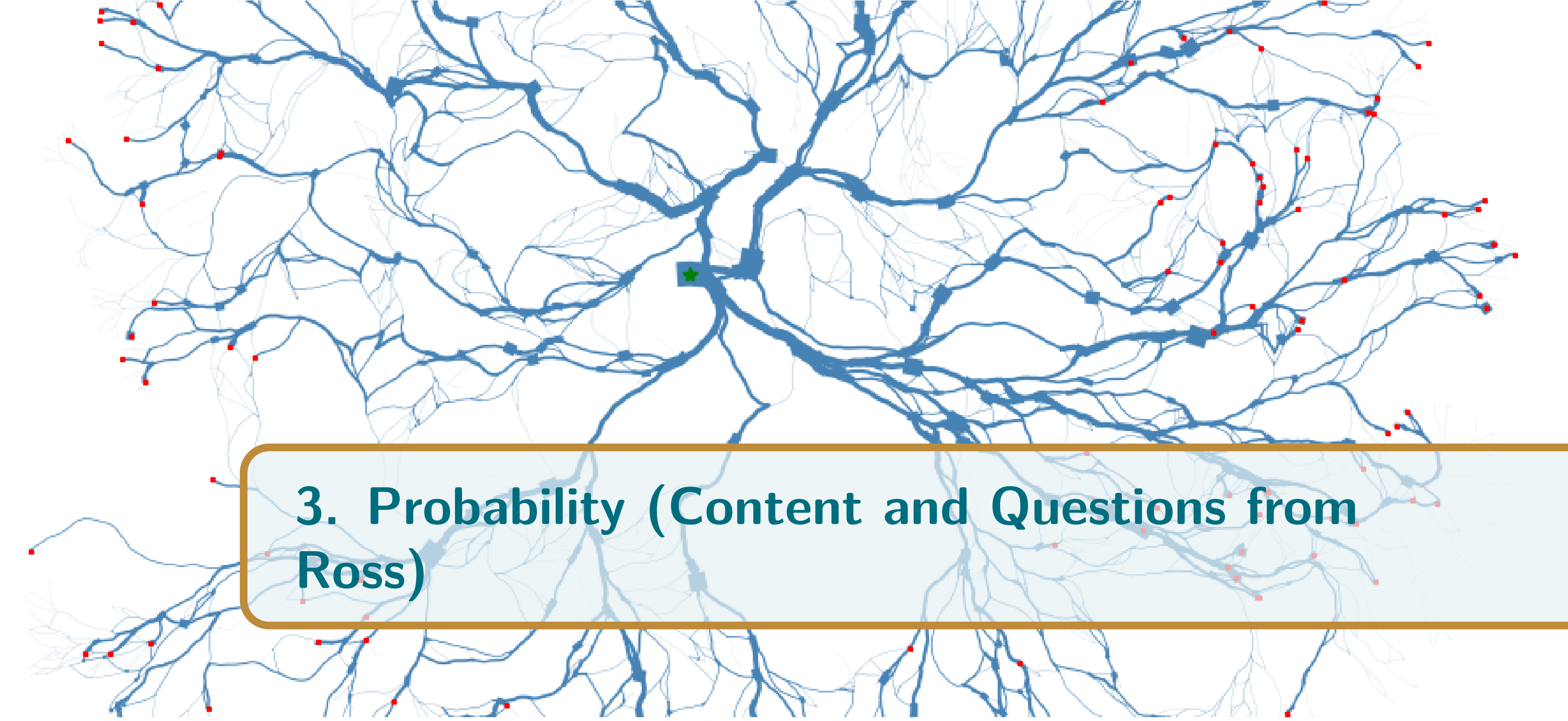
After calculation, we find that $\mu = 1.2$.

The variance σ^2 can be found by taking the second derivative of $G_1(s)$, evaluating it at $s = 1$, and using the formula $\sigma^2 = G_1''(1) + \mu - \mu^2$:

$$\sigma^2 = \frac{d^2}{ds^2}\left(\frac{7}{10} + \frac{2}{10}s^3 + \frac{1}{100}s^{10} + \frac{1}{100}s^{50}\right)\bigg|_{s=1} + \mu - \mu^2$$

After calculation, we find that $\sigma^2 = 26.36$.

(c)



3. Probability (Content and Questions from Ross)

3.1 Fundamentals

The main concept in the field of statistics and probability is the set theory. Basically all we deal with the sets. The whole theroy of statistics can be built on that. Let's discuss some foundamental concepts in statistics and then build the theory.

3.1.1 Random Experiment

To understand the meaning of random experiment, do not over think! The first thing that comes into our minds when we hear the word "random experiment" is its definition! In a nutshell, random experiment is an experiment that its outcome is unkown to us. Like:

- Tossing two coin
- Rolling a dice
- Measuring the number of possible ReadWrite operations on a piece of EEPROM chip

Do not overthink about that. Yes we can go further and discuss stuff like "we can compute the exact movement of dice or coin so it is not random but determenistic" and etc. Here I will not touch the philosophical topics that are very deep and do not necessarily converge the a unified point of view!

The random experiments can be modeled and despite the fact that a random experiment is random, we can deduce many useful information from modeling that. To model a random experiment, we use three important concepts: sample space, events, probability. In the following section, we will discuss each of them in detail.

3.1.2 Sample Space

Definition 3.1 — Sample Space. Sample space Ω is simply a set that contains *all possible outcomes* of a random experiment/

For each of random experiments described above, we can define a sample space. For example:

- Ω of Tossing Tow Coins:
$$\Omega = \{HH, HT, TH, TT\}$$
- Ω of Rolling a Dice:
$$\Omega = \{1, 2, 3, 4, 5, 6\}$$
- Ω of Rolling Two Dices:
$$\Omega = \{(1, 1), (1, 2), \ldots, (1, 6), \ldots, (6, 6)\}$$
- Ω of Number of possible ReadWrite operations on a EEPROM chip:
$$\Omega = \mathbb{N}$$

3.1.3 Events

Definition 3.2 — Events. Event E is a set of outcomes of a random experiment and is the subset of sample space Ω.
$$E \in \Omega$$

For example for any of the sample spaces specified above, we can define so many possible events. In fact any set that is a subset of the sample space is a valid event of that sample space. For example:

- Tossing Three Coins
  - There are at least on Heads:
  $$E = \{HHH, HHT, HTH, THH, HTT, THT, TTH\}$$
  - There are only two Tails:
  $$E = \{TTH, THT, HTT\}$$
- Rolling Two Dices
  - The sum of two dices is 4:
  $$E = \{(1, 3), (2, 2), (3, 1)\}$$
  - there are at least one prime number in the outcome:
  $$E = \{(1, 2), (1, 3), (1, 5), (2, 1), (3, 1), (5, 1), (2, 2), (2, 3), (2, 5), \ldots, (5, 5)\}$$

Since we have define everything on the basics of set theory, then now we can correspond the everyday concepts to specific operations in the set theory.

■ **Example 3.1** The Mapping Between Everyday Language and Sets in the Theory of Probability

- At least one of two events $A, B \in \Omega$ happens: $E = A \cup B$.
- Tow events $A, B \in \Omega$ occures at the same time: $E = A \cap B$.
- Event $A \in \Omega$ does not happen: $E = \overline{A} = \Omega - A$.
- The event A happens but B does not happen: $E = A - B$.

■

In probability and statistics, we are dealing with three important concepts: sample space Ω, event E, and probability P.

Definition 3.3 — Disjoint events. If two events has no common elements (i.e. $A \cap B = \varnothing$) then we say that two events are *disjoint*. Basically, if two sets in the venn diagram has nothing is common they are considerent to be disjoint sets.

For example for the random experiment of tossing two coins, the events 1) both coins are heads: $A = \{HH\}$ and 2) both coins are tails: $B = \{TT\}$. Two events A, B are two disjoint events. **Two events being disjoing is NOT the same as being independent**. We will talk about independet events in future.

Note that since the events are basically sets, we can use theorems of set theory to solve the problems.

Theorem 3.1 — De Morgan's Laws. If A, B are two sets then:

$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$

Proof. the proof is left as an excerise! □

3.1.4 Probability

The last foundamental ingeridient in modeling a random experiment, is to define a probability for each event. The probability should intuitively reflect how likely an event is probable to happen. This probability should satisfy some foundamental properties which are explained as follows.

Definition 3.4 — Axioms of probability (Kolmogorov axioms). Suppose that $A, B \in \Omega$ is an event and $\mathbb{P}$ is a probability function. Then $\mathbb{P}$ should satisfy the following properites:

(I) $0 \leq \mathbb{P}(A) \leq 1$

(II) $\mathbb{P}(\Omega) = 1$

(III) For the events $E_1, E_2, ..., E_n \in \Omega$ that are mutually exclusive (i.e. disjoint events):

$$\mathbb{P}(\bigcup_i E_i) = \sum_i \mathbb{P}(E_i).$$

The last property is known as the countable additivity of the probability measure.

■ **Remark** Note that we do not require the uncountable additivity property. That is because every set Ω can be written as a disjoint union of singletons $\Omega = \cup_{x \in \Omega}\{x\}$ and this leads to contradiction when Ω a continuous set (like the interval $[0, 1]$). We will see more about this later.

These axioms are called the foundamental axioms of probability and also the Kolmogorov axioms. We are free to define any kind of probabilty function that we want but it is important that 1) It should align with our common sense, 2) It should satisfy the Kolmogorov axioms.

Using the axioms above, we can observe and prove several interesting properties of the probability function. In the following box we have expressed some of them.

Theorem 3.2 — Basic Properties of the Probability Function. Suppose that $\mathbb{P}$ is a probability function and $A, B \in \Omega$ are events of the sample space Ω. We can show that the probability function has the following properties:

1. $\mathbb{P}(\varnothing) = 0$
2. If $A \subset B$ then $\mathbb{P}(A) \leq \mathbb{P}(B)$.
3. $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$.
4. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

Proof. The properties can be proved using the basic set theory theorems.

1. Since $\varnothing$ is the complement of Ω, so these two sets are disjoint (i.e. $\varnothing \cap \Omega = \varnothing$). On the other hand from the set theory we know that $\varnothing \cup \Omega = \Omega$. So $\mathbb{P}(\varnothing \cup \Omega) = \mathbb{P}(\Omega)$. On the other hand, using the third axiom we can write: $\mathbb{P}(\varnothing \cup \Omega) = \mathbb{P}(\varnothing) + \mathbb{P}(\Omega)$. Comparing the two recent equations we can conclude that $\mathbb{P}(\varnothing) = 0$.

The proofs for 2,3,4 are left as a exercise. However, the solutions can be found in the book "Statistical Modeling and Computation by Kroese" chapter 1.

□

■ **Example 3.2 — Defining a simple probability function.** Let's define a probability function for the rolling n dice experiment that is both aligned with our common sense and also satisfy the Kolmogorov equations. Suppose that the Ω is the sample space and $E \in \Omega$ is an event. Then let's define:

$$\mathbb{P}(E) = \frac{|E|}{|\Omega|}$$

in which the $|E|$ means the cardinality (number of elements) of the set E.

■

Utilizing the properties of the probability function, we can derive some very important notions, one of which is reflected in the following proposition.

Proposition 3.1 — Conditional expansion - Law of total probabilities. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\mathfrak{F}$ be a finite collection of events $\mathfrak{F} = \{F_1, F_2, \cdots, F_n\}$ that partitions Ω. I.e.

(i) $F_i \cap F_j = \varnothing \qquad i \neq j$,

(ii) $\bigcap_i F_i = \Omega$.

Let $E \in \mathcal{F}$ be any nonempty event. Then we can write

$$\mathbb{P}(E) = \sum_i \mathbb{P}(E|F_i)\mathbb{P}(F_i).$$

Proof. Since $\mathfrak{F}$ partitions Ω and $E \neq \varnothing$, then $\{E \cap F_i\}_i$ is a partition of E. Thus

$$\mathbb{P}(E) = \mathbb{P}(\bigcup_i (E \cap F_i)) = \sum_i \mathbb{P}(E \cap F_i) = \sum_i \mathbb{P}(E|F_i)\mathbb{P}(F_i).$$

This completes the proof. □

In dealing with random variables, either continuous or discrete, using the notion of the law of total probabilities helps us to simplify some of the calculations significantly. The following examples are some places that we use this idea to simplify calculations by a lot.

■ **Example 3.3** Let $X_1, X_2, X_3, \cdots$ be i.i.d. real-valued random variable. Let T be a positive integer valued random variable. Define the the real-valued random variable N as

$$N = \sum_{i=1}^{T} X_i.$$

What is the probability generating function for N.

Solution For the generating probability function we have

$$G_N(s) = \mathbb{E}\Big[s^N\Big] = \mathbb{E}\Big[s^{X_1+X_2+\cdots+X_T}\Big].$$

The problem in evaluating the expression above is that the number of random variables X_i to be summed up is also a random variable. So the first step is to make this a non-random variable buy conditional expansion.

$$G_N(s) = \mathbb{E}\Big[s^N\Big] = \sum_{i\in\mathbb{N}} \mathbb{E}\Big[s^{X_1+\cdots+X_T}\Big|T=i\Big]\mathbb{P}(T=i) = \sum_{i\in\mathbb{N}} \mathbb{E}\Big[s^{X_1+\cdots+X_i}\Big]\mathbb{P}(T=i)$$

Since X_i are all i.i.d., then we can write

$$G_N(s) = \sum_{i\in\mathbb{N}} (\mathbb{E}\Big[s^{X_1}\Big])^n \mathbb{P}(T=i) = G_T(G_{X_1}(s)).$$

■

■ **Example 3.4** Let X, Y be two independent random variables. Define $Z = X + Y$. Find the PDF of Z.

Solution First, We need to find $F_Z(z) = \mathbb{P}(Z < z)$. For this we can write

$$F_Z(z) = \mathbb{P}(X + Y < z)$$

Again, we can use conditional expansion to write

$$F_Z(z) = \int_{\mathbb{R}} \mathbb{P}(X+Y<z\Big|Y=y) f_Y(y)\ dy = \int_{\mathbb{R}} \mathbb{P}(X<z-y) f_Y(y)\ dy = \int_{\mathbb{R}} F_X(z-y) f_Y(y)\ dy.$$

Then differentiating F_z with respect to z we will get the PDF

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \int_{\mathbb{R}} f_X(z-y) f_Y(y)\ dy = (f_X * f_Y)(z).$$

■

■ **Example 3.5** Let X, Y be two real valued random variable, not necessarily independent. Calculate $\mathbb{P}(X < Y)$.

Solution To calculate this we can again use the law of total probabilities. In particular

$$\mathbb{P}(X < Y) = \int_{\mathbb{R}} \mathbb{P}(X < y) f_Y(y)\ dy = \int_{\mathbb{R}} F_{X|Y}(y) f_Y(y)\ dy.$$

■

3.1.5 Isomorphism between random experiments

Often, there is this intuition that certain random experiments are really the same, although they might look very different from each other. For instance, consider two random experiments. In one, we are playing a dice successively and asking what is the probability that after 5 plays, 1 is not appeared. The second experiment is that we have 6 Urns and we place balls in them successively, i.e. at each step one ball is placed in one of the urns and the chance of a ball to end up in any of the urns in equal. These two experiment, although very different, but looks very similar. There is one way that we can formalize this wage intuition, and that is the notion of isomorphism between sets. We say two sets are isomorphic if there is a bijection between them. And the reason that the previously mentioned experiments feel the same is that the sample space Ω of these two experiments are in fact isomorphic.

3.2 Random Variables

Often, we are interested in the some measurements of the outcome of a random experiment rather than knowing the outcome it self. For instance, if the experiment of tossing two dice, we might be interested in asking the question if the sum of two dice is 6, and not concerned over whether the actual outcome was (3,3) or (2,4), etc. These quantities of interest are called random variables. The following definition put this into a more formal definition.

Definition 3.5 Let $(\Omega, E, \mathbb{P})$ be a probability space. Then a random variable X is a function $X : \Omega \to S$, where S called the state space.

■ **Remark** The state space S must have some properties, i.e. being measurable, etc. You can read more about this on the Wikipedia of random variables. Also, the state space S if often $\mathbb{R}$, or in the case of a discrete time Markov chain, S is a finite set (that can be the edge set of a graph).

Since the value of a random variable is determined by the outcomes of the random experiment, we can assign probabilities to the possible values of the random variable. We use the following notation for this purpose.

Definition 3.6 — Notation for probability of random variables. Let X be a random variable. Then we define event

$$E = \{X = a\} = \{\omega \in \Omega : X(\omega) = a\}.$$

Then the following notations are usually used interchangeably:

$$\mathbb{P}(X = a) = \mathbb{P}(\{X = a\})$$

both of which is simply $\mathbb{P}(E)$.

■ **Example 3.6** Let X be a random variable defined to be the sum of two fair dice. Then

$$\begin{aligned}
&\mathbb{P}(\{X = 2\}) = \mathbb{P}(\{(1,1)\}) = \frac{1}{36}, \\
&\mathbb{P}(\{X = 3\}) = \mathbb{P}(\{(1,2),(2,1)\}) = \frac{2}{36}, \\
&\mathbb{P}(\{X = 13\}) = \mathbb{P}(\varnothing) = 0.
\end{aligned}$$

■

■ **Example 3.7** Suppose that we toss a coin having probability p of coming up heads. We continue tossing the coin until we see a heads. Let the random variable N be the number of times we toss the coin. Describe this random variable.

Solution Although, we can always solve this kind of questions in an ad hoc way by just simply following our intuition, but it is always a best practice to try to fine tune our abstract thinking with our intuitive understandings in these kind of example. Then we can use of abstract thinking capability to solve problems that are almost impossible to address by solely depending on the intuition. So, it is a good idea to try to see how does the set Ω look like. The set Ω will be the set of all finite string of all T letters terminated with H. In other words

$$\Omega = \{H, TH, TTH, TTTH, TTTTH, \cdots\}.$$

Then the random variable $N : \Omega \to \mathbb{Z}$ is basically the length of the string. For instance, if $\omega = TTH \in \Omega$, then $N(\omega) = 3$. Let's calculate

$$\mathbb{P}(N = 3) = \mathbb{P}(\{\omega \in \Omega : N(\omega) = 3\}).$$

To solve this, we need to define appropriate events and then condition our probability on those events. Define F_n be the event where the n first outcomes are tails. For instance

$$F_1 = \{TH, TTH, TTTH, \cdots\},\ F_2 = \{TTH, TTTH, TTTTH, \cdots\},\ \cdots.$$

And let $E = \{N = 3\} = \{TTH\}$. Then we can condition $\mathbb{P}(E)$ on F_2

$$\mathbb{P}(E) = \mathbb{P}(E|F_2)\mathbb{P}(F_2) + \mathbb{P}(E|F_2^c)\mathbb{P}(F_2^c).$$

Note that $F_2^c = \{H, TH\}$, this $\mathbb{P}(E|F_2^c) = \mathbb{P}(E \cap F_2^c)/\mathbb{P}(F_2^c) = 0$. Now we need to determine $\mathbb{P}(F_2)$. Again, we can condition this event on F_1. Then we can write

$$\mathbb{P}(F_2) = \mathbb{P}(F_2|F_1)\mathbb{P}(F_1) + \mathbb{P}(F_2|F_1^c)\mathbb{P}(F_1^c).$$

with the same argument as above $\mathbb{P}(F_2|F_1^c) = 0$. Combining these equations we will get

$$\mathbb{P}(E) = \mathbb{P}(E|F_2)\mathbb{P}(F_2|F_1)\mathbb{P}(F_1).$$

Now these probabilities are easy to calculate which leads to the final answer

$$\mathbb{P}(E) = (1-p)(1-p)p.$$

And by induction we can conclude

$$\mathbb{P}(\{N = n\}) = (1-p)^n p.$$

■

■ **Example 3.8** Suppose that independent trials, each of which results in m possible outcomes with respective probabilities $p_1, p_2, \ldots, p_m$ such that $\sum_{i=1}^m p_i = 1$. Are continually performed. Let X be the number of trials needed until each outcome has occurred at least once. Describe the properties of this random variable.

Solution It is sometime a good idea to try to imagine what does the sample space look like. Let $\Sigma = \{s_1, s_2, s_3, \ldots, s_m\}$ be a set of m distinct symbols. Then each time we are continually performing the experiment, we are getting each of these symbols with corresponding probability p_m. Thus the sample space will be the set of all infinite sequences of these symbols. In other words

$$\Omega = \{\text{all infinite sequence of symbols from } \Sigma\}.$$

Then the random number $X(\omega)$ for $\omega \in \Omega$ is basically the length of the prefix string of ω in which any of the symbols in Σ has been occurred at least once.

■

3.2.1 Cumulative Distribution of Random Variable

The notion of the cumulative distribution of a random variable comes handy in most of the future calculations. Also, this distribution can be used to derive other notions of distributions what are extremely important in applications.

Definition 3.7 — Cumulative distribution. Let X be a random variable $X : \Omega \to \mathbb{R}$. Then the cumulative distrubition $F : \mathbb{R} \to \mathbb{R}$ is defined as

$$F(x) = \mathbb{P}(\{X \leq x\}).$$

Proposition 3.2 The cumulative distribution of a random variable has the following properties.

(i) $\mathbb{P}(a < X \leq b) = F(b) - F(a)$.

(ii) $F(x)$ is a non-decreasing function of x.

Proof. (i)

$$\mathbb{P}(\{a < X \leq b\}) = \mathbb{P}(\{X \leq b\} \cap \{X \leq a\}^c) = -\mathbb{P}(\Omega) + \mathbb{P}(\{X \leq b\}) + \underbrace{\mathbb{P}(\{X \leq a\}^c)}_{1-\mathbb{P}(\{X \leq a\})} = F(b) - F(a).$$

(ii) Let $b_1, b_2 \in \mathbb{R}$ and $b_1 \leq b_2$. Then $\{X \leq b_1\} \subseteq \{X \leq b_2\}$. This implies

$$\mathbb{P}(\{X \leq b_1\}) \leq \mathbb{P}(\{X \leq b_2\}) \implies F(b_1) \leq F(b_2).$$

This implies that $F(x)$ is a non-decreasing function.

□

3.3 Probability Generating Function

In this section we will go through the details of the probability generating function. We start with the following definition.

Definition 3.8 — Probability Generating Function. Let X be a random variable with state space $S = \mathbb{Z}_+$. Then the probability generating function for this random variable is a function define as

$$G_X(s) = \mathbb{E}\left[s^X\right] = \sum_{x \in S} s^x \mathbb{P}(X = x).$$

In different areas of mathematics, we often can define something algebraic that is very easy to handle (like differentiation, etc) and carries the important information of the object under study. One of these algebraic symbolic objects is the Tutte polynomial, Chromatic polynomial, matching polynomial, etc. These polynomials are kind of modeling the object under study with tools that are easy to handle. The probability generating function is one of those symbolic objects. Because of the way that is crafted, it carries most of the information about the random variable, while the actual object as a function might have poor properties. This will be more clear in the following proposition. In a nutshell, the probability generating function is more of a symbolic thing rather than actual function with meaning full properties. That is why we generally evaluate this function (and its derivatives) at point 0 or 1.

Proposition 3.3 — Properties of the probability generating function. Let X be a random variable, and $G_X(s)$ its probability generating function. Then we have

(i) $G_X(1) = 1$.

(ii) $\mathbb{E}\left[X\right] = G'_X(1)$.

(iii) $\delta X = G''_X(1) - G'_X(1)^2 + G'_X(1)$

(iv) Let X, Y be independent random variables. Then we have

$$G_{X+Y}(s) = G_X(s)G_Y(s).$$

(v) Let $X_1, X_2, \cdots$ be iid random variables, and N be a random variable taking values in $\mathbb{Z}_+$. Define $T = X_1 + X_2 + \cdots + X_N$. Then we have

$$G_T(s) = (G_N \circ G_{X_1})(s).$$

Proof. The proof for part i,ii, iii, and iv basically follows immediately from the definition. So we will only provide the proof for part iv.
T is the sum of N iid random variables where N is itself a random variable. We can make it a normal variable by using the law of total probabilities.

$$G_T(s) = G_{\sum_i^N X_i}(s) = \sum_{n=0}^{\infty} G_{\sum_i^n X_i}(s)\mathbb{P}(N = n) = \sum_{n=0}^{\infty}(G_{X_1})^n \mathbb{P}(N = n) = G_N(G_{X_1})(s)$$

and this completes the proof. □

The item (iv) in the proposition above is very important, as it makes the hard calculations easy to do. See the following example for more details.

■ **Example 3.9** We select a number N from $\{1, 2, 3, \cdots, 100\}$ randomly and then generate N random numbers $X_1, X_2, \cdots X_N$ from the distribution $\mathrm{Unif}[0, 1]$. Then we compute $T = X_1+X_2+\cdots+X_N$. What is the average of T?

Solution We know that

$$\mathbb{E}\left[T\right] = G'_T(1).$$

Thus we need to calculate the probability generating function $G_T(s)$. From part (iv) of the proposition above we know that $G_T = G_N \circ G_{X_1}$. Thus we will have

$$G'_T = G'_{X_1} G'_N \circ G_{X_1}.$$

Thus evaluating at $s = 1$ we will have

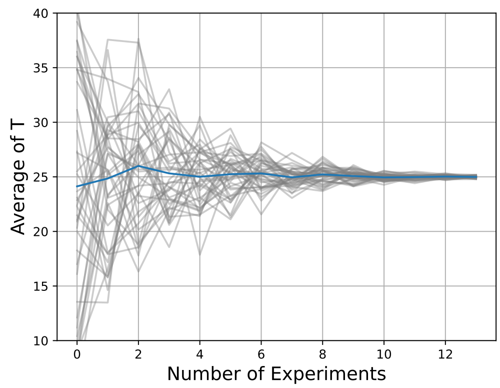
$$G'_T(1) = G'_{X_1}(1) G'_N(\underbrace{G_{X_1}(1)}_{1}) = \mathbb{E}\left[X_1\right]\mathbb{E}\left[N\right].$$

On the other hand we have $\mathbb{E}\left[N\right] = 50$ and $\mathbb{E}\left[X_1\right] = 1/2$. Then

$$\mathbb{E}\left[T\right] = 25.$$

The following figure shows this fact (i.e. convergence of the average value of T to 25 when we increase the number of experiments.)

■



3.4 Some more deep notes

This section has the definitions and discussions that I leaned through a course on the rigorous probability.

Definition 3.9 — Algebra or Field of a Set. A field of sets is a pair $(X, \mathcal{F})$, where X is a set and $\mathcal{F}$ is collection of subsets of X such that

(i) $\varnothing \in \mathcal{F}$,
(ii) $A \in \mathcal{F} \implies A^c \in \mathcal{F}$,
(iii) $A_1, A_2 \in \mathcal{F} \implies A_1 \cup A_2 \in \mathcal{F}$.

■ **Remark** It follows from the definition above that $\Omega \in \mathcal{F}$ and also $A_1 \cap A_2 \in \mathcal{F}$ if $A_1, A_2 \in \mathcal{F}$.

Note that in the notion of the algebra of a set, the algebra is closed only for finite intersections or finite unions.

Definition 3.10 — σ-algebra of a set. The σ-algebra of a set is an algebra of the set such that it is also closed under countable union and intersection.

Definition 3.11 — Semi-algebra. The collection $\mathcal{I}$ is a semi-algebra of the subsets of Ω if

(i) $\varnothing \in \mathcal{I}$,

(ii) $A_1, A_2 \in \mathcal{I} \implies A_1 \cap A_2 \in \mathcal{I}$,

(iii) For $A \in \mathcal{I}$, $A^c = \dot{\cup}_{i=1}^n A_i$ for some $n \in \mathbb{N}$.

Theorem 3.3 — Extension Theorem. Let Ω be a set, and $\mathcal{I}$ be a semi-algebra of the subsets of it. Let $P : \mathcal{I} \to [0, 1]$ such that is satisfies the following properties
(i) $P(\varnothing) = 0$,
(ii) $P(\Omega) = 1$,
(iii) $P(\dot{\cup}_{i=1}^n A_i) \geq \sum_{i=1}^n P(A_i)$ for $\quad A_1, \cdots, A_n, \dot{\cup}_{i=1}^n A_i \in \mathcal{I}$.
(iv) $P(A) \leq \sum_i P(A_i)$ for $A_1, \cdots \in I, A \subseteq \cup_i A_i$.
Then there exist a valid probability space $(\Omega, \mathcal{M}, \mathbb{P}^*)$ such that $\mathcal{M} \supset \mathcal{I}$ and $\mathbb{P}(A) = \mathbb{P}^*(A)$ for all $A \in \mathcal{I}$.

■ **Remark** By "there exists a valid probability space" we mean that $\mathcal{M}$ is a σ-algebra and $\mathbb{P}^*$ is a countably additive probability measure. Also, it is worth mentioning that the property (*iii*) is called *finite super additivity*.

In the following propositions we will see more easy-to-check characterizations of this theorem.

3.5 Summary and Tricks

Summary 3.1 — Improving finite additivity to countable additivity. In question 3.18 we improve the finite additivity property of $\mathbb{P}^*$ to countable additivity, and we used two facts for that purpose. First, use the monotnonicity property, and second, use the countable sub additivity property. I.e. if we know that for $A_1, A_2, \cdots, A_n \in \mathcal{M}$ disjoint we have

$$\mathbb{P}^*(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n \mathbb{P}(A_i).$$

Then let $A_1, \cdots$ be a countable collection of disjoint sets in $\mathcal{M}$. So for any $m \in \mathbb{N}$ we have

$$\sum_{n<m} \mathbb{P}^*(A_n) = \mathbb{P}^*(\bigcup_{n<m} A_n) \leq \mathbb{P}^*(\bigcup_n A_n),$$

where for the last inequality we used the monotonicity property of $\mathbb{P}^*$. Since this is true for all $m \in \mathbb{N}$ we can conclude that

$$\sum_n A_n \mathbb{P}^*(A_n) \leq \mathbb{P}^*(\bigcup_n A_n).$$

Now form the countable subadditivity we have

$$\sum_n A_n \mathbb{P}^*(A_n) \geq \mathbb{P}^*(\bigcup_n A_n).$$

This implies

$$\sum_n A_n \mathbb{P}^*(A_n) = \mathbb{P}^*(\bigcup_n A_n).$$

Summary 3.2 — Quantifiers to Cup and Cap!. Let $\{A_i\}_{i\in I}$ be a collection of sets. Then $x \in \bigcup_i A_i$ means that $\exists n \in I$ such that $x ß A_n$. Similarly, $x \in \bigcap_i A_i$ means that $\forall n \in I$ we have $x \in A_n$. According to this, we can construct much more complicated statements. For instance,

$$x \in \bigcup_{n=1}^{\infty} \bigcap_{k>n}^{\infty} \bigcup_{m>k}^{\infty} A_k$$

in words meas that

$$\exists n \in \mathbb{N} \text{ s.t. } \forall k > n \; \exists m > k \text{ s.t. } x \in A_k.$$

This makes these statements more easier to follow. We have used this in Problem 3.26.

Summary 3.3 — Pre-image v.s. Image. The pre-image preserves the union, intersection, and complement, i.e.

$$f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B), \quad f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B), \quad f^{-1}(A^c) = (f^{-1}(A))^c.$$

But the same is NOT true in the case of image. image preserves the union but not necessarily the intersection, or complement (for which we need extra properties like being injective or surjective).

3.6 Solved Problems

■ **Problem 3.1 — From Ross.** Ben can talk a course in compute science or chemistry. If she takes the computer science course, then she will get A grade with probability $\frac{1}{2}$. If she takes the chemistry course, then she will get A grade with probability $\frac{1}{3}$. She decides to base her decision on the flip of a fair coin. What is the probability that she gets an A in chemistry?

Solution We define the following events

A: she will get an A grade.
CO: she will take the computer science course.
CH: she will take the chemistry course.

Then the question is basically asking for $\mathbb{P}(A \cap CH)$. We can compute it by

$$\mathbb{P}(A \cap CH) = \mathbb{P}(A|CH)\mathbb{P}(CH) = \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}.$$

■ **Problem 3.2** And urn contains seven black balls and five white balls. We draw two times from the urn. Given that the each ball has the same probability to be drawn, what is the probability that both balls drawn are black?

Solution This question nicely demonstrates the fact that there are many ways to define the event spaces, and not all of them are very useful in computing the desired probability. Define

E: two drawn balls are black.

The question is in fact asking $\mathbb{P}(E)$. But this even is not very useful in any progress with the solution. Thus we need to define some finer events

E_1: The first drawn ball is black.
E_2: The second drawn ball is black.

It is clear that $E = E_1 \cap E_2$. These two finer events allows us to compute the probability of interest given the data we have in our hand.

$$\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_2|E_1)\mathbb{P}(E_1) = \frac{6}{11} \cdot \frac{7}{12}$$

■ **Problem 3.3 — From Ross.** Three men at a party through their hats into the center of the room, and then, after mixing the hats, each pick a hat randomly. What is the probability if non of them get their own hat back.

Solution There are a million ways to tack a probability problem. We can construct a suitable sample space and then compute the probabilities explicitly, or we can use the properties of the probability function to computer the desired probability without any need to construct the sample space. Here, we will demonstrate two ways.

Solving the problem by utilizing the properties of the probability function. First we need to define some suitable events. There are again many ways to define event sets and each have their own pros and cons. We proceed with the following definition.

E_i: The person i "selects" his own hat.

Also, with this particular construction of the event sets, it is much more easier to compute the complementary probability of the desired probability first and then compute the desired one by simply subtracting it from 1. The complement of the event "no men gets his own hat back" is "at least one man gets his hat back" which is $\mathbb{P}(E_1 \cup E_2 \cup E_3)$. To compute the terms of this we first need to calculate $\mathbb{P}(E_i)$, $\mathbb{P}(E_i \cap E_j)$ where $i \neq j$ and also $\mathbb{P}(E_1 \cap E_2 \cap E_3)$. We know that $\mathbb{P}(E_i) = 1/3$ for $i = 1, 2, 3$. That is because it is equally likely he selects any of the hats at the center. For $\mathbb{P}(E_i \cap E_j)$ we can write

$$\mathbb{P}(E_i \cap E_j) = \mathbb{P}(E_i|E_j)\mathbb{P}(E_j) = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}.$$

In which we used the fact that $\mathbb{P}(E_i|E_j)$ is $\frac{1}{2}$ for distinct i, j. That is because given person j selects his hat correctly, then there are two possibilities for E_i to select his hat (he can pick the correct one or the wrong one). Lastly for $\mathbb{P}(E_1 \cap E_2 \cap E_3)$ we write

$$\mathbb{P}(E_1 \cap E_2 \cap E_3) = \mathbb{P}(E_1|E_2 \cap E_3)\mathbb{P}(E_2 \cap E_3) = \mathbb{P}(E_1|E_2 \cap E_3)\mathbb{P}(E_2|E_3)\mathbb{P}(E_3) = 1 \cdot \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}.$$

Thus

$$\mathbb{P}(E_1 \cup E_2 \cup E_3) = (1) - (1/2) + (1/6) = \frac{4}{6}.$$

Then the probability of interest will be

$$\mathbb{P}(E) = 1 - \frac{4}{6} = \frac{1}{3}.$$

Solving by constructing a sample space. A suitable sample space for this problem can be the set of all permutations on three letters. This set is

$$\Omega = \{\begin{pmatrix} a & b & c \\ \boxed{a} & \boxed{b} & \boxed{c} \end{pmatrix}, \begin{pmatrix} a & b & c \\ \boxed{a} & c & b \end{pmatrix}, \begin{pmatrix} a & b & c \\ b & a & \boxed{c} \end{pmatrix}, \begin{pmatrix} a & b & c \\ b & c & a \end{pmatrix}, \begin{pmatrix} a & b & c \\ c & a & b \end{pmatrix}, \begin{pmatrix} a & b & c \\ c & \boxed{b} & a \end{pmatrix}\}.$$

Note that the elements in the box represents the fixed point of the permutation. The probability of interest is basically the number of permutations that has no fixed point. As it is clear from the set Ω, the probability is

$$\mathbb{P}(E) = \frac{2}{6} = \frac{1}{3}.$$

■ **Problem 3.4 — Conditional probability mass function (from Ross).** Let X, Y be two random variables with the joint probability mass function given as

$$P(1,1) = 0.5 \qquad P(1,2) = 0.1, \qquad P(2,1) = 0.1, \qquad P(2,2) = 0.3.$$

Calculate the conditional probability mass function of X given that $Y = 1$.

Solution We will use the following identity

$$P_{X|Y}(x|y) = \mathbb{P}(X = x|Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}.$$

Observe that

$$\mathbb{P}(Y=y)=\sum_x \mathbb{P}(Y=y,X=x)$$

thus $\mathbb{P}(Y=1)=0.5+0.1=0.6$. So we will have

$$P_{X|Y}(1|1)=\frac{0.5}{0.6}=\frac{5}{6},\qquad P_{X|Y}(2|1)=\frac{0.1}{0.6}=\frac{1}{6}.$$

■ **Problem 3.5 — Conditional probability mass function for geometric random variables (from Ross).** Let X_1, X_2 be two independent random variables with geometric distributions with parameters (n_1, p) and (n_2, p). Calculate the conditional probability mass function of X_1 given that $X_1 + X_2 = m$.

Solution First, observe that $Y = X_1 + X_2$ is a binomial distribution with parameter $(n_1 + n_2, p)$. Thus we can write

$$P_{X_1|Y}(k|m)=\mathbb{P}(X_1=k|Y=m)=\frac{\mathbb{P}(X_1=k,X_1+X_2=m)}{\mathbb{P}(X_1+X_2=m)}=\frac{\mathbb{P}(X_1=k,X_2=m-k)}{\mathbb{P}(Y=m)}$$

Since the random variables X_1 and X_2 are independent, we can write

$$P_{X_1|Y}(k|m)=\frac{\mathbb{P}(X_1=k)\mathbb{P}(X_2=m-k)}{\mathbb{P}(Y=m)}=\frac{\binom{n_1}{k}\binom{n_2}{m-k}}{\binom{n_1+n_2}{m}}.$$

■ **Problem 3.6 — Conditional probability mass function for Poisson random variables (from Ross).** Let X, Y be two independent Poisson random variables with parameters λ_1 and λ_2 respectively. Calculate the conditional probability mass function for X given that $X_1 + X_2 = n$.

Solution First observe that $Z = X + Y$ is a Poisson random variable with parameter $\lambda_1 + \lambda_2$. Thus we will have

$$P_{X|X+Y}(m|n)=\frac{\mathbb{P}(X=m|X+Y=n)}{\mathbb{P}(X+Y=n)}=\frac{\mathbb{P}(X=m,Y=n-m)}{\mathbb{P}(X+Y=n)}$$

Given that X, Y are independent random variables then we can write

$$P_{X|X+Y}(m,n)=\frac{\mathbb{P}(X=m)\mathbb{P}(Y=n-m)}{\mathbb{P}(X+Y=n)}=\frac{\lambda_1^n\lambda_2^{n-m}n!}{m!(n-m)!(\lambda_1+\lambda_2)^n}=\binom{n}{m}(\frac{\lambda_1}{\lambda_1+\lambda_2})^m(\frac{\lambda_2}{\lambda_1+\lambda_2})^{n-m}$$

Thus the conditional probability mass function of X given that $X + Y = n$ will be a binomial random variable with parameter $(n, \lambda_1/(\lambda_1 + \lambda_2))$. We can now easily compute the conditional expectation value as

$$\mathbb{E}\Big[X|X+Y=n\Big]=\frac{n\lambda_1}{\lambda_1+\lambda_2}$$

■ **Problem 3.7** Let X, Y be two discrete random variables. Prove that

$$\mathbb{E}\Big[\mathbb{E}\Big[X|Y\Big]\Big]=\mathbb{E}\Big[X\Big].$$

Solution We start with the definition of the expectation of a discrete random variable.

$$\begin{aligned}\mathbb{E}\Big[\mathbb{E}\Big[X|Y\Big]\Big]&=\sum_y \mathbb{E}\Big[X|Y=y\Big]\mathbb{P}(Y=y)=\sum_y\sum_x x\mathbb{P}(X=x|Y=y)\mathbb{P}(Y=y)\\&=\sum_{x,y}x\mathbb{P}(X=x,Y=y)=\sum_x x\sum_y\mathbb{P}(X=x,Y=y)=\sum_x x\mathbb{P}(X=x)=\mathbb{E}\Big[X\Big]\end{aligned}$$

■ **Problem 3.8 — The expectation of a random number of random variables (from Ross).** Let the expected number of injures in an industrial field be 4 per week. Also, assume that the number of workers injured at each incidence are independent random variables with average 2. Then what is the expected number of injuries in one week?

Solution Let $X_1, X_2, \cdots$ be i.i.d random variables representing the number of workers injured at each incidence. We are interested in

$$\mathbb{E}\Big[X_1 + \cdots + X_N\Big]$$

where N is a random variable representing the number of incidences occurred in a week. By the law of conditional expectation we can write

$$\mathbb{E}\Big[X_1 + \cdots + X_N\Big] = \sum_n \mathbb{E}\Big[X_1 + \cdots + X_n\Big]\mathbb{P}(N = n) = \sum_n n\mathbb{E}\Big[X\Big]\mathbb{P}(N = n) = \mathbb{E}\Big[X\Big]\mathbb{E}\Big[N\Big].$$

Thus the average number of workers injured in a week will be 8.

■ **Problem 3.9 — An alternative way to compute the expectation of a geometric random variable.** Consider a coin with probability p to fall heads. What is the expectation value of the number of tosses required until we get the first head?

Solution Let $X_1, X_2, \cdots$ be Bernoulli random variables with parameter p. Let N be a random variable denoting the number of tosses required until we get the first heads. We can condition the expected value of E to the first outcome.

$$\mathbb{E}\Big[N\Big] = \mathbb{E}\Big[N|X_1 = H\Big]\underbrace{\mathbb{P}(X_1 = H)}_{=p} + \mathbb{E}\Big[N|X_1 = T\Big]\underbrace{\mathbb{P}(X_1 = T)}_{=1-p}$$

Observe that

$$\mathbb{E}\Big[N|X_1 = H\Big] = 1, \qquad \mathbb{E}\Big[N|X_1 = T\Big] = 1 + \mathbb{E}\Big[N\Big].$$

Thus we will have

$$\mathbb{E}\Big[N\Big] = \frac{1}{p}.$$

■ **Problem 3.10 — Trapped miner (from Ross).** A miner is trapped in the mine and has three doors in front of him. He is equally likely to choose any of the three. The first door will take him to safety after 2 hours of walking, the second door will take him to the mine again after 3 hours of walking, and the third door will take him to the mine again after 5 hours of walking. What is the expected time that the miner will arrive to safety?

Solution Let $X_1, X_2, \cdots$ be random variables denoting the doors that the miner choose at each time that he attempts to escape. Furthermore, let T be a random variable showing the the time it takes for the miner to escape. To calculate $\mathbb{E}\Big[T\Big]$ we can condition it on the first door choice. I.e.

$$\mathbb{E}\Big[T\Big] = \mathbb{E}\Big[T|X_1 = 1\Big]\mathbb{P}(X_1 = 1) + \mathbb{E}\Big[T|X_1 = 2\Big]\mathbb{P}(X_1 = 2) + \mathbb{E}\Big[T|X_1 = 3\Big]\mathbb{P}(X_1 = 3)$$

Observe that

$$\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = 2) = \mathbb{P}(X_1 = 3) = 1/3.$$

Also

$$\mathbb{E}\Big[T|X_1 = 1\Big] = 2, \qquad \mathbb{E}\Big[T|X_1 = 2\Big] = 3 + \mathbb{E}\Big[T\Big], \qquad \mathbb{E}\Big[T|X_1 = 3\Big] = 3 + \mathbb{E}\Big[T\Big].$$

Thus we will have

$$\mathbb{E}\Big[T\Big] = 10.$$

So on average it will take the miner to exit the mine in 10 hours. Note that this does not guarantee that the miner will eventually escape. It is possible that we will get in trap by repeatedly choosing the door number 3.

■ **Problem 3.11 — From Rosenthal.** Suppose that $\Omega = \{1, 2\}$, with $\mathbb{P}(\varnothing) = 0$ and $\mathbb{P}(\{1, 2\}) = 1$. Suppose $\mathbb{P}(\{1\}) = \frac{1}{4}$. Prove that $\mathbb{P}$ is countably additive if and only if $\mathbb{P}(\{2\}) = \frac{3}{4}$.

Solution The proof has two parts

$\boxed{\Longrightarrow}$ Since $\mathbb{P}$ is countably additive, then

$$\mathbb{P}(\{1\}\dot{\cup}\{2\}) = \mathbb{P}(\{1\}) + \mathbb{P}(\{2\}) = 1.$$

This implies $\mathbb{P}(\{2\}) = 3/4$.

$\boxed{\Longleftarrow}$ Assume $\mathbb{P}(\{2\}) = 3/4$. Then it is very straightforward to check that for every disjoint subset $A, B \subset \Omega$, we have

$$\mathbb{P}(A\dot{\cup}B) = \mathbb{P}(A) + \mathbb{P}(B).$$

Thus we conclude that $\mathbb{P}$ is countably additive.

■ **Problem 3.12 — From Rosenthal.** Suppose $\Omega = \{1, 2, 3\}$ and $\mathcal{F}$ is a the collection of all subsets of Ω. Find (with proof) necessary and sufficient conditions on the real numbers x, y, z such that there exists a countably additive probability measure $\mathbb{P}$ on $\mathcal{F}$ such that $x = \mathbb{P}\{1, 2\}, y = \mathbb{P}\{2, 3\}, z = \mathbb{P}\{1, 3\}$.

Solution To find the necessary conditions, we assume that $\mathbb{P}$ is an additive probability measure. Let $a = \mathbb{P}\{1\}, b = \mathbb{P}\{2\}$, and $c = \mathbb{P}\{3\}$. Then the countable additivity implies

$$a + b = x, \qquad b + c = y, \qquad a + c = z.$$

Then due to countable additivity, and the fact that $\mathbb{P}$ is a probability measure (i.e. $\mathbb{P}\{1, 2, 3\} = 1$), we have $a + b + c = 1$, thus we need to have

$$x + y + z = 2. \tag{♪}$$

Further, we solve the a, b, c in terms of x, y, z are require the singleton probabilities to be positive. We have

$$a = \frac{x - y + z}{2}, \qquad b = \frac{x + y - z}{2}, \qquad c = \frac{-x + y + z}{2}.$$

One of the necessary conditions is also to have

$$x - y + z \geq 0, \qquad x + y - z \geq 0, \qquad -x + y + z \geq 0. \tag{♫}$$

The two conditions (♪) and (♫) together are the necessary and sufficient conditions for $\mathbb{P}$ to be a valid probability measure.

■ **Problem 3.13 — From Rosenthal.** Suppose that $\Omega = \mathbb{N}$ is the set of positive integers, and $\mathbb{P}$ is defined for all $A \subseteq \Omega$ by $\mathbb{P}(A) = 0$ if A is finite, and $\mathbb{P}(A) = 1$ if A is infinite. Is $\mathbb{P}$ finitely additive?

Solution Not it is not. Consider the partitioning of the set Ω by the even E and odd O integers.

$$\mathbb{P}(\Omega) = \mathbb{P}(E) + \mathbb{P}(O) \implies 1 = 2,$$

which is not true. Thus $\mathbb{P}$ is not finitely additive.

■ **Problem 3.14 — From Rosenthal.** Suppose that $\Omega = \mathbb{N}$, and $\mathbb{P}$ is defined for all $A \subseteq \Omega$ by $\mathbb{P}(A) = |A|$ if A is finite, and $P(A) = \infty$ if A is infinite. This P is of course not a probability measure (in fact it is counting measure), however we can still ask the following: (be the convention $\infty + \infty = \infty$)

(I) Is $\mathbb{P}$ finitely additive?

(II) Is $\mathbb{P}$ countably additive?

Solution (I) Yes. $\mathbb{P}$ being finitely additive is equivalent being additive for disjoint $A, B \subset \Omega$. There are three cases for these choices

(i) A, B are both finite. In this case $\mathbb{P}(A \dot\cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ since $0 = 0 + 0$.

(ii) A, B are both infinite. In this case $\mathbb{P}(A \dot\cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ since $\infty + \infty = \infty$.

(iii) One of the sets A, B is infinite. Then $\mathbb{P}(A \dot\cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ since $0 + \infty = \infty$.

(II) No. We will show this by counterexample. We can write $\Omega = \dot\bigcup_{i \in \mathbb{N}} \{i\}$. Then the countable additivity implies

$$\mathbb{P}(\Omega) = \mathbb{P}(\dot\bigcup_{i \in \mathbb{N}} \{i\}) \implies 1 = 0.$$

which is not true.

■ **Problem 3.15 — From Rosenthal.** Let $\mathcal{I}$ be the set of all intervals in $[0, 1]$ (open, closed, half-open, singleton, empty set). Show that $\mathcal{I}$ is a semi-algebra.

Solution By definition of $\mathcal{I}$ we have $\varnothing \in \mathcal{I}$. Let A_1, A_2 be two intervals in $[0, 1]$. If A_1, A_2 are disjoint, then $A_1 \cap A_2 \in \mathcal{I}$. If they are not disjoint, then without loss of generality we can assume that

$$A_1 = \{x \in [0, 1] \mid a < x < b\}, A_2 = \{x \in [0, 1] \mid c < x < d\},$$

where $a < c < b < d$. Thus $A_1 \cap A_2 = (c, b)$. So $\mathcal{I}$ is closed under finite intersection. The proof is the same for any other choices of A_1, A_2 (i.e. being closed set, etc). To show the third property, again, without the loss of generality, let $A = (a, b)$. Then $A^c = (-\infty, a] \dot\cup [b, \infty)$. For other choices of A (i.e. being closed, etc) we will have a similar argument. Thus we conclude that the collection $\mathcal{I}$ is a semi-algebra of the subsets of $[0, 1]$.

■ **Problem 3.16 — From Rosenthal.** Let $\mathcal{I}$ be the semi-algebra consisting of all intervals in $[0, 1]$. Define

$$\mathcal{B}_0 = \{\text{all finite unions of elements of } \mathcal{I}.\}$$

Show that $\mathcal{B}_0$ is not a σ-algebra.

Solution Along with many other sets, $\mathcal{I}$ contains all of the singletons, so does $\mathcal{B}_0$. Consider the following collection

$$\mathcal{A} = \{\{x\} \ : \ x \in \mathbb{Q} \cap [0, 1]\}.$$

By definition, all of the sets in the collection $\mathcal{A}$ belongs to $\mathcal{B}_0$. However, the following countable union

$$\dot\bigcup_{A \in \mathcal{A}} A = [0, 1] \cap \mathbb{Q}$$

does not belong to $\mathcal{B}_0$ (as it is not possible to generate with only finite unions of the elements of singletons in $\mathcal{I}$).

■ **Problem 3.17 — From Rosenthal.** Prove that the outer measure $\mathbb{P}^*$ is countably sub-additive, i.e.

$$\mathbb{P}^*(\bigcup_{n=1}^{\infty} B_n) \leq \sum_{n=1}^{\infty} \mathbb{P}^*(B_n) \qquad \text{for any } B_1, B_2, \cdots \in \Omega.$$

Solution This problem is the proof of Lemma 2.3.6 in Rosenthal. See the text for more context. A very quick review on the context is that we have a semi-algebra $\mathcal{I}$ of the subsets of Ω, and we have the function $\mathbb{P} : \mathcal{I} \to [0, 1]$ that satisfies the properties required for the extension theorem, hence there exist a valid probability space $(\Omega, \mathcal{M}, \mathbb{P}^*)$, where $\mathcal{M}$ is a σ-algebra and $\mathbb{P}^*$ is a probability measure (which is also the outer measure). The proof of this question is as follows.

Fix $\epsilon > 0$. Since the outer measure is the infimum of the sum of the probabilities on all $\mathcal{I}$ covers, then for each B_n we can find a collection $\{C_{nk}\}$ where $C_{nk} \in \mathcal{I}$ such that

$$\sum_k \mathbb{P}(C_{nk}) \leq \mathbb{P}^*(B_n) + \epsilon 2^{-n}.$$

On the other hand, since $\{C_{nk}\}_{nk}$ covers $\bigcup_n B_n$, then again from the properties of inf we have

$$\mathbb{P}^*(\bigcup_n B_n) \leq \sum_{nk} \mathbb{P}(C_{nk}).$$

combining these two we will get

$$\mathbb{P}^*(\bigcup_n B_n) \leq \sum_n \mathbb{P}^*(B_n) + \epsilon.$$

Since this is true for all $\epsilon > 0$, then it implies that

$$\mathbb{P}^*(\bigcup_n B_n) = \sum_n \mathbb{P}^*(B_n).$$

■ **Problem 3.18 — From Rosenthal.** If $A_1, A_2, \cdots \in \mathcal{M}$ are disjoint, then prove that

$$\mathbb{P}^*(\dot{\bigcup}_n A_n) = \sum_n \mathbb{P}^*(A_n).$$

Solution First, we start to show the finite additivity, and then using the properties of monotnonicity and sub-additivity, we will prove the countable additivity as well. Let $A_1, A_2 \in \mathcal{M}$ disjoint. In particular, since $A_1 \in \mathcal{M}$, from the definition of $\mathcal{M}$ (see page 12 Rosenthal), then

$$\mathbb{P}^*(A_1 \cup A_2) = \mathbb{P}^*(A_1^c \cap (A_1 \cup A_2)) + \mathbb{P}^*(A_1 \cap (A_1 \cup A_2)) = \mathbb{P}^*(A_2) + \mathbb{P}^*(A_1).$$

This implies that for any finite disjoint collection of A_i we have the additivity property (by induction). Now for any $m \in \mathbb{N}$ we have

$$\sum_{n<m} \mathbb{P}^*(A_n) = \mathbb{P}^*(\bigcup_{n \leq m} A_n) \leq \mathbb{P}^*(\bigcup_n A_n)$$

where the last inequality follows form the monotnonicity property of $\mathbb{P}^*$. Since this is true for all $m \in \mathbb{N}$, then we conclude that

$$\mathbb{P}^*(\bigcup_n A_n) \geq \sum_n \mathbb{P}^*(A_n).$$

On the other hand, from the sub-additivity property we have

$$\mathbb{P}^*(\bigcup_n A_n) \leq \sum_n \mathbb{P}^*(A_n).$$

These two implies that

$$\mathbb{P}^*(\bigcup_n A_n) = \sum_n \mathbb{P}^*(A_n).$$

■ **Problem 3.19 — From Rosenthal.** Let $\mathcal{M}$ be the σ-algebra we get from the extension theorem, where by definition it contains all of the sets like $A \in \Omega$ for which the outer measure is additive on the union of $A \cap E$ and $A^c \cap E$ for $\forall E \subset \Omega$. In other words

$$\mathcal{M} = \{A \subseteq \Omega \ : \ \mathbb{P}^*(A \cap E) + \mathbb{P}^*(A^c \cap E) = \mathbb{P}^*(E) \quad \text{for all } E \subset \Omega\}.$$

Prove that $\mathcal{M}$ is an algebra.

Solution Let $A = \Omega$. Then

$$\mathbb{P}^*(A \cap E) + \mathbb{P}^*(A^c \cap E) = \mathbb{P}^*(E) + \mathbb{P}^*(\varnothing) = \mathbb{P}^*(E).$$

So we conclude that $\Omega \in \mathcal{M}$. Also, it follows immediately from the definition of $\mathcal{M}$ that if $A \in \mathcal{M}$ then $A^c \in \mathcal{M}$. Now it remains to show if $A_1, A_2 \in \mathcal{M}$ then $A_1 \cap A_2 \in \mathcal{M}$. Let $E \subset \Omega$. Then

$$\begin{aligned}
&\mathbb{P}^*(E \cap (A_1 \cap A_2)) + \mathbb{P}^*(E \cap (A_1 \cap A_2)^c) \\
&= \mathbb{P}^*(E \cap (A_1 \cap A_2)) + \mathbb{P}^*(E \cap A_1^c \cap A_2) + \mathbb{P}^*(E \cap A_1 \cap A_2^c) + \mathbb{P}^*(E \cap A_1^c \cap A_2^c) \\
&\leq \mathbb{P}^*(E \cap A_1 \cap A_2) + \mathbb{P}^*(E \cap A_1^c \cap A_2) + \mathbb{P}^*(E \cap A_1 \cap A_2^c) + \mathbb{P}^*(E \cap A_1^c \cap A_2^c) \\
&= \mathbb{P}^*(E \cap A_2) + \mathbb{P}^*(E \cap A_2^c) \qquad (\text{because } A_1 \in \mathcal{M}) \\
&= \mathbb{P}^*(E) \qquad (\text{because } A_2 \in \mathcal{M}).
\end{aligned}$$

On the other hand, from the sub-additivity property we know that

$$\mathbb{P}^*(E \cap (A_1 \cap A_2)) + \mathbb{P}^*(E \cap (A_1 \cap A_2)^c) \geq \mathbb{P}^*(E).$$

Thus we conclude that

$$\mathbb{P}^*(E \cap (A_1 \cap A_2)) + \mathbb{P}^*(E \cap (A_1 \cap A_2)^c) = \mathbb{P}^*(E).$$

This implies that $A_1 \cap A_2 \in \mathcal{M}$ and this finishes the proof.

■ **Problem 3.20 — From Rosenthal.** Let $A_1, A_2, \dots \in \mathcal{M}$ be disjoint. For each $m \in \mathbb{N}$, let $B_m = \bigcup_{n \leq m} A_n$. Prove that for all $m \in \mathbb{N}$, and for all $E \subseteq \Omega$ we have

$$\mathbb{P}^*(E \cap B_m) = \sum_{n \leq m} \mathbb{P}^*(E \cap A_n).$$

Solution First, observe that this statement is true for $m = 1$ in a trivial way. For $m = 2$, since $A_2 \in \mathcal{M}$, then we can expand $E \cap B_2$ according to A_2, i.e.

$$\mathbb{P}^*(E \cap B_2) = \mathbb{P}^*((E \cap B_2) \cap A_2) + \mathbb{P}^*((E \cap B_2) \cap A_2^c)$$

On the other hand $(E \cap B_2) \cap A_2 = E \cap A_2$ and $(A \cap B_2) \cap A_2^c = E \cap B_1 = E \cap A_1$. Thus we can write

$$\mathbb{P}^*(E \cap B_2) = \mathbb{P}^*(E \cap A_1) + \mathbb{P}^*(E \cap A_2).$$

In general, for $m \in \mathbb{N}$ we can write

$$\mathbb{P}^*(E \cap B_m) = \mathbb{P}^*(E \cap A_m) + \mathbb{P}^*(E \cap B_{m-1}).$$

Thus using induction we can write

$$\mathbb{P}^*(E \cap B_m) = \sum_{n \leq m} \mathbb{P}^*(E \cap A_n).$$

■ **Problem 3.21 — From Rosenthal.** In this question covers some of the proves for constructing a uniform probability measure on $\Omega = [0,1]$. Let $\mathcal{I}$ be the set of all intervals in Ω, and let $\mathcal{P} : \mathcal{I} \to [0,1]$ be a function that assigns the length of an interval to that interval. We want to prove that $\mathcal{P}$ satisfies the property (2.3.3) of extension theorem (2.3.1) in Rosenthal. I.e. we want to prove that for $A, A_1, \cdots \in \mathcal{I}$ such that $A \subset \cup_i A_i$ satisfies

$$\mathbb{P}(A) \leq \sum_i \mathbb{P}(A_i).$$

Solution We will do this in three parts.

(i) Step 1. First, we prove that for any finite $A, A_1, \cdots, A_n \in \mathcal{I}$ collection where $A \subset \cup_{i=1}^n A_i$ we have

$$\mathbb{P}(A) \leq \sum_{i=1}^n \mathbb{P}(A_i).$$

To see this let $A_1, A_2, A \in \mathcal{I}$ such that $A \subset A_1 \cup A_2$. We also assume that $A_1 \cap A \neq \varnothing$ as well as $A_2 \cap A \neq \varnothing$. This is to ensure that we do not have redundant interval in our collection that does not cover A. Note that from any collection of intervals we can put aside the redundant intervals and do our reasoning here and then finally at the last step consider the redundant intervals as well. So our assumption above does not lose the generality of the proof. Let a_i, b_i represent the left (right) endpoints of the intervals A_i and a_0, b_0 represent the left (right) endpoints of the interval A. In order to the intervals A_1, A_2 to cover A white neither of them are redundant we need to have

$$\min\{a_1, a_2\} \leq a_0 \leq a_2 \leq b_1 \leq b_0 \leq \max\{b_1, b_2\}.$$

Then it follows that

$$b_0 - a_0 \leq (b_1 - a_1) + (b_2 - a_2).$$

Thus this implies that

$$\mathbb{P}(A) \leq \mathbb{P}(A_1) + \mathbb{P}(A_2).$$

We can generalize this by induction to any finite number of intervals.

(ii) Step 2. We now want to prove that for any countable *open* intervals $A_1, A_2, \cdots \in \mathcal{I}$ such that $A \subset \cup_n A_n$ for $A \in \mathcal{I}$ closed, we have

$$\mathbb{P}(A) \leq \sum_n \mathbb{P}(A_n).$$

To see this, we will use the Heine-Borel theorem. Since the collection $\{A_1, A_2, \cdots\}$ is an open cover for the closed set A. if A is the whole space Ω, then the inequality that we want to show follows immediately (LHS is 1 while RHS is infinite). However, if A is not the whole space, then it is bounded. Thus A is closed and bounded, hence compact (by Heine-Borel). So the open cover has a finite sub-cover and this completes the proof by reducing this case to case (i) above.

(iii) Step 3. We now want to show that if $A_1, A_2, \cdots \in \mathcal{I}$ is any countable collection of intervals, and if $A \subset \cup_n A_n$ for any $A \in \mathcal{I}$ then

$$\mathbb{P}(A) \leq \sum_{i=1}^\infty \mathbb{P}(A_i).$$

TODO: TO BE COMPLETED

■ **Problem 3.22 — From Rosenthal.** Let $\mathcal{A} = \{(-\infty, x] \ : \ x \in \mathbb{R}\}$. Prove that $\sigma(\mathcal{A}) = \mathcal{B}$, i.e. that the smallest σ-algebra of subsets of $\mathbb{R}$ which contains $\mathcal{A}$ is equal to the Borel σ-algebra of subsets of $\mathbb{R}$.

Solution By definition, we know that $\mathcal{B}$ is the smallest σ-algebra that contains all of the intervals. However, we claim that the σ-algebra $\sigma(\mathcal{A})$ also contains all of intervals. To see this, let I be an interval which will have different cases $(-\infty, a), (-\infty, a], (a, b), [a, b], (a, b], [a, b), (a, \infty), [a, \infty)$. Each of these sets can be constructing by using the sets in $s\sigma(\mathcal{A})$ and using its sigma-algebra properties. Thus we showed that $\sigma(\mathcal{A})$ contains $\mathcal{I}$ the set of all intervals. However, by definition $\mathcal{B}$ was the smallest σ-algebra containing all of intervals. Thus the σ-algebra generated by $\mathcal{A}$ (i.e. the smallest σ-algebra by definition) is equal to $\mathcal{B}$.

■ **Problem 3.23 — From Rosenthal.** Prove the following statements.

(a) Prove that the Cantor set K and its complement K^c is in $\mathcal{B}$, the Borel set of the subset of $[0, 1]$.

(b) Prove that $K, K^c \in \mathcal{M}$, where $\mathcal{M}$ is the σ-algebra by the extension applied for the uniform distribution on $[0, 1]$ (see theorem 2.4.4 Rosenthal).

(c) Prove that $K^c \in \mathcal{B}_1$ where $\mathcal{B}_1$ is defined by (2.2.6) Rosenthal (i.e. the set of all finite or countable unions of intervals $\mathcal{I}$).

(d) Prove that $\mathcal{B}_1$ is not a σ-algebra.

Solution (a) In the construction of the cantor set, at each step we remove the middle 1/3 of the intervals. So starting with $I_0 = [0, 1]$ at the first step we will get $I_1 = [0, 1/3] \cup [2/3, 1]$, etc. The Cantor set is $I_0 \cap I_1 \cap I_2 \cap \cdots$. Since $I_i \in \mathcal{B}$ for all $i \in \mathbb{N}$, and $\mathcal{B}$ is a σ-algebra and is closed under countable intersection, then $K \in \mathcal{B}$ as well. It follows immediately that $K^c \in \mathcal{B}$ as well, as $\mathcal{B}$ is closed under complement.

(b) Since $\mathcal{M} \supset \mathcal{B}$, and as we showed above that $K, K^c \in \mathcal{B}$, then $K, K^c \in \mathcal{M}$ as well.

(c) From the construction given in part (a), we have

$$K = I_0 \cap I_1 \cap I_2 \cap \cdots .$$

From the De Morgan's law we will have

$$K^c = I_0^c \cup I_1^c \cup I_2^c \cup \cdots .$$

Thus by definition of $\mathcal{B}_1$ we have $K^c \in \mathcal{B}_1$.

(d) First, observe that (see Rosenthal page 17) that the Cantor set is uncountable. So there is no way to construct it by finite or countable union of singletons (which are in $\mathcal{I}$). We can not construct it with the finite or countable union of any intervals as K is a nowhere dense set. I.e. for every interval $(a, b) \in \mathcal{I}$ containing $x \in K$ there exists, $y \in \mathbb{R}$ such that $y \notin K$. To put this precisely, let $I_1, I_2, I_3, \cdots$ is collection of intervals such that $I = \cup_n I_n$. Since K is nowhere dense, for any I_i in the collection that contains $x \in K$ we can find some $y \in \mathbb{R}$ that $y \notin K$. This is a contradiction and we have $I \subset \cup_n I_n$.

(e) As we saw above, $K^c \in \mathcal{B}_1$ but $K \notin \mathcal{B}_1$. Thus $\mathcal{B}_1$ is not closed under complement, thus it is not a σ-algebra.

■ **Problem 3.24 — An extension of the extension theorem (from Rosenthal).** Let $\mathcal{I}$ be a semialgebra of subsets of Ω. Let $\mathbb{P} : \mathcal{I} \to [0,1]$ with $\mathbb{P}(\varnothing) = 0$ and $\mathbb{P}(\Omega) = 1$, satisfying

$$\mathbb{P}(\bigcup_n A_n) \geq \sum_n \mathbb{P}(A_n) \qquad \text{for } A_1, A_2, \dots \in \mathcal{I} \textit{ disjoint}, \text{ and } \bigcup_n A_n \in \mathcal{I},$$

as well as

$$\mathbb{P}(A) \leq \mathbb{P}(B), \qquad A \subseteq B,$$

and

$$\mathbb{P}(\bigcup_n B_n) \leq \sum_n \mathbb{P}(B_n) \qquad \text{for } B_1, B_2, \dots \in \mathcal{I}, \text{ and } \bigcup_n B_n \in \mathcal{I}.$$

Then there exist a valid probability space $(\Omega, \mathcal{M}, \mathbb{P}^*)$ such that $\mathbb{P}$ and $\mathbb{P}^*$ agree on the elements of $\mathcal{I}$

Solution According to the extension theorem 2.3.1 we need to prove that these alternative statements implies 2.3.3. Let $A, A_1, A_2, \dots \in \mathcal{I}$ (note that the union of A_is do not belong to $\mathcal{I}$ necessarily). Define

$$B_n = A_n \cap A.$$

Then we will have $A = \cup_n B_n$, thus $\cup_n B_n \in \mathcal{I}$. So

$$\mathbb{P}(A) = \mathbb{P}(\bigcup_n B_n) \leq \sum_n \mathbb{P}(B_n) \leq \sum_n \mathbb{P}(A_n).$$

Observation 3.6.1 In the prove above, one might attempt

$$\mathbb{P}(A) \leq \mathbb{P}(\bigcup_i A_i) \leq \sum_i \mathbb{P}(A_i)$$

where for the first inequality we use the monotnonicity (as $A \subset \bigcup_i A_i$) and for the second inequality we use the countable sub additivity given in the statement of the theorem. However, this prove is *wrong!*. Because we are not allowed to use the second inequality, as it does not satisfies the requirements for the sub-additivity statement to work. That is because $\bigcup_n A_n$ may not be in $\mathcal{I}$ necessarily.

■ **Problem 3.25 — Uniqueness property of the extension theorem (from Rosenthal).** Let $\mathcal{I}, \mathbb{P}, \mathbb{P}^*$ be the same as in Theorem 2.3.1 (Rosenthal). Let $\mathcal{F}$ be any σ-algebra with $\mathcal{I} \subseteq \mathcal{F} \subseteq \Omega$. Let $\mathbb{Q}$ be any probability measure on $\mathcal{F}$, such that $\mathbb{Q}(A) = \mathbb{P}(A)$ for all $A \in \mathcal{I}$. Then prove that $\mathbb{Q}(A) = \mathbb{P}^*(A)$ for all $A \in \mathcal{F}$.

Solution Let $A \in \mathcal{F}$, and $A_i \in \mathcal{I}$ for $i = 1, 2, \dots$ such that $A \subseteq \bigcup_i A_i$. Since $\mathbb{Q}$ is a probability measure, then

$$\mathbb{Q}(A) \leq \mathbb{Q}(\bigcup_i A_i) \leq \sum_i \mathbb{Q}(A_i).$$

Then we can write

$$\begin{aligned}
\mathbb{P}^*(A) &= \inf_{\substack{A_1,A_2,\cdots\in\mathcal{I}\\ A\subseteq\bigcup_i A_i}} \sum_i \mathbb{P}(A_i)\\
&= \inf_{\substack{A_1,A_2,\cdots\in\mathcal{I}\\ A\subseteq\bigcup_i A_i}} \sum_i \mathbb{Q}(A_i)\\
&\geq \inf_{\substack{A_1,A_2,\cdots\in\mathcal{I}\\ A\subseteq\bigcup_i A_i}} \mathbb{Q}(\bigcup_i A_i)\\
&\geq \inf_{\substack{A_1,A_2,\cdots\in\mathcal{I}\\ A\subseteq\bigcup_i A_i}} \mathbb{Q}(A)\\
&= \mathbb{Q}(A).
\end{aligned}$$

However, we could do the same with $A^c \in \mathcal{F}$, which would lead to

$$\mathbb{P}^*(A^c) \geq \mathbb{Q}(A^c).$$

Since $\mathbb{P}^*(A^c) = 1 - \mathbb{P}^*(A)$ and $\mathbb{Q}(A^c) = 1 - \mathbb{Q}(A)$, then this implies $\mathbb{P}^*(A) \geq \mathbb{Q}(A)$. Thus $\mathbb{P}^*(A) = \mathbb{Q}(A)$.

■ **Problem 3.26 — Properties of Random Variables (from Rosenthal).** Prove the followings.

(i) If X, Y are random variables and $c \in \mathbb{R}$, then $X + c, cX, X^2, X + Y$, and XY are random variables.

(ii) If $Z_1, Z_2, \cdots$ are random variables such that $\lim_{n\to\infty} Z_n(\omega)$ exists for all $\omega \in \Omega$,, and $Z(w) = \lim_{n\to\infty} Z_n(\Omega)$, then Z is also a random variable.

Solution the proves are as follows.

(i) (a) $(X + c)^{-1}((-\infty, x]) = \{w : X(w) + c \leq x\} = \{w : X(w) \leq x - c\} = X^{-1}((0\text{-}\infty, x - c]) \in \mathcal{F}$.

(b) Assume $c \neq 0$. Then $(cX)^{-1}((-\infty, x]) = \{w : cX(w) \leq x\} = \{w : X(w) \leq x/c\} = X^{-1}((-\infty, x/c]) \in \mathcal{F}$. For the case were $c = 0$, then $cX \equiv 0$ on all Ω, and this is a random variable as $(cX^{-1})((0, x]) = \varnothing$ if $x < 0$ and $(cX^{-1})((0, x]) = \Omega$ if $x \geq 0$.

(c) $X^2((-\infty, a]) = \{w : X^2 \leq a\} = \{w : X \in [-\sqrt{a}, \sqrt{a}]\} \in \mathcal{F}$.

(d) $(X + Y)^{-1}((\infty, x]) = \{w : X(w) + Y(w) < x\} = \bigcup_{r\in\mathbb{Q}}(\{X < r\} \cap \{Y < x - r\})$. Since this is a countable union, thus in $\mathcal{F}$.

(e) I have the following prove but I am not sure if it is a correct one or not. I feel that this is a correct proof as there seems to be nothing that can make it not to work.

$$(XY)^{-1}((-\infty, x]) = \{XY < x\} = \bigcup_{n\in\mathbb{N}}(\{X < n\} \cap \{Y < x/n\}).$$

The following prove is the idea by Rosenthal. Once we know that X^2, $X + Y$, and cX are random variables, then we can deduce XY is also a random variable as

$$XY = \frac{1}{2}((X + Y)^2 - X^2 - Y^2))$$

(ii) In a nutshell this statements claims that the point-wise convergence of a sequence of random variables is a random variable. We need to show that the event $\{Z \leq r\} \in \mathcal{F}$. To see this, let $w \in \{Z \leq r\}$. This means that $Z(w) \leq r$. Since $Z_n(\omega) \to Z(\omega)$ as $n \to \infty$, then this implies that $\forall m \in \mathbb{N}$ there exists $N \in \mathbb{N}$ such that $\forall n > N$ we have $Z_n(\omega) \leq r + \frac{1}{m}$. Thus we can write

$$\{Z(w) \leq r\} = \bigcap_{m=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \{Z_n(\omega) \leq r + \frac{1}{m}\}$$

■ **Problem 3.27** Let $f : \mathbb{R} \to \mathbb{R}$ be continouse or piece-wise continuous function. Then f is a Borel function.

Solution We first prove the statement for a continuous function. To show that f is a Borel function we need to show $f^{-1}(B) \in \mathcal{B}$ for $B \in \mathcal{B}$. Since the pre-image preserves the union, intersection, and complements, and by definition for a continuous function the pre-image of an open set is an open set, and using the fact that we can write any Borel set as a countable intersection, union, or complements of open sets then we conclude that $f^{-1}(B) \in \mathcal{B}$. A second way to show this is observe that $f^{-1}((x,\infty)) \in \mathcal{B}$ as (x,∞) is open and f is continuous, thus its pre-image is also an open set thus a Borel set. Then its complement is also a Borel set, i.e.

$$(f^{-1}((x,\infty)))^c = f^{-1}((-\infty, x]) \in \mathcal{B}.$$

Thus shows that f is a Borel function (since the pre-image of $(0,x]$ is a Borel set.

For the case where f is piece-wise continuous, by the definition of the piece-wise continuoity, f has at most countably many discontinuities. The we can write f as

$$f(x) = f_1(x)\mathbb{1}_{I_1}(x) + f_2(x)\mathbb{1}_{I_2}(x) + f_3(x)\mathbb{1}_{I_3}(x) + \cdots + f_n(x)\mathbb{1}_{I_n}(x),$$

where $I_1, I_2, I_3 \cdots, I_n$ are disjoint intervals on which $f_1, f_2, f_3, \cdots, f_n$ are continuous respectively. By the statement for the first part of the proof, we know that f_i is a Borel function (since it is continuous) as well as the indicator function $\mathbb{1}_{I_i}$. Their multiplication is also a Borel function and the sum of these Borel functions is a also a Borel function, thus $f(x)$ is a Borel function.

■ **Problem 3.28** Prove that if A, B are two independent events, then (A^c, B), (A, B^c), and (A^c, B^c) are pairwise independent.

Solution Since A, B are independent, then $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. From the identity $B = (B \cap A)\dot{\cup}(B \cap A^c)$. From the properties of the probability measure we have

$$\mathbb{P}(B) = \underbrace{\mathbb{P}(A \cap B)}_{\mathbb{P}(A)\mathbb{P}(B)} + \mathbb{P}(A^c \cap B).$$

Then we can write

$$\mathbb{P}(A^c \cap B) = \mathbb{P}(B) - \mathbb{P}(A \cap B) = \mathbb{P}(B)(1 - \mathbb{P}(A)) = \mathbb{P}(B)\mathbb{P}(A^c).$$

Thus A^c, B are also independent events. We use a similar argument for A, B^c. To show that the events A^c, B^c are also independent, we use the inclusion-exclusion principle.

$$\mathbb{P}(A^c \cap B^c) = 1 - \mathbb{P}(A \cup B) = 1 - (\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B)) = (1 - \mathbb{P}(A))(1 - \mathbb{P}(B)) = \mathbb{P}(A^c)\mathbb{P}(B^c).$$

Another way of showing this without the inclusion-exclusion principle is to use the identity

$$A^c = (A^c \cap B)\dot{\cup}(A^c \cap B^c).$$

Then

$$\mathbb{P}(A^c) = \underbrace{\mathbb{P}(A^c \cap B)}_{\mathbb{P}(A^c)\mathbb{P}(B)} + \mathbb{P}(A^c \cap B^c).$$

We can write

$$\mathbb{P}(A^c \cap B^c) = \mathbb{P}(A^c)(1 - \mathbb{P}(B)) = \mathbb{P}(A^c)\mathbb{P}(B^c).$$

■ **Problem 3.29** Let X, Y be random variables, and $f, g : \mathbb{R} \to \mathbb{R}$ be Borel functions. Then if X, Y are independent, $f(X)$ and $g(X)$ are also independent.

Solution Since X, Y are independent, thus for any $S_1, S_2 \in \mathcal{B}$ we have

$$\mathbb{P}(\{X \in S_1\} \cap \{Y \in S_2\}) = \mathbb{P}(\{X \in S_1\})\mathbb{P}(\{Y \in S_2\}).$$

Consider

$$\mathbb{P}(\{f(X) \in S_1\} \cap \{g(Y) \in S_2\}) = \mathbb{P}(\{X \in f^{-1}(S_1)\} \cap \{Y \in g^{-1}(S_2)\}) = \mathbb{P}(\{f(X) \in S_1\})\mathbb{P}(\{g(Y) \in S_2\}).$$

The equality above hods because for any $S_1, S_2 \in \mathcal{B}$ we have $f^{-1}(S_1), g^{-1}(S_2) \in \mathcal{B}$, that is because f, g are Borel functions.

■ **Problem 3.30 — Continuity of probabilities.** Prove that the probability measure function is continuous from bellow and above. I.e. for the continuity from below, let $A, A_1, A_2, \cdots \in \mathcal{F}$ such that $\{A_n\} \nearrow A$, i.e. $A_1 \subseteq A_2 \subseteq \cdots$ and $A = \bigcup_n A_n$. Then $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$. For the continuity from above, let $A, A_1, A_2, \cdots \in \mathcal{F}$ such that $\{A_n\} \searrow A$, i.e. $A_1 \subseteq A_2 \subseteq \cdots$ and $A = \bigcap_n A_n$, then $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$.

Solution First, we prove the probability from below. In general, one can observe that if $\{A_n\} \nearrow A$, then $\{\mathbb{P}(A_n)\}$ indeed converges as this is a bounded monotone sequence in $\mathbb{R}$. However, to show that this sequence converges to $\mathbb{P}(A)$, we do as following. Consider the following sets

$$B_1 = A_1, \qquad B_2 = A_2 \backslash A_1, \qquad B_3 = A_3 \backslash A_2, \cdots .$$

Then $A = \dot{\cup} B_n$. Thus $\mathbb{P}(A) = \sum_{n=1}^{\infty} \mathbb{P}(B_n)$. Thus the series on the rand hand side converges. This implies that the corresponding partial sums also converges. However, by the construction we have

$$\mathbb{P}(A_1) = \mathbb{P}(B_1), \quad \mathbb{P}(A_2) = \mathbb{P}(B_1) + \mathbb{P}(B_2) + \cdots .$$

Thus the converges of the partial sums implies the convergence of $\{\mathbb{P}(A_n)\}$.

For the proof for the continuity from above, first observe that if for a collection $A, A_1, A_2, \cdots \in \mathcal{F}$ we have $\{A_n\} \searrow A$, then this is equivalent to $\{A_n^c\} \nearrow A^c$. This follows from the De Morgan's law as well as the change of the direction of the inclusion $\subseteq$ under taking complements. By hypothesis we have $\{A_n\} \searrow A$, which is equivalent to$\{A_n^c\} \nearrow A^c$. From the first part of the proof, it follows that $\lim_{n\to\infty} \mathbb{P}(A_n^c) = \mathbb{P}(A^c)$. Thus

$$\lim_{n\to\infty} (1 - \mathbb{P}(A_n)) = 1 - \mathbb{P}(A).$$

Note that $\mathbb{P}(A_n)$ converges to some real number as $n \to \infty$, because it is a bounded decreasing sequence. Thus by the laws of the limit we can write

$$1 - \lim_{n\to\infty} \mathbb{P}(A_n) = 1 - \mathbb{P}(A).$$

Then it follows that

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A).$$

■ **Problem 3.31 — From Rosenthal.** Let $A_1, A_2, \cdots, A_n \in \mathcal{F}$. Generalize the principle of inclusion-exclusion to:

$$\mathbb{P}(A_1 \cup \cdots \cup A_n) = \sum_{i=1}^{n} A_i - \sum_{1\leq i<j\leq n} \mathbb{P}(A_i \cap A_j) + \sum_{1\leq i<j<k\leq n} \mathbb{P}(A_i \cap A_j \cap A_k) - \cdots \pm \mathbb{P}(A_1 \cap \cdots \cap A_n).$$

Hint: Expand $1 - \prod_{i=1}^{n}(1 - \mathbb{1}_{A_i})$, *and take expectation of both sides.*

Solution First, observe that $\mathbb{1}_{A^c} = 1 - \mathbb{1}_A$. So

$$1 - \prod_{i=1}^{n}(1 - \mathbb{1}_{A_i}) = 1 - \prod_{i=1}^{n} \mathbb{1}_{A_i^c} = 1 - \mathbb{1}_{A_1^c \cap \cdots \cap A_n^c} = \mathbb{1}_{A_1 \cup \cdots \cup A_n}.$$

On the other hand,

$$1 - \prod_{i=1}^{n}(1 - \mathbb{1}_{A_i}) = \sum_{i=1}^{n} \mathbb{1}_{A_i} - \sum_{1\leq i<j\leq n} \mathbb{1}_{A_i \cap A_j} + \cdots \pm \mathbb{1}_{A_1 \cap \cdots \cap A_n}.$$

So we have

$$\mathbb{1}_{A_1 \cup \cdots \cup A_n} = \sum_{i=1}^{n} \mathbb{1}_{A_i} - \sum_{1\leq i<j\leq n} \mathbb{1}_{A_i \cap A_j} + \cdots \pm \mathbb{1}_{A_1 \cap \cdots \cap A_n}.$$

By applying the expectation to both sides we will get

$$\mathbb{P}(A_1 \cup \cdots \cup A_n) = \sum_{i=1}^{n} A_i - \sum_{1\leq i<j\leq n} \mathbb{P}(A_i \cap A_j) + \sum_{1\leq i<j<k\leq n} \mathbb{P}(A_i \cap A_j \cap A_k) - \cdots \pm \mathbb{P}(A_1 \cap \cdots \cap A_n).$$

■ **Problem 3.32 — From Rosenthal.** Let $f(x) = ax^2 + bx + c$ be a second-degree polynomial function where $a, b, c \in \mathbb{R}$.

(a) Find necessary and sufficient condition on a, b and c such that the equation $\mathbb{E}\Big[f(\alpha x)\Big] = \alpha^2 \mathbb{E}\Big[f(X)\Big]$ holds for all $\alpha \in \mathbb{R}$ and all random variable X.

(b) Find necessary and sufficient condition on a, b and c such that the equation $\mathbb{E}\Big[f(x - \beta)\Big] = \mathbb{E}\Big[f(x)\Big]$ holds of all $\beta \in \mathbb{R}$ and all random variable X.

(c) Do parts (a) and (b) account for the properties of the variance function? Why or why not?

Solution (a) For the LHS we have

$$\mathbb{E}\Big[\alpha^2 a x^2 + \alpha b x + c\Big] = a\alpha^2 \mathbb{E}\Big[x^2\Big] + b\alpha \mathbb{E}\Big[x\Big] + c.$$

And for the RHS we have

$$\alpha^2 \mathbb{E}\Big[f(x)\Big] = a\alpha^2 \mathbb{E}\Big[x^2\Big] + b\alpha^2 \mathbb{E}\Big[x\Big] + c\alpha^2.$$

Thus the necessary and sufficient condition for the equality to hold for every $\alpha \in \mathbb{R}$ is to have

$$b = 0, \qquad c = 0.$$

(b) For the LHS we have

$$\mathbb{E}\Big[f(x-\beta)\Big] = a\mathbb{E}\Big[x^2\Big] + (b-2a\beta)\mathbb{E}\Big[x\Big] + a\beta^2 - b\beta + c.$$

For the RHS we have

$$\mathbb{E}\Big[f(x)\Big] = a\mathbb{E}\Big[x^2\Big] + b\mathbb{E}\Big[x\Big] + c.$$

This implies that we need to have

$$a = 0, \qquad b = 0.$$

I.e. the polynomial should be a constant polynomial.

(c) For the property $\operatorname{Var}(\alpha C) = \alpha^2 \operatorname{Var}(C)$, it follows from the fact that in part (b) we found that for the polynomial we need to have $b = 0$ and $c = 0$. However, part (b) does not account for the property of variance that $\operatorname{Var}(X + b) = \operatorname{Var}(X)$. Because in the case of Var the constants of the polynomial depends on the random variable under consideration. I.e. we have

$$b = -2\mathbb{E}\Big[X\Big], \qquad c = \mathbb{E}\Big[X\Big]^2.$$

■ **Problem 3.33 — From Rosenthal.** Let $X_1, X_2, \cdots$ be independent, each with mean μ and variance σ^2, and let N be an integer-valued random variable with mean m and variance v, with N independent of all the X_i. Let $S = X_1 + X_2 + \cdots + X_n = \sum_{i=1}^{\infty} X_i \mathbb{1}_{N\geq i}$. Compute $\operatorname{Var}(S)$ and $\mathbb{E}\Big[S\Big]$.

Solution — Using the conditional expectation. For this problem we can either use the conditional expectation or use the first principles.

$$\mathbb{E}\Big[S\Big] = \mathbb{E}\Big[X_1 + \cdots + X_N\Big] = \sum_n \mathbb{E}\Big[X_1 + \cdots + X_n\Big]\mathbb{P}(N=n) = \mu \sum_n n\mathbb{P}(N=n) = \mu m.$$

Similarly we have

$$\begin{aligned}
\mathbb{E}\Big[S^2\Big] &= \sum_n \mathbb{E}\Big[S^2 \mid N=n\Big]\mathbb{P}(N=n) \\
&= \sum_n \mathbb{E}\Big[(X_1+\cdots+X_n)^2\Big]\mathbb{P}(N=n) \\
&= \sum_n \mathbb{E}\Big[\sum_i X_i^2 + \sum_{i<j} X_iX_j\Big]\mathbb{P}(N=n) \\
&= \sum_n (\sum_i \mathbb{E}\Big[X_i^2\Big] + 2\sum_{i<j}\mathbb{E}\Big[X_iX_j\Big])\mathbb{P}(N=n) \\
&= \sum_n ((\sigma^2+\mu^2)n + n(n-1)\mu^2)\mathbb{P}(N=n) \\
&= (\sigma^2+\mu^2)m + \mu^2(v + m^2 - m).
\end{aligned}$$

Where we have used $\mathbb{E}\Big[X_i^2\Big] = \operatorname{Var}(X_i) + \mathbb{E}\Big[X_i\Big]^2 = \mu^2 + \sigma^2$, and also used the fact that the sum $\sum_{i<j}$ has $\binom{n}{2}$ terms. Now we can compute the variance

$$\operatorname{Var}(S) = \mathbb{E}\Big[S^2\Big] - \mathbb{E}\Big[S\Big]^2 = (\sigma^2+\mu^2)m + \mu^2(v+m^2-m) - \mu^2m^2 = \sigma^2 m + \mu^2 v.$$

■ **Problem 3.34 — from Rosenthal.** Let X, Z be independent random variables each with standard normal distribution. Let $a, b \in \mathbb{R}$ (not both 0), and let $Y = aX + bZ$.

(a) Compute $\mathrm{Corr}(X, Y)$.

(b) Show that $|\mathrm{Corr}(X, Y)| \leq 1$.

(c) Given necessary and sufficient conditions on the values of a and b such that $\mathrm{Corr}(X, Y) = 1$.

(d) Given necessary and sufficient conditions on the values of a and b such that $\mathrm{Corr}(X, Y) = -1$.

Solution (a) First, observe that

$$\mathbb{E}\Big[X\Big] = \mathbb{E}\Big[Z\Big] = 0, \quad \mathrm{Var}(X) = \mathrm{Var}(Y) = 1, \quad \mathbb{E}\Big[Y\Big] = a + b, \quad \mathrm{Var}(Y) = a^2 + b^2$$

where we have used the properties of variance to compute $\mathrm{Var}(Y)$. By the definition of correlation we have

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} = \frac{\mathbb{E}\Big[XY\Big] - \mathbb{E}\Big[X\Big]\mathbb{E}\Big[Y\Big]}{\sqrt{a^2 + b^2}} = \frac{a}{\sqrt{a^2 + b^2}}.$$

(b) It follows immediately by

$$|\mathrm{Corr}(X, Y)| = |\frac{1}{\sqrt{1 + b^2/a^2}}| \leq 1.$$

(c) From our solution in part (a), the necessary and sufficient condition for $\mathrm{Corr}(X, Y) = 1$ is that $b = 0$ and $a > 0$.

(d) From the solution in part (a). the necessary and sufficient condition for $\mathrm{Corr}(X, Y) = -1$ is that $b = 0$ and $a < 0$.

■ **Problem 3.35 — From Rosenthal.** Let X, Y be independent general non-negative random variables, and let $X_n = \Psi_n(X)$, where $\Psi_n(x) = \min(n, 2^{-n}\lfloor 2^n x \rfloor)$.

(a) Give an example of a sequence of functions $\Phi_n : [0, \infty) \to [0, \infty)$ other than $\Phi_n = \Psi_n$, such that for all x we have $0 \leq \Phi_n(x) \leq x$ and $\{\Phi_n(x)\} \nearrow x$ as $n \to \infty$.

(b) Suppose that $Y_n = \Phi_n(Y)$ with Φ_n as in part (a). Must X_n and Y_n be independent.

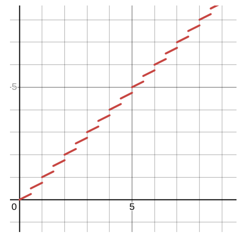
(c) Suppose $\{Y_n\}$ is an arbitrary collection of non-negative random variables such that $\{Y_n\} \nearrow Y$. Must X_n and Y_n be independent?

(d) Under the assumption of part (c), determine which quantities in 4.2.7 are necessarily equal?

Solution (a) Define

$$f_n(x) = \min\{n, \frac{1}{2^n}(\frac{1}{2}(2^n x + \lfloor 2^n x \rfloor))\}.$$

The graph of this function will be as follows.



(b) First observe that Φ_n is a Borel measurable function as the pre-image of any open interval is a union of half open intervals or the empty set. Since X, Y are independent, then X_n, Y_n are also independent.

(c) No. We demonstrate a counterexample. Let

$$Y_n = \max\{\Psi_n(Y) - \frac{1}{n^2}\Psi_n(X), 0\}.$$

(d) Since $\{X_n\} \nearrow X$ and $\{Y_n\} \nearrow Y$, then $\{X_nY_n\} \nearrow XY$. Thus by the monotone convergence theorem we have

$$\mathbb{E}\Big[XY\Big] = \lim_n \mathbb{E}\Big[X_nY_n\Big].$$

Also since (by monotone convergence theorem) $\mathbb{E}\Big[X\Big] = \lim_n \mathbb{E}\Big[X_n\Big]$ and $\mathbb{E}\Big[Y\Big] = \lim_n \mathbb{E}\Big[Y_n\Big]$ then by the limit laws we have

$$\lim_n \mathbb{E}\Big[X_n\Big]\mathbb{E}\Big[Y_n\Big] = \mathbb{E}\Big[X\Big]\mathbb{E}\Big[Y\Big].$$

■ **Problem 3.36 — From Rosenthal.** Give examples of a random variable X defined on Lebesgue measure on $[0, 1]$, such that

(a) $\mathbb{E}\Big[X^+\Big] = \infty$ and $0 < \mathbb{E}\Big[X^-\Big] < \infty$.

(b) $\mathbb{E}\Big[X^-\Big] = \infty$ and $0 < \mathbb{E}\Big[X^+\Big] < \infty$.

(c) $\mathbb{E}\Big[X^+\Big] = \mathbb{E}\Big[X^-\Big] = \infty$.

(d) $0 < \mathbb{E}\Big[X\Big] < \infty$ but $\mathbb{E}\Big[X^2\Big] = \infty$.

Solution (a) Define

$$X^+ = 2 \cdot \mathbb{1}_{(1/2,3/4)} + \sum_{n=2}^{\infty} 2^n \cdot \mathbb{1}_{(2^{-n},2^{-n+1})}, \qquad X^- = -2 \cdot \mathbb{1}_{(3/4,1)}.$$

Then $\mathbb{E}\Big[X^+\Big] = 1/2 + 1 + 1 + \cdots = \infty, \qquad \mathbb{E}\Big[X^-\Big] = 1/2.$

(b) Similar to part (a) but exchange X^- and X^+.

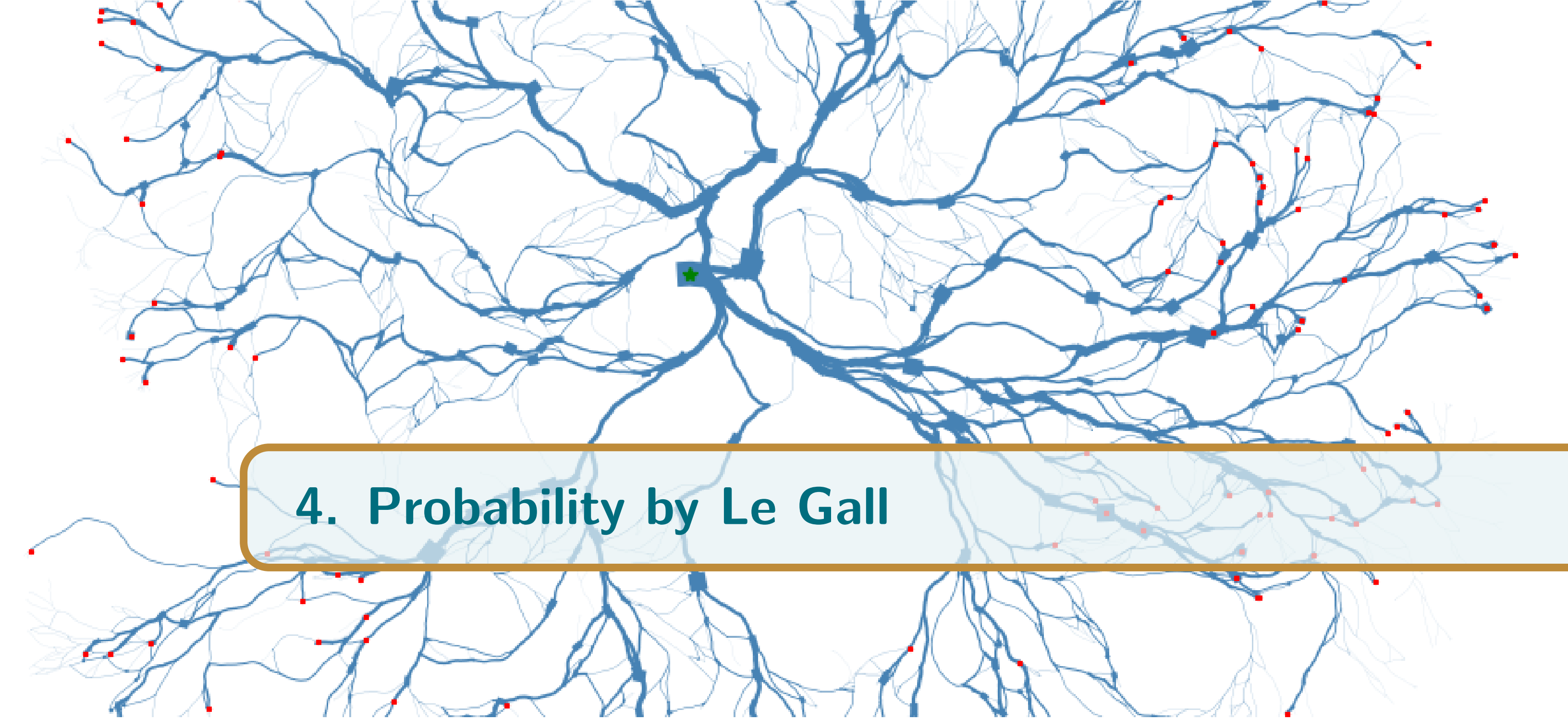
(c) Define

$$X^+ = \sum_{n \text{ even}} 2^n \cdot \mathbb{1}_{(2^{-n},2^{-n+1})}, \qquad X^+ = \sum_{n \text{ odd}} 2^n \cdot \mathbb{1}_{(2^{-n},2^{-n+1})}.$$

Clearly

$$\mathbb{E}\Big[X^+\Big] = \mathbb{E}\Big[X^-\Big] = \infty.$$

(d) TODO: TO BE ADDED.



4. Probability by Le Gall

4.1 Integration of Measurable Functions

The following properties become very useful in the following chapters.

Proposition 4.1 Let f be a *non-negative* measurable function.

(i) We have
$$\int f d\mu = 0 \Leftrightarrow f = 0,\ \mu\ a.e.$$

Proof. For the proof we will have

(i) Define
$$A_n = \{f \geq 1/n\}.$$
Then by the Markov inequality
$$\mu(A_n) = \mu(\{f \geq 1/n\}) \leq \frac{\int f \, \mathrm{d}\, \mu}{1/n} = 0.$$
Also observe that $\{f > 0\} = \bigcup_n \{f \geq 1/n\} = \bigcup_n A_n$. Thus
$$\mu(\bigcup_n A_n) \leq \sum_n \mu(A_n) = 0.$$
This implies $\mu(\{f > 0\}) = 0$, thus $f = 0$ almost everywhere.

□

■ **Remark** For proof for part (i) in the proposition above, I also have the following proof using the conditional expectation. But I don't know how that translates to the integrals. Note that we have $\mathbb{E}\big[X\big] = 0$ and we want to show $\mathbb{P}(X = 0) = 1$. We can write

$$\begin{aligned} 0 = \mathbb{E}\big[X\big] &= \mathbb{E}\big[X\mathbb{1}_{X=0} + X\mathbb{1}_{X\neq 0}\big] = \mathbb{E}\big[X\mathbb{1}_{X=0}\big] + \mathbb{E}\big[X\mathbb{1}_{X\neq 0}\big] \\ &= \underbrace{\mathbb{E}\big[X|X=0\big]\mathbb{E}\big[\mathbb{1}_{X=0}\big]}_{0} + \mathbb{E}\big[X|X\neq 0\big]\mathbb{E}\big[\mathbb{1}_{X\neq 0}\big] \\ &= \mathbb{E}\big[X|X\neq 0\big]\mathbb{E}\big[\mathbb{1}_{X\neq 0}\big]. \end{aligned}$$

Observe that we have $\mathbb{E}\Big[X|X \neq 0\Big] > 0$ since X is a non-negative random variable. Thus we need to have $\mathbb{E}\Big[I_{X\neq 0}\Big] = \mathbb{P}(X \neq 0) = 1$.

4.2 Convergence of Random Variables

In the following we will have a quick review of the different definitions of convergence.

Definition 4.1 — Convergence almost surely. Let $X, X_1, X_2, \cdot$ be random variables. We say X_n converges to X almost surely if it converges to X point-wise on a set of measure 1. In other words

$$\mathbb{P}(X_n \to X) = 1,$$

where $X_n \to X$ should be interpreted by point wise convergence as $n \to \infty$.

Definition 4.2 — Convergence in Probability. Let $X, X_1, X_2, \cdots$ be random variables. We say that X_n converges to X in probability if for any $\epsilon > 0$ we have

$$\mathbb{P}(|X_n - X| \geq \epsilon) \to 0 \qquad \text{as } n \to \infty.$$

■ **Remark** Note that if for all $\epsilon > 0$ the convergence $\mathbb{P}(|X_n - X| \geq \epsilon) \to 0$ is fast enough so that it is summable, i.e. $\sum \mathbb{P}(|X_n - X| \geq \epsilon) < \infty$, then by Borel Cantelli we can conclude that $\mathbb{P}(|X_n - X| \geq \epsilon \ i.o.) = 0$, hence $\mathbb{P}(|X_n - X| < \epsilon \ a.a.) = 1$, this $X_n \to X$ almost surely.

Definition 4.3 — Convergence in L^p. Let $X, X_1, X_2, \cdots$ be random variables with $\mathbb{E}\Big[|X|^p\Big] < \infty$ as well as $\mathbb{E}\Big[|X_n|^p\Big] < \infty$ for all $n \in \mathbb{N}$, then we say that X_n converges to X in L^p if

$$\mathbb{E}\Big[|X_n - X|^p\Big] \to 0 \qquad \text{as } n \to \infty.$$

■ **Remark** With the notation $\|X\|_p = (\mathbb{E}\Big[|X|^p\Big])^{1/p}$ we can state the theorem above as: if $\|X\|_p < \infty$ and $\|X_n\|_p < \infty$ for all n, then $X_n \to X$ in L^p if

$$\|X_n - X\|_p \to 0.$$

4.2.1 Important Theorems

Theorem 4.1 — A.s. Convergence implies convergence in probability. Let $\{X_n\}$ be a collection of random variable such that $X_n \to X$ $a.s.$. Then $X_n \to X$ in probability.

Proof. Let $\epsilon > 0$ be given. Define

$$A_n = \{|X_n - X| \geq \epsilon\}.$$

Then if $\omega \in \limsup_n A_n$ we have $X_n(\omega) \not\to X(\omega)$. I.e. the set $\limsup_n A_n$ precisely contains all the points for which X_n does not converge to X on those points. Since we have a.s. convergence, then we have

$$\mathbb{P}(\limsup_n A_n) = 0.$$

Observe that since the sequence of events $\{\bigcup_{k\geq n} A_k\}_n$ is decreasing, then by continuity of probabilities

$$\lim_{n\to\infty} \mathbb{P}(\bigcup_{k\geq n} A_k) = \mathbb{P}(\bigcap_n \bigcup_{k\geq n} A_k) = \mathbb{P}(\limsup_n A_n) = 0.$$

Because $A_n \subset \bigcup_{k\geq n} A_n$, then $\mathbb{P}(A_n) \leq \mathbb{P}(\bigcup_{k\geq n} A_n)$ for all $n \in \mathbb{N}$ which implies $\mathbb{P}(A_n) \to 0$ as $n \to \infty$. □

Theorem 4.2 — Convergence in probability implies a.s. convergence along a subsequence. Let $\{X_n\}$ be a collection of random variables that converges to X in probability. Then there exists a subsequence $\{X_{n_k}\}$ that convergence a.s. to X.

Proof. Since $\{X_n\}$ converges to X in probability then for all $\epsilon > 0$

$$\mathbb{P}(A_n) \to 0,$$

where $A_n = \{|X_n - X| \geq \epsilon\}$. Since $\{\mathbb{P}(A_n)\}$ is a non-negative sequence that converges to 0, then there exists a subsequence $\{\mathbb{P}(A_{n_k})\}$ that is summable, i.e.

$$\sum_k \mathbb{P}(A_{n_k}) < \infty.$$

By Borel-Cantelli this implies

$$\mathbb{P}(\limsup_k A_{n_k}) = 0.$$

As stated in the proof of Theorem 4.1 $\limsup_n A_n$ is the set of all points $\omega \in \Omega$ such that $X_n(\omega)$ does not converge to $X(\omega)$. So $\{X_{n_k}\}$ converges to X a.s.. □

Theorem 4.3 — Convergence in probability vs Almost sure convergence. Let $\{X_n\}$ be a sequence of random variables converging to X in probability. Then if for any choice of $\epsilon > 0$

$$\sum_n \mathbb{P}(|X_n - X| \geq \epsilon) < \infty,$$

then $X_n \to X$ almost surely. Furthermore, if

$$\sum_n \mathbb{P}(|X_n - X| \geq \epsilon) < \infty = \infty,$$

and $\{X_n\}$ are independent, then $X_n \not\to X$ almost surely.

Proof. The proof follows immediately from Borel-Cantelli lemma by considering the facts that if for all $\epsilon > 0$

$$\mathbb{P}(|X_n - X| \geq \epsilon \ i.o.) = 0 \quad \text{then} \quad X_n \to X \text{ a.s..}$$

And if for all $\epsilon > 0$

$$\mathbb{P}(|X_n - X| \geq \epsilon \ i.o.) = 1 \quad \text{then} \quad X_n \not\to X \text{ a.s..}$$

□

■ **Remark** An important remark is that sumability of $\mathbb{P}(|X_n - X| \geq \epsilon)$ has information about a.s. convergence. But unsummability of $\mathbb{P}(|X_n - X| \geq \epsilon)$ has information about a.s. convergence only when $\{X_n\}$ are independent. For instance let X_n be defined as $X_n = \mathbb{1}_{[0,1/n]}$. Then $\mathbb{P}(|X_n - X| \geq \epsilon)$ is not summable. However since $\{X_n\}$ are not independent, we can not conclude that X_n is not converging to X almost surely (because it does converge a.s.). However, if we let $\{X_n\}$ be independent with $\mathbb{P}(X_n = 1) = 1/n$ and $\mathbb{P}(X_n = 0) = 1 - 1/n$, then X_n converges to X in probability, but not almost surely. Because for any $\epsilon > 0$, $\mathbb{P}(|X_n - X| \geq \epsilon \ i.o.) = 1$ (by Borel Cantelli Lemma).n

Theorem 4.4 — Convergence in L^p implies convergence in probability. Let $\{X_n\}$ be a sequence of random variables that converge in L^p. Then $X_n \to X$ in probability.

Proof. Let $\epsilon > 0$ be given. Then by Markov inequality we have

$$\mathbb{P}(|X_n - X| \geq \epsilon) \leq \mathbb{E}\Big[|X_n - X|\Big]/\epsilon = \|X_n - X\|_p/\epsilon.$$

On the other hand using Hölder's inequality (i.e. Corollary 4.2 in the book) we can write $\|X_n - X\|_1 \leq \|X_n - X\|_p$. So

$$\mathbb{P}(|X_n - X| \geq \epsilon) \leq \|X_n - X\|_p/\epsilon \to 0 \quad \text{as } n \to \infty,$$

where we have used the fact that $X_n \to X$ in probability.xc $\square$

■ **Remark** We can also do this without using the Hölder's inequality:

$$\mathbb{P}(|X_n - X| \geq \epsilon) = \mathbb{P}(|X_n - X|^p \geq \epsilon^p) \leq \frac{\mathbb{E}\Big[|X_n - X|^p\Big]}{\epsilon^p} = \frac{\|X_n - X\|_p^p}{\epsilon^p} \to 0 \quad \text{as } n \to \infty.$$

The following proposition is very important as it connects the convergence in probability to convergence in L^p (in moments), and its proof uses the Fatou's Lemma, Hölder's inequality as well as Minkowski's inequality.

Proposition 4.2 — Convergence in probability and convergence in L^p. Let $\{X_n\}$ be a sequence of random variables that converge to X in probability. Then X_n converges to X in L^p if there exists $r > p$ such that $\mathbb{E}\Big[|X_n|^r\Big] < \infty$ for all $n \in \mathbb{N}$.

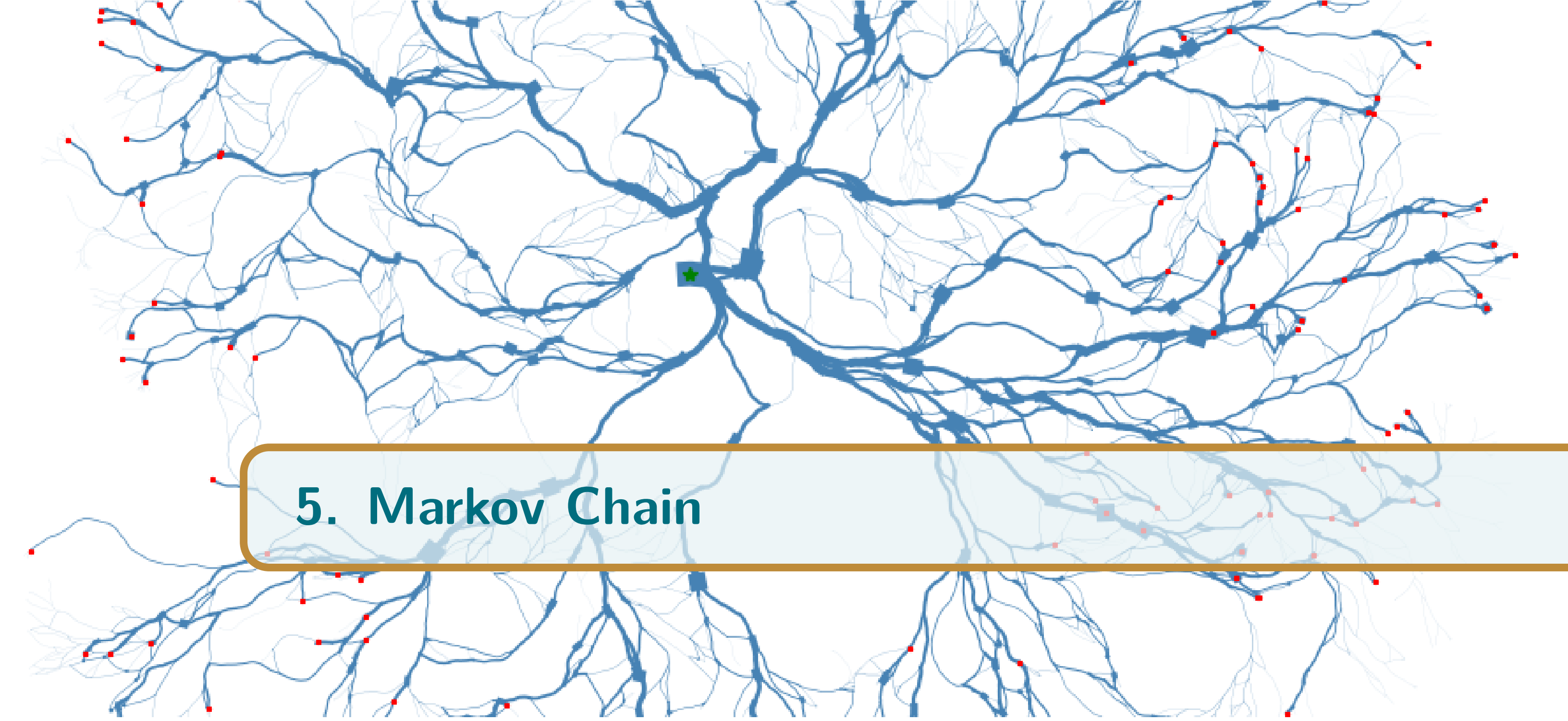
Proof. Since $X_n \to X$ in probability, there exists a subsequence $\{X_{n_k}\}$ that converges to X almost surely. In particular we can write $|X| = \liminf |X_n|$. So

$$\mathbb{E}\Big[|X|^r\Big] = \mathbb{E}\Big[\liminf_n |X_n|^r\Big] \leq \liminf \mathbb{E}\Big[|X_n|^r\Big] \leq C < \infty,$$

where we have used the hypothesis in the last inequality. Consider

$$\begin{aligned}
\mathbb{E}\Big[|X_n - X|^p\Big] &= \mathbb{E}\Big[|X_n - X|^p \mathbb{1}_{|X_n - X| < \epsilon}\Big] + \mathbb{E}\Big[|X_n - X|^p \mathbb{1}_{|X_n - X| \geq \epsilon}\Big] \\
&\leq \epsilon^p + \mathbb{E}\Big[(|X_n - X|^p)^{r/p}\Big]^{p/r} \mathbb{P}(|X_n - X| \geq \epsilon)^{1-p/r} \\
&= \epsilon^p + (\mathbb{E}\Big[|X_n - X|^r\Big]^{1/r})^p \mathbb{P}(|X_n - X| \geq \epsilon)^{1-p/r} \\
&= \epsilon^p + \|X_n - X\|_r^p \mathbb{P}(|X_n - X| \geq \epsilon)^{1-p/r} \\
&\leq \epsilon^p + (\|X\|_r + \|X_n\|_r)^p \mathbb{P}(|X_n - X| \geq \epsilon)^{1-p/r} \\
&\leq \epsilon^p + (2C)^p \mathbb{P}(|X_n - X| \geq \epsilon)^{1-p/r}.
\end{aligned}$$

Using the fact that $\mathbb{P}(|X_n - X| \geq \epsilon) \to 0$ as $n \to \infty$, then we conclude that $\mathbb{E}\Big[|X_n - X|^p\Big] \to 0$ as $n \to \infty$. So $X_n \to X$ in L^p. $\square$



5. Markov Chain

NOTE TO MYSELF: I prefer to develop whole theory of discrete Markov chains by defining the state space to be the set of symbols Σ (at most countable). This is beneficial, because then the sample space of any Markov chain will be the set of all infinite sequences (strings) from the symbols from Σ. At some point in the future, I might rewrite this chapter, working consistently with Σ as the state space.

We start with the definition of a Markov Chain.

Notation Let $(X_n)_{n\geq 0}$ be a Markov chain on the state space S, $x \in S$, and let E be an event. Then

$$\mathbb{P}_x(E) = \mathbb{P}(E|X_0 = x).$$

The following proposition will be one of our main tools throughout the chapter.

Proposition 5.1 — Conditional expansion. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\mathfrak{F}$ be a finite collection of events $\mathfrak{F} = \{F_1, F_2, \cdots, F_n\}$ that partitions Ω. I.e.

(i) $F_i \cap F_j = \varnothing \qquad i \neq j$,

(ii) $\bigcap_i F_i = \Omega$.

Let $E \in \mathcal{F}$ be any nonempty event. Then we can write

$$\mathbb{P}(E) = \sum_i \mathbb{P}(E|F_i)\mathbb{P}(F_i).$$

Proof. Since $\mathfrak{F}$ partitions Ω and $E \neq \varnothing$, then $\{E \cap F_i\}_i$ is a partition of E. Thus

$$\mathbb{P}(E) = \mathbb{P}(\bigcup_i (E \cap F_i)) = \sum_i \mathbb{P}(E \cap F_i) = \sum_i \mathbb{P}(E|F_i)\mathbb{P}(F_i).$$

This completes the proof. □

Proposition 5.2 — First step argument. Let $(X_n)_{n\geq 0}$ be a Markov chain on the state space S.

Let $x \in S$, and $W, Z \subset S$. Let B be any event. Then

$$\mathbb{P}_x(B) = \sum_{y:x\sim y} \mathbb{P}_y(B)P(x,y).$$

Proof. To prove the proposition above, we Let $E_i = \{X_0 = x, X_1 = y_i\}$ where $y_i \sim x$. So, in words, we say that the event E_i has occurred if $X_1 = y_i$. It is clear that $E_i \cap E_j = \varnothing$ where $i \neq j$. Thus $\bigcup_i (B \cap E_i) = B$. Thus

$$\mathbb{P}_x(B) = \sum_i \mathbb{P}_x(B \cap E_i) = \sum_i \mathbb{P}_x(B|E_i)\mathbb{P}_x(E_i).$$

In which $\mathbb{P}_x(E_i) = \mathbb{P}(E_i|X_0 = x) = \mathbb{P}(X_1 = y_i|X_0 = x) = P(x, y_i)$. Also

$$\mathbb{P}_x(B|E_i) = \mathbb{P}(B|X_1 = y_i, X_0 = x) = \mathbb{P}(B|X_1 = y_i) = \mathbb{P}_{y_i}(B),$$

in which we have used the Markov property. Thus we can write

$$P_x(B) = \sum_i \mathbb{P}_{y_i}(B)P(x, y_i).$$

□

5.1 Dissecting an Experiment

The idea of the Markov chain, random variables, probability spaces, etc. might be quite confusing when the setting of a particular random experiment becomes large. Here in this section, we are going to explain the details of a random experiment explicitly. The random experiment is the following

> Assume we have 6 urns, and we put ball at each urn successively. What is the probability that there will be exactly 3 non-empty urns after 9 balls have been distributed?

Sample Space

First, note that there are two things happening, that we can call experiments. First is that we are successively doing something, throwing dice and putting a ball inside the urn, and the second thing is that we can consider the whole thing to be a giant experiment by its own. Our convention, from now on, will be that we will call the whole thing as experiment, and we will consider each sub-experiment as sub-steps of the process. Intuitively, an experiment is something that we can repeat to observe different outcomes. It is true that the whole experiment is actually successive repetition of throwing dice, but we actually consider the largest meaningful setup to be our experiment and call the sub-experiment as the steps of the process. It is kind of confusing at first glance, but becomes more natural after a while. By the definition, the sample space of a random experiment, is the set of all possible outcomes. But what are the possible outcomes in our experiment? If we perform an experiment, then we can get an outcome like

$$134234521133456665342431255553214532145123453124563412314566654 35\dots.$$

This outcome basically is saying that the outcome of the first dice throw was 1, the second dice throw was 3, the third was 4, and etc. If we repeat the experiment, we will get other outcomes.

So the sample space is the set of all sequence of numbers consisting of $\{1, 2, 3, 4, 5, 6, 7, 8\}$. Thus, we can write

$$\Omega = \{13424\ldots, 54321\ldots, 65432\ldots, 12345\ldots, \cdots\}.$$

So, each outcome, consists of sequence of random variables $\{Y_t\}$, that is defined on the sample space $\Omega_{\text{dice}} = [1, 2, 3, 4, 5, 6]$. That is Y_{10} means the random variable associated with the dice throwing experiment at the time t of our main experiment. These random variables are all independent and identically distributed (i.e. they are iid).

Markov Chain

Let X_t be a random variable defined on Ω that represents the number of full urns at time t. Let $\omega = 2312345623421324534 23\ldots$ and $t = 10$. Then $X_10(\Omega) = 6$, since by the time t, we have put at least one ball at each urn and all of the urns are full. Since this is a Markov chain with state space $S = \{0, 1, 2, 3, 4, 5, 6\}$, we can draw a transition diagram, and analyze the question more carefully. This is what we have done in Example 5.3.

5.2 Solved Problems

■ **Example 5.1** An urn always contains 2 balls. Ball colors are red and blue. At each stage a ball is randomly chosen and then replaced by a new ball, which with probability 0.8 is the same color, and with probability 0.2 is the opposite color, as the ball it replaces. If initially both balls are red, find the probability that the fifth ball selected is red. [This question is from Ross]

Solution First, we need to translate this problem to a suitable Markov chain. There are many ways we can do so, each with its own pros and cons. The difference between all of these formulations come down to our choice for the state space (i.e. the co-domain of the random variable). For instance, we can assume that the state space is $S = \{RR, RB, BB\}$ that is the content of the Urn, or we can simply say that the state space is $S = \{0, 1, 2\}$ that is the number of red ball inside the Urn. Since these two sets are isomorphic (as there is a bijection between these two sets), but the actual choice depends on personal preference. Let's proceed with $S = \{0, 1, 2\}$. Then, we need to determine the transition matrix. We can do so by doing the first step argument. We start with $P(0, 0)$.

$$P(0,0) = \mathbb{P}(X_1 = 0|X_0 = 0) = \mathbb{P}(X_1 = 0|X_0 = 0, E_R)\underbrace{\mathbb{P}(E_R|X_0 = 0)}_{0} + \underbrace{\mathbb{P}(X_1 = 0|X_0 = 0, E_B)}_{0.8}\underbrace{\mathbb{P}(E_B|X_0 = 0)}_{1},$$

where E_R is the event at which a red balls is drawn from the Urn, while E_B is the event where a blue ball is drawn. The reason behind the values for the term above are very straight forward. For instance $\mathbb{P}(E_R|X_0 = 0) = 0$ because given the fact that number of red balls in the Urn is zero ($X_0 = 0$), then the probability that we draw a red ball is zero (as there is not red balls in the Urn). For the term $\mathbb{P}(X_1 = 0|X_0 = 0, E_B) = 0.8$, because given there is no red balls inside the urn, and also given the fact that the drawn ball is blue, the probability of ending up at the state $X_1 = 0$ (i.e. still no red balls) is that probability is that we replaced the drawn ball with a blue ball (same color) which has the probability 0.8. Similarly, we can calculate the first step transition probabilities.

$$P(0,1) = \mathbb{P}(X_1 = 1|X_0 = 0) = \mathbb{P}_0(X_1 = 1) = \mathbb{P}_0(X_1 = 1|E_R)\underbrace{\mathbb{P}_0(E_R)}_{0} + \underbrace{\mathbb{P}_0(X_1 = 1|E_B)}_{0.2}\underbrace{\mathbb{P}_0(E_B)}_{1} = 0.2,$$

$$P(0,2) = \mathbb{P}(X_1 = 2|X_0 = 0) = \mathbb{P}_0(X_1 = 2) = \mathbb{P}_0(X_1 = 2|E_R)\underbrace{\mathbb{P}_0(E_R)}_{0} + \underbrace{\mathbb{P}_0(X_1 = 2|E_B)}_{0}\underbrace{\mathbb{P}_0(E_B)}_{1} = 0,$$

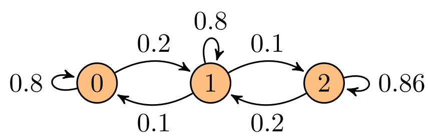
$$P(1,0) = \mathbb{P}(X_1 = 0|X_0 = 1) = \mathbb{P}_1(X_1 = 0) = \underbrace{\mathbb{P}_1(X_1 = 0|E_R)}_{0.2}\underbrace{\mathbb{P}_1(E_R)}_{0.5} + \underbrace{\mathbb{P}_1(X_1 = 0|E_B)}_{0}\underbrace{\mathbb{P}_1(E_B)}_{0.5} = 0.1.$$

$$P(1,1) = \mathbb{P}(X_1 = 1|X_0 = 1) = \mathbb{P}_1(X_1 = 1) = \underbrace{\mathbb{P}_1(X_1 = 1|E_R)}_{0.8}\underbrace{\mathbb{P}_1(E_R)}_{0.5} + \underbrace{\mathbb{P}_1(X_1 = 1|E_B)}_{0.8}\underbrace{\mathbb{P}_1(E_B)}_{0.5} = 0.8.$$

and so on. Then we will have the following transition matrix for this problem.

$$M = \begin{pmatrix} 0.8 & 0.2 & 0 \\ 0.1 & 0.8 & 0.1 \\ 0 & 0.2 & 0.8 \end{pmatrix}$$

with the following graph



Now, we need to compute the probability that the fifth ball drawn is red. This means that we have already drawn four balls, and now we want to draw the fifth one. So, we need to consider the 4 step transition matrix, i.e. M^4. Then

$$M^4 = \begin{pmatrix} 0.4872 & 0.4352 & 0.0776 \\ 0.2176 & 0.5648 & 0.2176 \\ 0.0776 & 0.4352 & 0.4872 \end{pmatrix}$$

Given that we have started with 2 red balls, then the probability of finding the Urn with 0 red balls is 0.0776, with 1 red ball is 0.4352, and with 2 red balls is 0.4872. So the probability that the next drawn balls is red is

$$\mathbb{P}(E_R) = \underbrace{\mathbb{P}(E_R|X_4 = 0)}_{0}\underbrace{\mathbb{P}(X_4 = 0)}_{0.0776} + \underbrace{\mathbb{P}(E_R|X_4 = 1)}_{0.5}\underbrace{\mathbb{P}(X_4 = 1)}_{0.4352} + \underbrace{\mathbb{P}(E_R|X_4 = 2)}_{1}\underbrace{\mathbb{P}(X_4 = 2)}_{0.4872} = 0.7048.$$

■

■ **Example 5.2 — Turning non-Markov processes to Markov-chain.** Suppose that whether or not it rains today depends on previous weather conditions through the last two days. Specifically, suppose that if it has rained for the past two days, then it will rain tomorrow with probability 0.7; if it rained today but not yesterday, then it will rain tomorrow with probability 0.5; if it rained yesterday but not today, then it will rain tomorrow with probability 0.4; if it has not rained in the past two days, then it will rain tomorrow with probability 0.2. [This question is from Ross]. Given that it rained on Monday and Tuesday, what is the probability that it will rain on Thursday?

Solution This random process is not a Markov chain, the value of the random variable at the next state, depends on two previous states. However, we can turn this into a Markov chain. Define the following states

RR: Rained yesterday and today.
$R\overline{R}$: Rained yesterday, but not today.
$\overline{R}R$: Not rained yesterday, but rained today.
$\overline{RR}$: Not rained yesterday and today.

Suppose that we are at state RR. Suppose that it rained yesterday and also today. Thus we are at state RR. If it rains tomorrow, then we will be still at state RR. That is because, That is because the yesterday of tomorrow is today! So if it rains tomorrow, since today (yesterday of tomorrow) was also rainy, thus if it rains tomorrow then we will stay at state RR. If it does not rain tomorrow, then we will get to state $\overline{R}R$. The following matrix is the transition matrix for this Markov chain

$$M = \begin{pmatrix} 0,7 & 0.3 & 0 & 0 \\ 0 & 0 & 0.4 & 0.6 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.2 & 0.8 \end{pmatrix}$$

Now, to calculate the probability of raining on Thursday, given it rained on Monday and Tuesday, we first need to calculate the two step transition probability.

$$M^2 = \begin{pmatrix} \boxed{0.49} & 0.21 & \boxed{0.12} & 0.18 \\ 0.2 & 0.2 & 0.12 & 0.48 \\ 0.35 & 0.15 & 0.2 & 0.3 \\ 0.1 & 0.1 & 0.16 & 0.64 \end{pmatrix}$$

The probability to rain on Thursday is the sum of the boxed elements in the matrix above. So the desired probability is

$$p = 0.61.$$

■

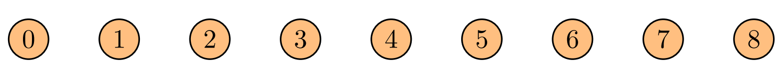
■ **Example 5.3** Suppose that balls are successively distributed among 8 urns, with each ball being equally likely to be put in any of these urns. What is the probability that there will be exactly 3 nonempty urns after 9 balls have been distributed? [Question from Ross]

Solution Before going through the solution, it might be more informative to explicitly write down what is the sample space Ω. At each time step, we basically throwing a 8 sided dice, and then put a ball at the urn number i if the output of the dice is i. So, each time we repeat the experiment, we will get a sequence of number each of which is one of $1, 2, \cdots, 8$. So the sample space will be the set of all sequences consisting of number $1, \cdots, 8$.

$$\Omega = \{21342\ldots, 44513\ldots, 11234\ldots, 88432\ldots, \cdots\}.$$

So, the outputs of the throwing dice at different steps are independent and identically distributed random variables. I.e. for a fixed $\omega \in \Omega$, The t-th element of the sequence is a random variable Y_t and all of the random variables $\{Y_t\}_t$ are independent and identically distributed. Note that the sample space associated with these random variables is $\{1, 2, 3, 4, 5, 6, 7, 8\}$ (i.e. the sample space of a 8 sided dice experiment).w

Let the random variable X_n be the number of filled (non-empty) urns at step n. So the state space will be $S = \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$, which is represented in the following graph.



This picture is not yet complete and we need to include the transition probabilities. We will do so by the first step argument. First, observe that $P(0, 0) = 0$, because if we start with all of the urns empty, then after one step, we have put a ball somewhere, thus it is impossible to end up with zero filled urn. Similarly, $P(8,8)=1$, that is because if all of the urns are filled, then adding any new ball somewhere to any of the urns will keep the number of filled urns at 8. Then for $X_0=n$, i.e. starting with n filled urns, we have

$$\mathbb{P}_n(X_1=n-1)=0.$$

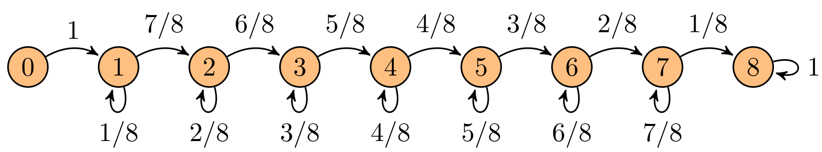
That is because starting with n filled urns, after doing one step, it is not possible to have less urns filled. I.e. after each step, we can either end up with more filled urns or the same number of filled urns. For $P(n,n)$, define the event E be the event of putting the ball in any of the filled urns. Thus E^c will be the probability of putting the ball at one of the empty urns.

$$\mathbb{P}_n(X_1=n)=\underbrace{\mathbb{P}_n(X_1=n|E)}_{1}\underbrace{\mathbb{P}_n(E)}_{n/8}+\underbrace{\mathbb{P}_n(X_1=n|E^c)}_{0}\underbrace{\mathbb{P}_n(E^c)}_{(8-n)/8}=\frac{n}{8}.$$

Now for $\mathbb{P}(n,n+1)$ we can write

$$\mathbb{P}_n(X_1=n+1)=\underbrace{\mathbb{P}_n(X_1=n+1|E)}_{0}\underbrace{\mathbb{P}_n(E)}_{n/8}+\underbrace{\mathbb{P}_n(X_1=n+1|E^c)}_{1}\underbrace{\mathbb{P}_n(E^c)}_{(8-n)/8}=1-\frac{n}{8}.$$

Thus the completed graph will be



The corresponding transition matrix will be

$$M=\begin{pmatrix}0&1&0&0&0&0&0&0&0\\0&1/8&7/8&0&0&0&0&0&0\\0&0&1/8&6/8&0&0&0&0&0\\0&0&0&3/8&5/8&0&0&0&0\\0&0&0&0&4/8&4/8&0&0&0\\0&0&0&0&0&5/8&3/8&0&0\\0&0&0&0&0&0&6/8&2/8&0\\0&0&0&0&0&0&0&7/8&1/8\\0&0&0&0&0&0&0&0&1\end{pmatrix}$$

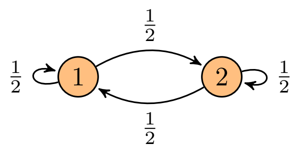
The probability that after 9 steps, there are exactly three empty urns is $(M^9)_{(0,3)}$, which is

$$p=(M^9)_{(0,3)}\approx 0.007572.$$

■

■ **Example 5.4** It is a good practice to derive the value of the transition probability of a simple Markov chain using the first principles. Consider the Markov chain representing a lamp that turns on with probability 1/2 and turns off with probability 1/2, and stays at the old state with probability 1/2. Thus we will have the following diagram for this Markov chain.
In this example, the state space is $S=\{0,1\}$, and the sample space is

$$\Omega=\{(x_1,x_2,\cdots):x_i\in S\}$$



which is basically the set of all sequences of one's and zero's. Given this, the random variables $(X_n)_n$ defined t be

$$X_n(\omega) = x_n,$$

where $\omega \in \Omega$ and x_n is the n-th letter in ω. Intuitively speaking, we know that

$$P(1,0) = \mathbb{P}(X_{n+1} = 1 | X_n = 0) = \frac{1}{2}.$$

However, here we want to derive that number more explicitly by working directly with the elements of the probability space. First, we need to determine the event associated with $X_{n+1} = 1$. This is the event that has elements where the $n+1$-th position is 1. I.e.

$$E = \{(x_1, x_2, \cdots, x_n, 1, x_{n+2}, \cdots) : x_i \in S\}.$$

Similarly, we have

$$F = \{(x_1, x_2, \cdots, x_{n-1}, 0, x_{n+1}, \cdots) : x_i \in S\}.$$

So we have

$$\mathbb{P}(X_{n+1} = 1 | X_n = 0) = \mathbb{P}(E|F) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)} = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F \cap E) + \mathbb{P}(F \cap E^c)} = \frac{\frac{1}{|\Omega|}}{\frac{1}{|\Omega|} + \frac{1}{|\Omega|}} = \frac{1}{2}.$$

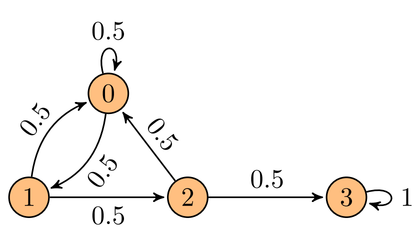
Note that $\mathbb{P}(E \cap F) = \frac{1}{|\Omega|}$, since out of many combinations of the sequence of zeros and ones, there is one one sequence whose n-th place is 0 and $n+1$-th place is 1. Furthermore, $\mathbb{P}(F \cap E^c) = \frac{1}{|\Omega|}$ as there is only one string where its n-th and $(n+1)$-th string are both zero. ■

■ **Example 5.5** In a sequence of independent flips of a fair coin, let N denote the number of flips until there is a run of three consecutive heads. Find

(a) $\mathbb{P}(N \leq 8)$,

(b) $\mathbb{P}(N = 8)$.

Solution Let X_n denote the number of consecutive heads at step n. For instance for the outcome $\omega \in \Omega$ where $\omega = HTTHTTHHHTTHT\ldots$, $X_2(\omega) = 0$ since the second symbol is T thus there is no consecutive heads. But $X_4(\omega) = 1$, as there is one consecutive heads at step 4. Lastly $X_9(\omega) = 3$, since there is three consecutive heads at step 9. This Markov chain will have the following transition diagram.



The transition probabilities are simply computed by the first step argument. For instance, for $P(0,1)$ we have

$$\mathbb{P}_0(X_1 = 1) = \underbrace{\mathbb{P}_0(X_1 = 1|H)}_{1}\underbrace{\mathbb{P}_0(H)}_{1/2} + \underbrace{\mathbb{P}_0(X_1 = 1|T)}_{0}\underbrace{\mathbb{P}_0(T)}_{1/2},$$

where H is the event that the flipped coin is heads and $H^c = T$. The transition matrix for this Markov chain will be

$$M = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0.5 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

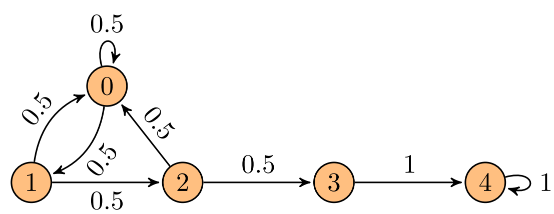
Since the state 3 is an absorbing state, then if we get there we will be there for the rest of our life! Thus the probability that the random walker has got there for $N \leq 8$ is simply $(M^8)_{(0,3)}$. Then

$$\mathbb{P}(N \leq 8) = 0.4180.$$

Now for part (b), the probability that the random walker has arrived at the state 3 right at the step 8, is

$$\mathbb{P}(N = 8) = \mathbb{P}(N \leq 8) - \mathbb{P}(N \leq 7) = 0.0508.$$

There is yet another approach that we can compute the probability $\mathbb{P}(N = 8)$. To do this, we need to consider 4 states $S = \{0, 1, 2, 3, 4\}$ where the state 4 is of when 3 consecutive heads has occurred at the past. So when the random walker enters the state 3 at some time, it moves to the state 4 at the next time and remains there forever. The state diagram will be



Then the transition matrix will be

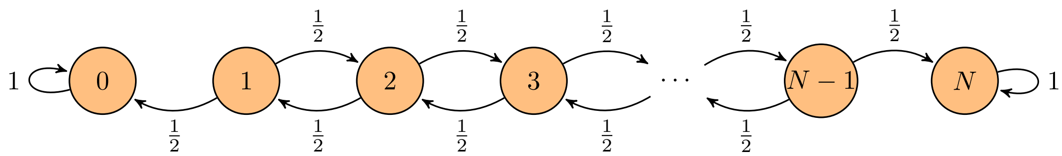
$$M = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0.5 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Then the probability $\mathbb{P}(N = 8) = (M^8)_{(0,3)} = 0.05080$.

■

■ **Example 5.6 — Gambler's Ruin.** Suppose Alice and Bob have in total of N coins. Alice and Bob play a game with a fair coin. When Alice wins, gets a coin from Bop, and vise versa. What is the probability that Alice wins if she starts with $0 \leq a \leq N$ coins.

Solution There are many ways to tackle a probability problem like this and the solution presented here is not the only way to find the solution to this problem. We want to model this with Markov chain whose state space is $\{0, 1, 2, \cdots, N\}$. Thus X_n represents the fortune of Alice after playing the games for n times.



Let p_a be the probability of Alice wining if she starts with a coins. First, observe that $p_0 = 0$ and $p_N = 1$. Let E denote that event of Alice wining the whole game. Also, let F_1 be the event in which she looses the first game and F_2 the event in which she wins the first game. Then

$$p_a = \mathbb{P}_a(E) = \underbrace{\mathbb{P}_a(E|F_1)}_{\mathbb{P}(E|F_1, X_0=a)} \mathbb{P}(F_1) + \underbrace{\mathbb{P}_a(E|F_1^c)}_{\mathbb{P}(E|F_1^c, X_0=a)} \mathbb{P}(F_1^c)$$

(note that this identity is actually true for any set F_1, but here F_1 is the specific event explained above). The probability that she looses or wins the first game is $\frac{1}{2}$. Also, observe that $\mathbb{P}_a(E|F_1) = p_{a+1}$ (since if she wins the first game she will have one more coin) and $\mathbb{P}_a(E|F_1^c) = p_{a-1}$. Thus

$$p_a = \frac{1}{2}p_{a+1} + \frac{1}{2}p_{a-1}.$$

Now we can solve this recurrent equation with the characterization polynomial which is $2 = X + 1/X$ or $X^2 - 2X + 1 = (X-1)^2 = 0$. Thus the characteristic polynomial has a double root. Thus

$$p_a = (Aa + B)(1)^a = Aa + B.$$

Since $p_0 = 0,\ p_N = 1$, then it turns out that

$$p_a = \frac{a}{N}.$$

■

■ **Example 5.7 — Gambler's Ruin with Draw.** Let Alice and Bob play Rock-Paper-Scissors. If Alice and Bob has a total of N coins, and at each play, the winner gets one coin from the loser, what is the probability that Alice will win the game if he starts with a coins. When they draw, then they repeat the game (or equivalently, they play another game without any coins exchange).

Solution We need to do a first step analysis similar to what we did before. Let E be the event that Alice wins the whole game, and the event $F = F_{-1} \cup F_0 \cup F_1$ where

F_{-1}: Alice loses the first game,
F_0: Alice draws the first game,
F_1: Alice wins the first game.

It is clear that $\mathbb{P}(F) = 1$, since the components are mutually disjoint. Thus $E \cap F_{-1},\ E \cap F_0,\ E \cap F_1$ are also mutually disjoints where. Thus we can write

$$\mathbb{P}_a(E) = \mathbb{P}_a(E \cap F_{-1}) + \mathbb{P}_a(E \cap F_0) + \mathbb{P}_a(E \cap F_1) = \mathbb{P}_a(E|F_{-1})\mathbb{P}_a(F_{-1}) + \mathbb{P}_a(E|F_0)\mathbb{P}_a(F_0) + \mathbb{P}_a(E|F_1)\mathbb{P}_a(F_1).$$

Since the game is fair we know

$$\mathbb{P}_a(F_{-1}) = \mathbb{P}_a(F_0) = \mathbb{P}_a(F_1) = \frac{1}{3}.$$

Furthermore, we know

$$\mathbb{P}_a(E|F_{-1}) = p_{a-1}, \qquad \mathbb{P}_a(E|F_0) = p_a, \qquad \mathbb{P}_a(E|F_1) = p_{a+1}.$$

Thus the first step analysis will lead to the following identity.

$$\mathbb{P}_a(E) = p_a = \frac{1}{3}(p_{a-1} + p_a + p_{a+1}),$$

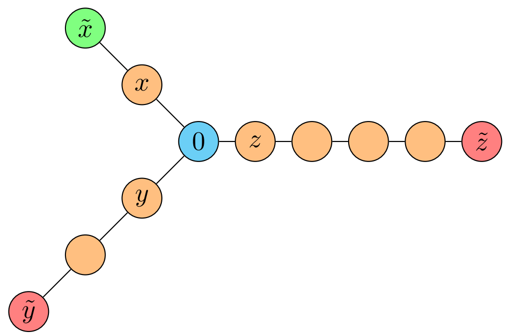
which after simplification becomes

$$2p_a = p_{a-1} + p_{a+1},$$

which is the same recursive formula we got in the previous example. So the possibility of the draw, will not change the behaviour of the system.

■

■ **Example 5.8** Consider the a simple random walker on the following graph. Let $B = \{T_{\tilde{x}} < T_{\{\tilde{z},\tilde{y}\}}\}$. Compute the probability $\mathbb{P}_0(B)$.



Solution This problem is simply asking what is the probability that we hit $\tilde{x}$ state before hitting any of $\tilde{y}$ or $\tilde{z}$ states, given the fact that the random walker starts from the state 0. To keep unnecessary details out of the way, we have only labeled the vertices that we will use in our analysis. We will have the following notation to simplify the solution

$$p_v = \mathbb{P}_v(B),$$

where v is any vertex in the graph. Note that starting at 0, i.e. $X_0 = 0$, then going to any of the states x, y, or z, are mutually disjoint events, and the probability of the union of these events is one. With our first time step analysis (see Proposition 5.2) we can write

$$\mathbb{P}_0(B) = \frac{1}{3}(p_x + p_y + p_z).$$

Now we need to analyze each of terms in the RHS. Let's start with p_z. Consider two events $\{T_0 < T_{\tilde{z}}\}$ and $\{T_0 > T_{\tilde{z}}\}$, where the first time is the event where the random walker hits the 0 state before hitting the $\tilde{z}$ step first, and the second one is the vice versa. These two events are disjoint and the probability of the union is 1. Thus we write the conditional expansion of p_z based on these events

$$p_z = \mathbb{P}_z(B) = \mathbb{P}_z(B|T_0 < T_{\tilde{z}})\mathbb{P}_z(T_0 < T_{\tilde{z}}) + \mathbb{P}_z(B|T_0 > T_{\tilde{z}})\mathbb{P}_z(T_0 > T_{\tilde{z}}).$$

We know that $\mathbb{P}_z(B|T_0 > T_{\tilde{z}}) = \mathbb{P}(B|X_0 = z, X_i = \tilde{z})$ for some $i > 0$. From Markov property it follows that

$$\mathbb{P}(B|X_0 = z, X_i = \tilde{z}) = \mathbb{P}(B|X_i = \tilde{z}) = \mathbb{P}(B|X_0 = \tilde{z}) = p_{\tilde{z}}.$$

Also $\mathbb{P}_z(B|T_0 < T_{\tilde{z}}) = \mathbb{P}_0(B) = p_0$ by the Markov property. Lastly, $\mathbb{P}_z(T_0 < T_{\tilde{z}})$ is determined by the Gambler's ruin method we say before, which is basically

$$\mathbb{P}_z(T_0 < T_{\tilde{z}}) = \frac{5}{4}, \qquad \mathbb{P}_z(T_0 > T_{\tilde{z}}) = \frac{1}{5}.$$

By doing the same kind of analysis for p_x as well as p_y we will get

$$p_z = \frac{4}{5}p_0, \qquad p_y = \frac{2}{3}p_0, \qquad p_x = \frac{1}{2}p_0 + \frac{1}{2}.$$

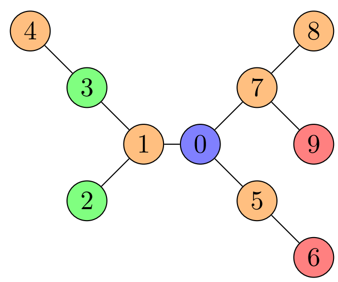
Now by substituting in the identity we got from the first time step argument, we can fine that

$$p_0 = \frac{15}{31},$$

And this completes our solution for the problem.

■

■ **Example 5.9** Consider the graph $\gamma = (V, E)$ drawn below. Set $Z = \{2, 3\}$, and $W = \{6, 9\}$. Compute $\mathbb{P}_0(T_Z < T_W)$. In colors: we start at blue, win if we reach green, and lose of we reach red.



■

Solution As always, we start with our powerful tool in hand, which is the first step argument (which is basically a special form of the more general conditional expansion). We start with first step argument at state 0. We will get

$$\mathbb{P}_0(B) = \frac{1}{3}(\mathbb{P}_1(B) + \mathbb{P}_7(B) + \mathbb{P}_5(B)),$$

and now we need to analyze each of the terms in the right hand side. We start with $\mathbb{P}_5(B)$ which is the most straight forward one. As we saw in the last example, we can analyze this state with a conditional expansion on the two disjoint events, whose union probability is 1. Let those two events be $\{T_6 < T_0\}$ (where the random walker hits the state 6 before hitting the state 0), and $\{T_6 > T_0\}$, where the random walker hits the state 0 before hitting the state 6. Thus the expansion will be

$$\mathbb{P}_5(B) = \mathbb{P}_5(B|T_6 < T_0)\mathbb{P}_5(T_6 < T_0) + \mathbb{P}_5(B|T_6 > T_0)\mathbb{P}_5(T_6 > T_0).$$

We know that if we hit the state 6 before 0, we have no chance to hit any of the green states (we will lose). Thus

$$\mathbb{P}_5(B|T_6 < T_0) = 0.$$

And from the Gambler's ruin we know that $\mathbb{P}_5(T_6 > T_0) = 1/2$, and from the Markov property we know that $\mathbb{P}_5(B|T_6 > T_0) = \mathbb{P}_0(B)$, because the conditional probability $\mathbb{P}_5(B|T_6 > T_0)$ is basically stating what is the probability of B happening, if we start from 5 and $X_i = 0$ for some i in the future. Thus

$$\mathbb{P}_5(B) = \frac{\mathbb{P}_0(B)}{2}.$$

Now, we need to analyze the term $\mathbb{P}_1(B)$. Again, at this step, we do another first step analysis.

$$\mathbb{P}_1(B) = \frac{1}{3}(\underbrace{\mathbb{P}_3(B)}_{=1} + \underbrace{\mathbb{P}_2(B)}_{=1} + \mathbb{P}_0(B)) = \frac{2+\mathbb{P}_0(B)}{3}.$$

Note that from the assumption, we know that if we reach any of green states, then we are declared winner, that is why we have $\mathbb{P}_3(B) = \mathbb{P}_2(B) = 1$. Now it only remains to analyze the term $\mathbb{P}_7(B)$. Again, similar to the case above, we do a first time step argument

$$\mathbb{P}_7(B) = \frac{1}{3}(\mathbb{P}_0(B) + \underbrace{\mathbb{P}_8(B)}_{=\mathbb{P}_7(B)} + \underbrace{\mathbb{P}_9(B)}_{=0}) \implies \mathbb{P}_7(B) = \frac{\mathbb{P}_0(B)}{2}.$$

Note that $\mathbb{P}_8(B) = \mathbb{P}_7(B)$ by a first stem analysis when starting at the state 8. Putting all of these terms back to the original identity we derived the first, we can conclude that

$$p_0 = \mathbb{P}_0(B) = \frac{2}{5}.$$

5.3 Solved Problems

■ **Problem 5.1** The French roulette game has slots numbered from 0 to 36. The slot 0 is green, Among the slots from 1 to 36, 18 are black and 18 are red. Alex goes to a casino to play roulette. Their strategy is to always bet "red". They start with 50 coins, play 1 coin each turn, and stop when reaching 100 or getting broke.

(a) What is the probability that Alex reaches 100?

(b) How many coins should Alex start with to have about 50% chance to reach 100?

Solution (a) Let B be the event $B = \{T_{100} < T_0\}$ and we are looking for $\mathbb{P}_a(B)$ where $0 \leq a \leq 100$ and indicates the number of coins we are starting with. First observe that

- $p_0 = 0$: Since if we start with zero coins we are already broken and the game is over.
- $p_{100} = 1$: Since if we start with 100 coins then we won the game and the game is finished.

To compute the probability for intermediate values of a, we do the first step argument. Let WF be the event where Alex wins the first bet, and LF the event where Alex loses the first bet. Then we can write

$$p_a = \mathbb{P}_a(B) = \mathbb{P}_a(B|WF)\mathbb{P}_a(WF) + \mathbb{P}_a(B|LF)\mathbb{P}_a(LF).$$

Since there are 18 red spots, then the chance to win the first bet is

$$\mathbb{P}_a(WF) = \frac{18}{37}.$$

and since there are 19 non-red spots in total, then the chance to win is

$$\mathbb{P}_a(LF) = \frac{19}{37}.$$

Also, from Markov property, we know that

$$\mathbb{P}_a(B|WF) = p_{a+1}, \qquad \mathbb{P}_a(B|LF) = p_a.$$

Thus the first step argument formula will be

$$p_a = \frac{18}{37}p_{a+1} + \frac{19}{37}p_{a-1} \implies \boxed{37p_a = 18p_{a+1} + 19p_{a-1}}.$$

The characteristic equation for the recursive equation is

$$37 = 18x + \frac{19}{x} \implies \boxed{18x^2 - 37x + 19 = 0}.$$

We can write it as $(x-1)(18x-19) = 0$. Thus the roots will be

$$r_1 = 1, \qquad r_2 = \frac{19}{18}.$$

So

$$p_a = A + Br_2^a.$$

To fine A and B we use the fact $p_0 = 0$, and $p_{100} = 1$. Then $A = -B$, and $A = 1/(1 - r_2^{100})$. Thus

$$\boxed{p_a = \frac{1 - r_2^a}{1 - r_2^{100}}}.$$

(b) We basically need to compute find a for which $p_a = 1/2$. Thus we need to solve for a

$$\frac{1 - r_2^a}{1 - r_2^{100}} = \frac{1}{2}.$$

After some algebra we will find

$$\boxed{a = \frac{\ln\left(\frac{1+r_2^{100}}{2}\right)}{\ln(r_2)} \approx 87.26}.$$

Thus we need to start with at least 88 coins to have a 50% chance of winning.

□

■ **Problem 5.2** There are 6 coins on a table, each showing heads (H) or tails (T). In each step we

- Select uniformly one of the coins.
- If it is heads, toss it and replace on the table (with random side).
- If it sis tails, toss it. If it comes up heads, leave it at that. If it comes up tails, toss it a second time, and leave the result as it is. Let X_n be the number of heads showing after n such steps. Answer the following questions

  (a) Determine the transition probabilities for this Markov chain.

  (b) Draw the transition diagram and write the transition matrix.

  (c) What is $\mathbb{P}(X_2 = 4 | X_0 = 5)$?

Solution (a) To compute the transition probabilities, we need to perform the first step analysis. Let the events

$$I = \{X_1 = a+1\}, \qquad S = \{X_1 = a\}, \qquad D = \{X_1 = a-1\},$$

where $0 \le a \le 6$ is the number of heads. So to compute the transition probabilities, we need to compute

$$P(a, a+1) = \mathbb{P}_a(I), \qquad P(a,a) = \mathbb{P}_a(S), \qquad P(a, a-1) = \mathbb{P}_a(D).$$

We start with $\mathbb{P}_a(I)$. Let ST be the event where the selected coin is tails, and SH be the event where the selected coin is heads. These two events are disjoint and the probability of their union is 1, thus

$$\mathbb{P}_a(I) = \underbrace{\mathbb{P}_a(I|SH)}_{\text{see Eq (2.I.1)}} \underbrace{\mathbb{P}_a(SH)}_{\frac{a}{6}} + \underbrace{\mathbb{P}_a(I|ST)}_{\text{see Eq (2.I.2)}} \underbrace{\mathbb{P}_a(ST)}_{\frac{6-a}{6}}. \tag{2.I}$$

Note that if we start with a coins heads, then the chance we choose a random coin from the table and find it heads is $\frac{a}{6}$, hence $\mathbb{P}_a(SH) = \frac{a}{6}$, and $\mathbb{P}_a(ST) = \frac{6-a}{6}$. Now we need to expand the remaining terms with appropriate conditioning. Let TT be the event where we toss a coin and find it tails and TH be the event where we toss a coin and find it heads. Thus we can write

$$\mathbb{P}_a(I|SH) = \underbrace{\mathbb{P}_a(I|SH,TH)}_{0} \underbrace{\mathbb{P}_a(TS)}_{\frac{1}{2}} + \underbrace{\mathbb{P}_a(I|SH,TT)}_{0} \underbrace{\mathbb{P}_a(TT)}_{\frac{1}{2}}. \tag{2.I.1}$$

Note that $\mathbb{P}_a(TT) = \mathbb{P}_a(TH) = \frac{1}{2}$, since the coin tossing is fair. Also, note that $\mathbb{P}_a(I|SH,TH) = \mathbb{P}_a(I|SH,TT) = 0$ since if we select a heads, and then toss it, finding it either heads or tails will not increase the total number of heads on the table. Similarly, for the other term in (2.1) we have

$$\mathbb{P}_a(I|ST) = \underbrace{\mathbb{P}_a(I|ST,TH)}_{1} \underbrace{\mathbb{P}_a(TH)}_{\frac{1}{2}} + \underbrace{\mathbb{P}_a(I|ST,TT)}_{\text{see Eq (2.I.3)}} \underbrace{\mathbb{P}_a(TT)}_{\frac{1}{2}}. \tag{2.I.2}$$

Now we need to expand the remaining terms in the equation above.

$$\mathbb{P}_a(I|ST,TT) = \underbrace{\mathbb{P}_a(I|ST,TT,TH)}_{1} \mathbb{P}_a(TH) + \underbrace{\mathbb{P}_a(I|ST,TT,TT)}_{0} \mathbb{P}_a(TT) = \frac{1}{2}. \tag{2.I.3}$$

Putting all together we can write

$$\boxed{P(a, a+1) = \mathbb{P}_a(I) = \frac{6-a}{8}}.$$

Similarly, we can compute other transition probabilities. For instance for $\mathbb{P}_a(S)$ we can write

$$\mathbb{P}_a(S) = \underbrace{\mathbb{P}_a(S|SH)}_{\text{see Eq (2.S.1)}} \underbrace{\mathbb{P}_a(SH)}_{\frac{a}{6}} + \underbrace{\mathbb{P}_a(S|ST)}_{\text{see Eq (2.S.2)}} \underbrace{\mathbb{P}_a(ST)}_{\frac{6-a}{6}}. \tag{2.S}$$

and for the remaining terms we can write

$$\mathbb{P}_a(S|SH) = \underbrace{\mathbb{P}_a(S|SH,TH)}_{1} \underbrace{\mathbb{P}_a(TS)}_{\frac{1}{2}} + \underbrace{\mathbb{P}_a(S|SH,TT)}_{0} \underbrace{\mathbb{P}_a(TT)}_{\frac{1}{2}}, \tag{2.S.1}$$

and

$$\mathbb{P}_a(S|ST) = \underbrace{\mathbb{P}_a(S|ST,TH)}_{0} \underbrace{\mathbb{P}_a(TH)}_{\frac{1}{2}} + \underbrace{\mathbb{P}_a(S|ST,TT)}_{\text{see Eq (2.S.3)}} \underbrace{\mathbb{P}_a(TT)}_{\frac{1}{2}}. \tag{2.S.2}$$

And for the remaining term above

$$\mathbb{P}_a(S|ST,TT) = \underbrace{\mathbb{P}_a(S|ST,TT,TH)}_{0}\mathbb{P}_a(TH) + \underbrace{\mathbb{P}_a(S|ST,TT,TT)}_{1}\mathbb{P}_a(TT) = \frac{1}{2}. \tag{2.S.3}$$

and by putting all together we will get

$$\boxed{P(a,a) = \mathbb{P}_a(S) = \frac{6+a}{24}}.$$

Finally, since $\mathbb{P}_a(I \cup S \cup D) = 1$, and I, S, D are mutually disjoint, we can write

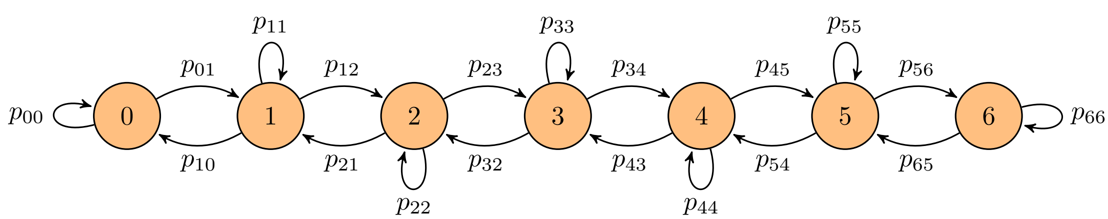
$$\mathbb{P}_a(D) = 1 - (\mathbb{P}_a(I) + \mathbb{P}_a(S)),$$

hence

$$\boxed{P(a,a-1) = \mathbb{P}_a(D) = \frac{a}{12}}.$$

so the transition probabilities are as calculated.

(b) The transition diagram is plotted below.



And the transition matrix is

$$M = \begin{pmatrix} 1/4 & 3/4 & 0 & 0 & 0 & 0 & 0 \\ 1/12 & 7/24 & 5/8 & 0 & 0 & 0 & 0 \\ 0 & 1/6 & 1/3 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/4 & 3/8 & 3/8 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 5/12 & 1/4 & 0 \\ 0 & 0 & 0 & 0 & 5/12 & 11/24 & 1/8 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix}$$

(c) $\mathbb{P}(X_2 = 4|X_0 = 5)$ is the second transition probability $P_2(5,4)$. To compute this, we need to fine the element in the 6-th row and 5-th column in the M^2 matrix, which is basically the inner product between the vectors formed by the 6-th row and the 5-th column.

$$P_2(5,4) = (\frac{5}{12})^2 + \frac{11}{24} \cdot \frac{5}{12} = \frac{35}{96}$$

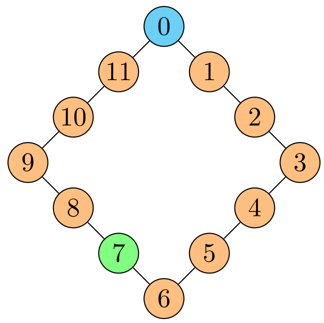
which after simplification becomes

$$\boxed{P_2(5,4) = \frac{35}{96}}.$$

□

■ **Problem 5.3** A clock is broken. It has only one hand which moves every hour either clockwise with probability 1/2 or counter-clockwise with probability 1/2 (the numbers are from 0 to 11 and the hand moves by one full hour when it moves). Assume it starts at 0. What is the probability that it reaches 7 before coming back to 0 for the first time?

Solution First, let's draw the graph representing the state space of the random variable of interest.



Define the event B be $B = \{T_0^+ > T_7\}$. We are interested in finding $\mathbb{P}_0(B)$. Now we can perform the first step argument as follows

$$p_0 = \frac{1}{2}(p_1 + p_{11}). \tag{3.1}$$

Then we analyze each term in the right hand side of the equation above. For p_1 we have

$$\mathbb{P}_1(B) = \underbrace{\mathbb{P}_1(B|T_0 > T_7)}_{1}\underbrace{\mathbb{P}_1(T_0 > T_7)}_{1/5} + \underbrace{\mathbb{P}_1(B|T_0 < T_7)}_{0}\underbrace{\mathbb{P}_1(T_0 < T_7)}_{6/7} = \frac{1}{5}.$$

Note that $\mathbb{P}_1(B|T_0 > T_7) = 1$ since it literally means the random walker reaches 7 before 0. Also $\mathbb{P}_1(B|T_0 < T_7) = 0$ since the event B is conditioned on reaching 0 before 7, which is clearly 0. The term $\mathbb{P}_1(T_0 > T_7)$ is computed using the Gambler's ruin analysis. Similarly, for the p_{11} term we have

$$\mathbb{P}_{11}(B) = \underbrace{\mathbb{P}_{11}(B|T_0 > T_7)}_{1}\underbrace{\mathbb{P}_{11}(T_0 > T_7)}_{1/7} + \underbrace{\mathbb{P}_{11}(B|T_0 < T_7)}_{0}\mathbb{P}_{11}(T_0 < T_7) = \frac{1}{7}.$$

The rationale behind the values of the terms are the same as the ones discussed above. Now we can substitute everything in (3.1)

$$\boxed{p_0 = \frac{1}{2}(\frac{12}{35}) = \frac{6}{35}}.$$

■ **Problem 5.4** The Fibonacci sequence is the sequence $(F_n)_{n\geq 0}$ defined by $F_0 = 0, F_1 = 1$ and

$$F_{n+2} = F_{n+1} + F_n \quad \text{for } n \geq 0.$$

Find a general formula for F_n

Solution First, we construct the characteristic polynomial of the sequence. From the recursive formula we can write

$$X^2 = X + 1 \implies \boxed{X^2 - X - 1 = 0}.$$

The roots of the equation is

$$r_1, r_2 = \frac{1 \pm \sqrt{5}}{2}.$$

Now the general formula will be

$$F_n = Ar_1^n + Br_2^n.$$

To find the coefficients, we utilize the first two terms

$$0 = A + B, \qquad 1 = \frac{1}{2}(A+B) + \frac{\sqrt{5}}{2}(A-B).$$

This system of equations implies that

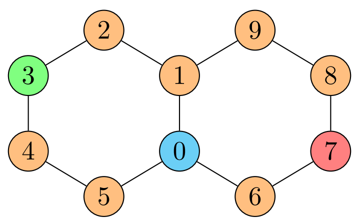
$$A = \frac{1}{\sqrt{5}}, \qquad B = \frac{-1}{\sqrt{5}}.$$

Thus the general formula will be

$$\boxed{F_n = \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right)}.$$

□

■ **Problem 5.5** Let (X_n) be the simple random walk on the following graph. Compute $\mathbb{P}_0(T_3 < T_7)$.



Solution For a much more simpler solution, let's define the two following notations

$$B = \{T_3 < T_7\}, \qquad p_v = \mathbb{P}_v(B).$$

Then, by first step argument at state 0, we can write

$$p_0 = \frac{1}{3}(p_5 + p_6 + p_1). \tag{5.1}$$

Now we need to evaluate each of the terms in the right hand side. We start with p_5.

$$p_5 = \mathbb{P}_5(B) = \underbrace{\mathbb{P}_5(B|T_3 < T_0)}_{1}\underbrace{\mathbb{P}_5(T_3 < T_0)}_{1/3} + \underbrace{\mathbb{P}_5(B|T_3 > T_0)}_{p_0}\underbrace{\mathbb{P}_5(T_3 > T_0)}_{2/3} = \frac{1}{3} + \frac{2}{3}p_0.$$

note that $\mathbb{P}_5(B|T_3 < T_0) = 1$, since if we get to state 3, before getting to state 0, then it means that we have reached the state 3 before reaching the state 7, thus the event B occurs with probability 1. Also $\mathbb{P}_5(T_3 < T_0) = 1/3$ from the Gambler's ruin. Furthermore $\mathbb{P}_5(B|T_3 > T_0) = p_0$ by using the Markov property, and finally $\mathbb{P}_5(T_3 > T_0) = 2/3$ by the Gambler's ruin.
Now, we need to evaluate the term p_6. To analyze this term, we will do a first step argument starting at this point

$$p_6 = \mathbb{P}_6(B) = \frac{1}{2}(\underbrace{p_7}_{0} + p_0) = \frac{p_0}{2}.$$

Note that $p_7 = 0$, since then the event B has not occurred.
Finally, we need to analyze the term p_1. Again, by first step argument on this state we have

$$p_1 = \frac{1}{3}(p_0 + p_9 + p_2).$$

By doing a analysis on p_9 similar to the one we did for 5, we can write

$$p_9 = \mathbb{P}_9(B) = \underbrace{\mathbb{P}_9(B|T_7 < T_1)}_{0}\mathbb{P}_9(T_7 < T_1) + \underbrace{\mathbb{P}_9(B|T_7 > T_1)}_{p_1}\underbrace{\mathbb{P}_9(T_7 > T_1)}_{2/3} = \frac{2}{3}p_1.$$

The rationale behind the values for each term in the equation above, is exactly the same as in analyzing the terms of p_5.
Now, we analyze the term p_2 by performing another first step analysis, similar to the one we did for state 6.

$$p_2 = \frac{1}{2}(\underbrace{p_3}_{1} + p_1) = \frac{1}{2}(1 + p_1).$$

Now we can calculate p_1 in terms of p_0 which turns out to be

$$p_1 = \frac{6}{11}p_0 + \frac{3}{11}.$$

Now we insert all of the terms in the equation (5.1) to get

$$\begin{aligned} 3p_0 &= \frac{1}{3} + \frac{2}{3}p_0 + \frac{1}{2}p_0 + \frac{6}{11}p_0 + \frac{3}{11} \\ \implies 3p_0 - \frac{113}{66}p_0 &= \frac{40}{33} \\ \implies p_0 &= \frac{66}{85} \cdot \frac{40}{33} = \frac{16}{17} \\ \implies & \boxed{p_0 = \frac{16}{17}}. \end{aligned}$$

$\square$

6. Stationary Distributions of Markov Chains

6.1 Time Evolution of Distributions

Let $\{X_n\}_n$ be a discrete Markov with finite state space $S = \{1, 2, \cdots, N\}$. Then, we know that starting at a state $X_0 = 1$, then the probability to be at state j after one step is the $(1, j)$ element of the transition matrix. To be more concrete, let's assume that the Markov chain is defined on $\{1, 2, 3\}$, and assume $X_0 = 1$. Then $P(1, 3) = \mathbb{P}_1(X_1 = 3)$ is the element $(1, 3)$ of the transition matrix. Don't forget that $\{X_n\}$ are all random variables. Thus while we can talk about the cases that what will happen if, for example X_0, be $X_0 = 1$ and etc. We can also talk about the probability that the random variable has specific values, which is the idea of distribution. Let $\mu_{X_0}(i)$ for $i \in \{1, 2, 3\}$ be the distribution of X_0. In other words, we have

$$\mu_{X_0}(i) = \mathbb{P}(X_0 = i) \qquad \text{for } i \in \{1, 2, 3\}.$$

We can also introduce the vector notation for the distribution. Note that we can drop the subscript X_0 as shown in the following notation.

$$\mu_0 = \mu_{X_0} = (\mu_{X_0}(1), \mu_{X_0}(2), \mu_{X_0}(3)).$$

Now, suppose we want to find the distribution of X_1 given the distribution of X_0, i.e. we want to calculate the time evolution of the distribution after one step. Let's calculate what happens for μ_0 after one step:

$$\mu_1 = \mu_{X_1} = (\mu_1(1), \mu_1(2), \mu_1(3)).$$

To calculate $\mu_1(1)$ we can write

$$\mu_1(1) = \mathbb{P}(X_1 = 1) = \underbrace{\mathbb{P}_1(X_1 = 1)}_{P(1,1)} \underbrace{\mathbb{P}(X_0 = 1)}_{\mu_0(1)} + \underbrace{\mathbb{P}_2(X_1 = 1)}_{P(2,1)} \underbrace{\mathbb{P}(X_0 = 2)}_{\mu_0(2)} + \underbrace{\mathbb{P}_3(X_1 = 1)}_{P(3,1)} \underbrace{\mathbb{P}(X_0 = 3)}_{\mu_0(3)}.$$

In summary

$$\mu_1(1) = P(1, 1)\mu_0(1) + P(2, 1)\mu_0(2) + P(3, 1)\mu_0(3).$$

By a similar argument, we can quickly see that

$$\mu_1 = \mu_0 M$$

where M is the transition matrix.

Proposition 6.1 Let $X_0 \sim \mu_0$. Then $X_n \sim \mu_n$ where

$$\mu_n = \mu_0 M^n$$

where M is the transition matrix.

Now we can state the following important observation.

Observation 6.1.1 For a given discrete Markov chain $\{X_n\}$ defined on a *finite* state space $S = \{1, 2, \cdots, N\}$, the sequence of distributions of the random variables at each time step

$$\mu_{X_n} = \mu_n = (\mu_n(1), \mu_n(2), \cdots, \mu_n(N)),$$

defines a discrete time Markov chain with continuous state space $\mathbb{R}^{N-1}$. The transition matrix for $\{\mu_n\}_n$ will be the same as the original Markov chain. The state space will in fact be an affine hyperplane at $\mathbb{R}^N$, that intersects each axis at 1. That is because we require the distributions to sum up to 1. Thus we will have a discrete map

$$\mu_{n+1} = \mu_n M.$$

Observation 6.1.2 — Be careful here!. You need to be careful here and pay special attention to the notations and conventions here. We said that any Markov chain defines another Markov chain $\{\mu_n\}$ which is the distribution of the Markov chain random variable at each step. And also we said that this Markov chain has the same Transition matrix as the original Markov chain. However, you need to note that μ is defined to be a *row* vector

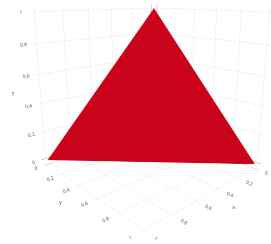
$$\mu_n = (\mu_n(1) \quad \mu_n(2) \quad \cdots \quad \mu_n(N)).$$

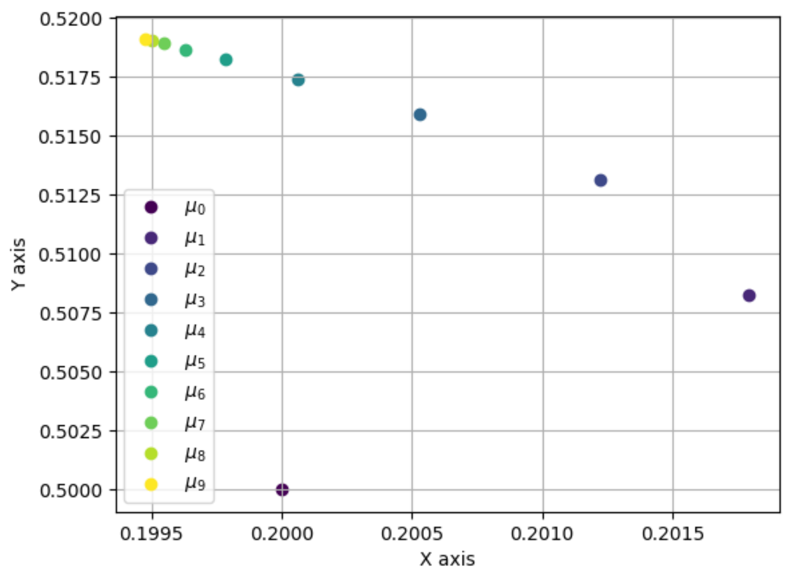
Thus value of μ_{n+1} will be μn multiplied by the transition matrix from *left*. We can develop the whole theory using the notion of transpose, but here we will keep this convention as it is more straight forward.

■ **Example 6.1** Consider the Markov chain $\{X_n\}_n$ defined on the state space $S = \{1, 2, 3\}$, with the following transition probability

$$M = \begin{bmatrix} 0.43 & 0.30 & 0.27 \\ 0.10 & 0.77 & 0.12 \\ 0.22 & 0.20 & 0.58 \end{bmatrix}$$

Now assume that the distribution of X_0 is $\mu_0 = (0.2, 0.5, 0.3)$. The $\{\mu_n\}$ is a Markov chain defined on $\mathbb{R}^3$. To be more precise, since all of the distributions should be normalized, the state space if $\{\mu_n\}$ is in fact the 2D plane that cuts through each axis at 1, as shown in the figure below. This is basically a two dimensional manifold that is embedded in 3 dimensional Euclidean space. We can consider the projection of μ_n on the $x - y$ plane as the 2d atlas (this projection is the diffeomorphism). The following is the time evolution of the distribution with $\mu_0 = (0.2, 0.5, 0.3)$. This example demonstrates that the time evolution of the distribution of a Markov chain is a Markov chain with the same transition matrix. ■





The following is the definition of the stationary distribution for a Markov chain.

Definition 6.1 Let $\{X_n\}$ be a Markov chain defined on the state space S. The distribution vector π is a stationary distribution if we have

$$\pi = \pi P.$$

■ **Remark** Given a distribution, we can do the following test to check if it is a stationary distribution. First, it needs to be a distribution, i.e.

$$\sum_{x \in S} \pi(x) = 1,$$

And secondly, it needs to satisfy the definition for a stationary distribution, i.e. for all $x \in S$ we have

$$\pi(x) = \sum_{y \in S} \pi(y) P(y, x).$$

Observation 6.1.3 A stationary distribution is a left eigenvector for the transition matrix with eigenvalue 1.

Proposition 6.2 Let $\Gamma = (E, V)$ be a connected graph with at least two vertices. Then the stationary distribution for a simple random walk on the graph is given as

$$\pi(x) = \frac{\deg(x)}{2|E|}$$

where $|E|$ is the number of edges of the graph.

■ **Remark** We can show that this is true by two simple checks of the stationary distributions. First, we need to check this is indeed distribution, i.e. sums up to 1.

$$\sum_{x \in S} \pi(x) = \frac{1}{2|E|} \sum_{x \in S} \deg(x) = \frac{2|E|}{2|E|} = 1.$$

Now, we need to check if it is a stationary distribution, i.e. $\forall x \in S$ we have

$$\sum_{y \in S} \pi(y) P(y, x) = \sum_{x \sim y} \frac{\pi(y)}{\deg(y)} = \sum_{x \sim y} \frac{\deg(y)}{2|E|} \frac{1}{\deg(y)} = \frac{\deg(x)}{2|E|}.$$

Observation 6.1.4 — Intuition behind the distribution. Intuitively speaking, the notion of distribution for a Markov chain on a graph, is intuitively speaking similar to the idea of considering the vertices as containers that has some sort of liquid in it, and the transition probability as the rate at which a flow moves between these vertices. Thus at each step, the liquid moves around and the stationary distribution is a distribution where the input and output flow of each vertex is just balanced, that we see no change in the liquid content of each vertex through time.

6.1.1 Uniqueness of the stationary distribution on the irreducible and finite Markov Chains

For this section, we will need the notion of a harmonic function on a graph, as defined below.

Definition 6.2 Let $\{X_n\}$ be a Markov chain defined on the *finite* state space S. Then a function $h : S \to \mathbb{R}$ is harmonic if it satisfies

$$h(x) = \sum_{y \sim x} P(x, y) h(y) \qquad \forall x \in S$$

The notion of a harmonic function on a graph has more intuitive characterization, as follows.

Observation 6.1.5 Let $\Gamma = (V, E)$ be a connected graph. Then a harmonic function defined on Γ is $h : V \to \mathbb{R}$ where $\forall x \in V$ we have

$$h(x) = \sum_{y \sim x} P(x, y) h(y) = \frac{1}{\deg(x)} \sum_{y \sim x} h(y).$$

Thus we can see that a function defined on a graph is harmonic if its value at a particular vertex x is the average of its values at the neighborhood vertices.

Lemma 6.1 Let $\{X_n\}$ be a Markov chain defined on a finite state space S, and denote the transition matrix of this Markov chain as P. Then h is a harmonic function on S if it satisfies

$$Ph = h,$$

i.e. h is the right eigenvector of P with eigenvalue 1.

The following proposition shows a deep connection between the harmonic functions and the first step analysis.

Proposition 6.3 Let $\{X_n\}$ be a Markov chain defined on a finite state space S. Let $W, Z \subset S$ disjoint. Then the following function

$$h(x) = \mathbb{P}_x(T_W < T_Z)$$

is harmonic on $S \backslash (W \cup Z)$.

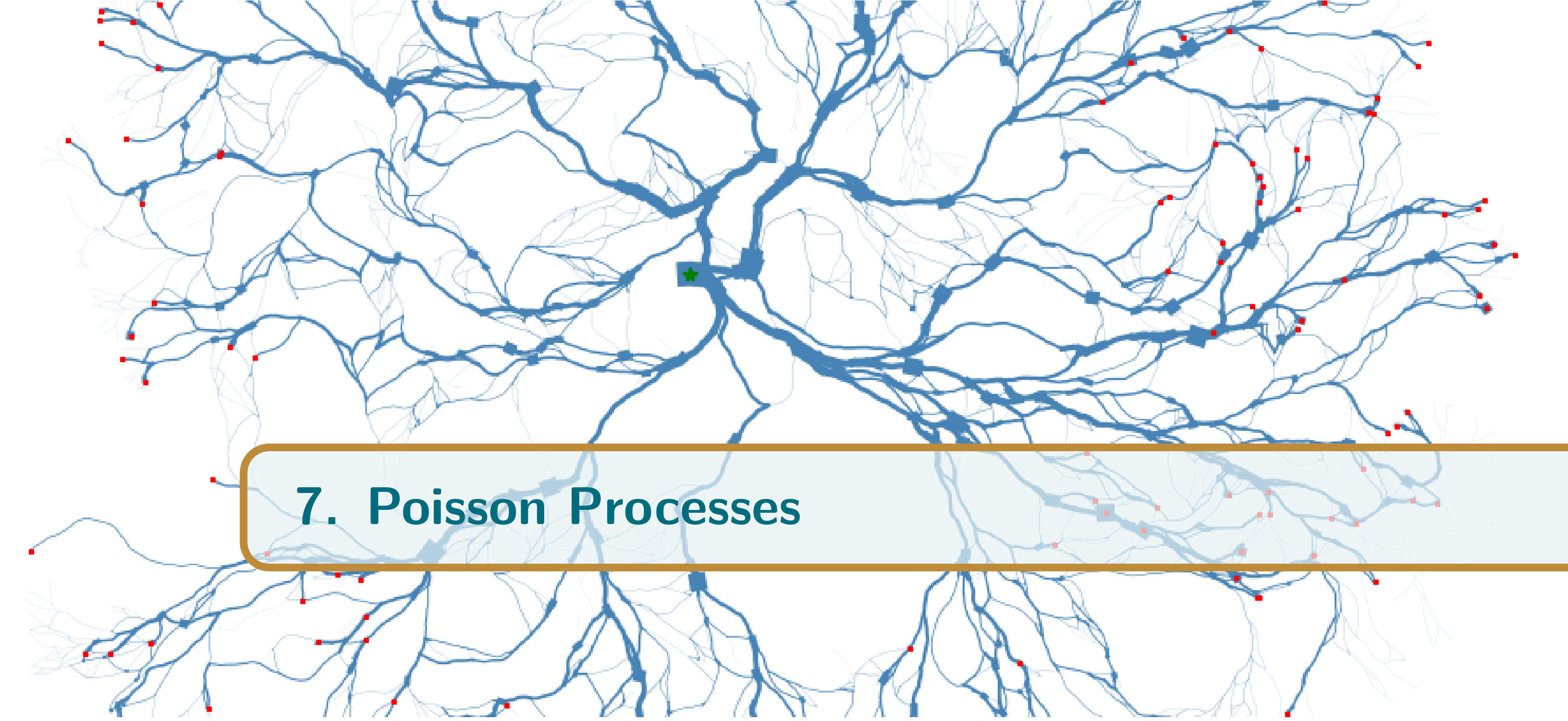
Proof. This immediately follows from the first step argument. Define the event $E = T_W < T_Z$. Then by the first step analysis we have

$$\mathbb{P}_x(E) = \sum_{y \in S} P(x,y)\mathbb{P}_y(E).$$

Thus we can see that the function h satisfies

$$h(x) = \sum_{y \in S} P(x,y)h(y) \qquad \forall x \in S.$$

Thus we conclude that the function h is harmonic. $\square$



7. Poisson Processes

7.1 Solved Problems

■ Problem 7.1 Smith enters a bank with two tellers who just started processing Allen and Yang. Assume that the processing time depends on the special kind of online authentication whose waiting time is a random variable that has a exponential distribution with some constant λ that is constant for all costumers. What is the probability that Smith will leave the bank the last?

Solution Smith should wait until either Allen or Yang is finished. After either is processed (without loss of generality, assume Allen got served and Yang is still waiting), then one of the tellers will give start serving Smith. Let T be a random variable showing the waiting time of Smith and S be a random variable representing the waiting time of Yang, both of which are i.i.d. random variables with exponential distribution. First note that since exponential random variables are memory less, then we have

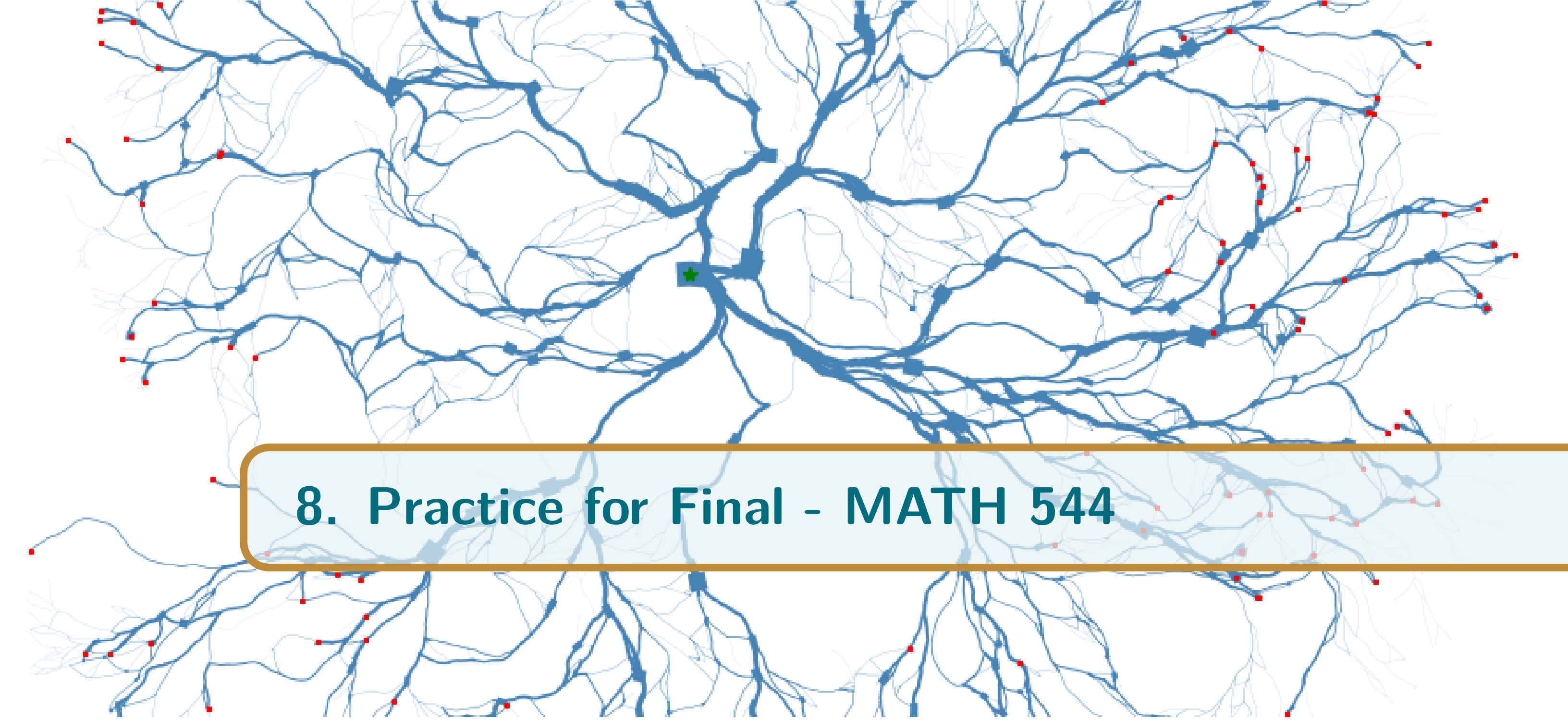
$$\mathbb{P}(S > s + t | S > s) = \mathbb{P}(S > t).$$

We can formulate the Smith leaving the first as

$$\mathbb{P}(T < S).$$

By the law of total probabilities we have

$$\mathbb{P}(T < S) = \int_{-\infty}^{\infty} \mathbb{P}(T < s) f(s) \; ds = \int_{0}^{\infty} (1 - e^{\lambda s}) \lambda e^{-\lambda s} \; ds = 1 - \frac{1}{2} = \frac{1}{2}.$$



8. Practice for Final - MATH 544

These are my solved problems that I was practicing for the final exam.

■ **Problem 8.1** Let $X_1, X_2, \cdots$ be independent $\mathcal{N}(0,1)$ (standard normal) random variables. Prove that

$$\mathbb{P}(\limsup_n \frac{|X_n|}{\sqrt{\log n}} = \sqrt{2}) = 1.$$

You may use the fact that the cumulative distribution function Φ of the standard normal obeys $1 - \Phi(x) \sim \frac{1}{x}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ as $x \to \infty$.

Solution Let $\epsilon \geq 0$. Then as $n \to \infty$ we have

$$\mathbb{P}(\frac{|X_n|}{\sqrt{\log n}} \geq (\epsilon+1)\sqrt{2}) = \mathbb{P}(|X_n| \geq \sqrt{2\log n}(\epsilon+1)) \sim \frac{C}{\sqrt{2\log n}\cdot n^{(\epsilon+1)^2}} \tag{♫}$$

When $\epsilon = 0$ the RHS is not summable. By Borel Cantelli lemma, and using the fact that X_n are independent, we conclude

$$\mathbb{P}(|X_n| \geq \sqrt{2\log n}\ i.o.) = 1.$$

Using that following fact

$$\limsup_n \{\frac{|X_n|}{\sqrt{\log n}} \geq \sqrt{2}\} = \{\limsup_n \frac{|X_n|}{\sqrt{\log n}} \geq \sqrt{2}\},$$

we conclude that

$$\mathbb{P}(\limsup_n \frac{|X_n|}{\sqrt{\log n}} \geq \sqrt{2}) = 1. \tag{♩}$$

Let $\epsilon > 0$. Then the RHS of (♫) is summable. By Borel-Cantelli we get

$$\mathbb{P}(|X_n| \geq (1+\epsilon)\sqrt{2\log n}\ i.o.) = 0.$$

This implies

$$\mathbb{P}(\frac{|X_n|}{\sqrt{\log n}} \leq (1+\epsilon)\sqrt{2}\ a.a.) = 1.$$

Using the fact that $\liminf\{f_n \leq a\} = \{\limsup f_n \leq a\}$ for any sequence of measurable function f_n we can write

$$\mathbb{P}(\frac{|X_n|}{\sqrt{\log n}} \leq (1+\epsilon)\sqrt{2}\ a.a.) = \mathbb{P}(\limsup \frac{|X_n|}{\sqrt{\log n}} \leq (1+\epsilon)\sqrt{2}) = 1.$$

Taking $\epsilon \to 0$ and using (♩) we will get

$$\mathbb{P}(\limsup \frac{|X_n|}{\sqrt{\log n}} = \sqrt{2}) = 1.$$

■ **Problem 8.2** This problem shows that the conclusion of the second Borel-Cantelli Lemma holds under the assumption that the events are pairwise independent only. It is due to Erdos and Renyi 1959. Suppose that the events $A_1, A_2, \cdots$ are pairwise independent, i.e., that $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$. Let $S_n = \sum_{j=1}^n \mathbb{1}_{A_j}$ and $S_\infty = \sum_{j=1}^\infty \mathbb{1}_{A_j}$.

(a) Prove that $\mathrm{Cov}(\mathbb{1}_{A_i}, \mathbb{1}_{A_j}) = \mathbb{P}(A_i \cap A_j) - \mathbb{P}(A_i)\mathbb{P}(A_j)$.

(b) Prove that

$$\mathrm{Var}(S_n) = \sum_{j=1}^n (\mathbb{P}(A_j) - \mathbb{P}(A_j)^2) \le \sum_{j=1}^n \mathbb{P}(A_j).$$

(c) Prove that

$$\mathbb{P}(S_n \ge 1/2 \sum_{j=1}^n \mathbb{P}(A_j)) \le \frac{4}{\sum_{j=1}^n \mathbb{P}(A_j)}$$

(d) Prove that for all $n \in \mathbb{N}$ we have

$$\mathbb{P}(S_\infty \ge 1/2 \sum_{j=1}^n \mathbb{P}(A_j)) \le \frac{4}{\sum_{j=1}^n \mathbb{P}(A_j)}$$

(e) Show that if $\sum_{j=1}^\infty \mathbb{P}(A_j) = \infty$ then $\mathbb{P}(S_\infty < \infty) = 0$ and then conclude that $\mathbb{P}(A_n\ i.o.) = 0$.

Solution (a) This follows immediately from the definition.

$$\begin{aligned}\mathrm{Cov}(\mathbb{1}_{A_j}, \mathbb{1}_{A_j}) &= \mathbb{E}\Big[\mathbb{1}_{A_i}\mathbb{1}_{A_j}\Big] - \mathbb{E}\Big[\mathbb{1}_{A_i}\Big]\mathbb{E}\Big[\mathbb{1}_{A_j}\Big] \\ &= \mathbb{E}\Big[\mathbb{1}_{A_i \cap A_j}\Big] - \mathbb{P}(A_i)\mathbb{P}(A_j) \\ &= \mathbb{P}(A_i \cap A_j) - \mathbb{P}(A_i)\mathbb{P}(A_j).\end{aligned}$$

(b) Starting with the definition of Var we can write

$$\begin{aligned}\mathrm{Var}(S_n) &= \mathbb{E}\Big[S_n^2\Big] - \mathbb{E}\Big[S_n\Big]^2 = \mathbb{E}\Big[(\sum_{j=1}^n \mathbb{1}_{A_j})^2\Big] - \mathbb{E}\Big[\sum_{j=1}^n \mathbb{1}_{A_j}\Big]^2 \\ &= \sum_{i,j=1}^n (\mathbb{E}\Big[\mathbb{1}_{A_i}\mathbb{1}_{A_j}\Big] - \mathbb{E}\Big[\mathbb{1}_{A_i}\Big]\mathbb{E}\Big[\mathbb{1}_{A_j}\Big]) \\ &= \sum_{i,j=1}^n \mathrm{Cov}(\mathbb{1}_{A_i}, \mathbb{1}_{A_j}) \\ &= \sum_{i,j=1}^n (\mathbb{P}(A_i \cap A_j) - \mathbb{P}(A_i)\mathbb{P}(A_j)) \\ &= \sum_{i=1}^n (\mathbb{P}(A_i) - \mathbb{P}(A_i)^2).\end{aligned}$$

(c) First, observe that

$$\mathbb{E}\Big[S_n\Big] = \sum_{j=1}^{n} \mathbb{P}(A_j).$$

So

$$\begin{aligned}
\mathbb{P}(S_n \le 1/2 \sum_{j=1}^{n} \mathbb{P}(A_j)) = \mathbb{P}(-S_n \ge -1/2 \sum_{j=1}^{n} \mathbb{P}(A_j)) &= \mathbb{P}(\mathbb{E}\Big[S_n\Big] - S_n \ge 1/2 \sum_{j=1}^{n} \mathbb{P}(A_j)) \\
&\le \mathbb{P}(|\mathbb{E}\Big[S_n\Big] - S_n| \ge 1/2 \sum_{j=1}^{n} \mathbb{P}(A_j)) \\
&\le \frac{4\operatorname{Var}(S_n)}{(\sum_{j=1}^{n} \mathbb{P}(A_j))^2} \\
&\le \frac{4\sum_{j=1}^{n} \mathbb{P}(A_j)}{(\sum_{j=1}^{n} \mathbb{P}(A_j))^2} \\
&= \frac{4}{\sum_{j=1}^{n} \mathbb{P}(A_j)}
\end{aligned}$$

(d) First, observe that $S_\infty \ge S_n$ for all $n \in \mathbb{N}$. This implies $\{S_\infty \le a\} \subset \{S_n \le a\}$ for all $n \in \mathbb{N}$ and $a \in \mathbb{R}$. So using the monotonicity of probability for any $n \in \mathbb{N}$ we can write

$$\mathbb{P}(S_\infty \le 1/2 \sum_{j=1}^{n} \mathbb{P}(A_j)) \le \mathbb{P}(S_n \le 1/2 \sum_{j=1}^{n} \mathbb{P}(A_j)) \le \frac{4}{\sum_{j=1}^{n} \mathbb{P}(A_j)}.$$

(e) It immediately follows from above that

$$\mathbb{P}(S_\infty < \infty) = 0.$$

Thus we can write $\mathbb{P}(S_\infty = \infty) = 1$. Note that $\{S_\infty = \infty\}$ should be interpreted in the sense that $\omega \in \{S_\infty = \infty\}$ then $S_\infty(\omega) \ge M$ for all $M \in \mathbb{R}$. Observe that from definition $\{S_\infty = \infty\} \subset \{A_i i.o.\}$. Using monotonicity

$$\mathbb{P}(A_n i.o.) = 1.$$

■ **Problem 8.3** Let $X \ge 0$ be a non-negative random variable with $\mathbb{E}\Big[X^2\Big] < \infty$.

(a) Prove that

$$\mathbb{P}(X > 0) \ge \frac{\mathbb{E}\Big[X\Big]^2}{\mathbb{E}\Big[X^2\Big]}.$$

(b) Prove that for $\theta \in [0, 1]$ we have

$$\mathbb{P}(X > \theta\mathbb{E}\Big[X\Big]) \ge (1-\theta)^2 \frac{\mathbb{E}\Big[X\Big]^2}{\mathbb{E}\Big[X^2\Big]}.$$

Solution (a) We can write

$$\mathbb{E}\Big[X\Big] = \mathbb{E}\Big[X\mathbb{1}_{X>0}\Big] + \mathbb{E}\Big[X\mathbb{1}_{X=0}\Big] \le \mathbb{E}\Big[X^2\Big]^{1/2}\mathbb{P}(X > 0)^{1/2},$$

where we have used the Cauchy-Schwartz inequality. This implies

$$\mathbb{P}(X > 0) \geq \frac{\mathbb{E}\Big[X\Big]^2}{\mathbb{E}\Big[X^2\Big]}.$$

(b) Let $\mu = \mathbb{E}\Big[X\Big]$ for easier notation. Using similar idea as above we can write

$$\begin{aligned}\mu = \mathbb{E}\Big[X\Big] &= \mathbb{E}\Big[X\mathbb{1}_{X\leq\theta\mu}\Big] + \mathbb{E}\Big[X\mathbb{1}_{X>\theta\mu}\Big] &\quad (8.0.1)\\ &= \theta\mu + \mathbb{E}\Big[X^2\Big]^{1/2}\mathbb{P}(X > \theta\mu)^{1/2}, &\quad (8.0.2)\end{aligned}$$

where we have used the fact that $\mu\theta \geq \mathbb{1}_{X<\mu\theta}$ for the first term and Cauchy-Schwartz inequality for the second term. This implies

$$\mathbb{P}(X > \theta\mathbb{E}\Big[X\Big]) \geq (1-\theta)^2\frac{\mathbb{E}\Big[X\Big]^2}{\mathbb{E}\Big[X^2\Big]}.$$

■ **Problem 8.4** Let X be a random variable such that $\mathbb{E}\Big[|X|\Big] < \infty$. Prove that

$$\mathbb{E}\Big[|X|\mathbb{1}_{|X|\geq n}\Big] \to 0 \quad \text{as } n \to \infty.$$

Solution Observe that for all $\omega \in \Omega$ as $n \to \infty$ we have

$$|X(\omega)|\mathbb{1}_{|X(\omega)|\geq n} \to 0,$$

because $X(\omega)$ is a real number and is eventually less that n as $n \to \infty$. Further, observe that for all $n \in \mathbb{N}$

$$|X|\mathbb{1}_{|X|\geq n} \leq |X|.$$

Since $\mathbb{E}\Big[|X|\Big] < \infty$, by dominated convergence theorem we have

$$\mathbb{E}\Big[|X|\mathbb{1}_{|X|\geq n}\Big] \to \mathbb{E}\Big[0\Big] = 0 \quad \text{as } n \to \infty.$$

■ **Problem 8.5** Let $X_1, X_2, \cdots$ be i.i.d. with $\mathbb{E}\Big[X^+\Big] = \infty$ and $\mathbb{E}\Big[X^-\Big] < \infty$. Let $S_n = X_1 + \cdots + X_n$. Prove that $S_n/n \to \infty$ a.s. as $n \to \infty$.

Solution We will consider the truncated random variables $X_{i,N} = \min\{X_i, N\}$. First, observe that $X_{i,N}$ has finite mean because

$$\mathbb{E}\Big[|X_{i,N}|\Big] = \mathbb{E}\Big[X_{i,N}^+\Big] + \mathbb{E}\Big[X_{i,N}^-\Big] < \infty$$

where we have used the fact that $X_{i,N}^- = X_i^-$ thus $\mathbb{E}\Big[X_{i,N}^-\Big] < \infty$ and $X_{i,N}^+ \leq N$, thus $\mathbb{E}\Big[X_{i,N}^+\Big] < N$. By applying the SLLN to $X_{i,N}$ we see that

$$S_{n,N}/n \to \mathbb{E}\Big[X_{1,N}\Big] \quad \text{as } n \to \infty.$$

Since $S_{n,N} \leq S_n$ for all $n, N \in \mathbb{N}$ we have

$$\liminf S_n/n \geq \liminf S_{n,N}/n = \mathbb{E}\Big[X_{1,N}\Big].$$

On the other hand observe that $X_{i,N}^{+} \uparrow X_i^{+}$ and $X_{i,N}^{-} = X_i^{-}$. Using Monotone Convergence theorem with the first statement we see $\mathbb{E}\Big[X_{i,N}^{+}\Big] \to \infty$ as $N \to \infty$ and for the second statement we have $\mathbb{E}\Big[X_{i,N}^{-}\Big] \to \mathbb{E}\Big[X_i^{-}\Big] < \infty$. This implies

$$\mathbb{E}\Big[X_{1,N}\Big] \to \infty \quad \text{as } n \to \infty.$$

So we will have

$$\liminf S_n/n \geq \infty.$$

■ **Problem 8.6** Let X and Y be random variables with finite mean. Suppose that for any $A \in \mathcal{F}$ we have $\mathbb{E}\Big[X\mathbb{1}_A\Big] \leq \mathbb{E}\Big[Y\mathbb{1}_A\Big]$. Prove that $X \leq Y$ a.s.

Proof. First, observe that for all $A \in \mathcal{F}$ we have

$$\mathbb{E}\Big[(Y - X)\mathbb{1}_A\Big] \geq 0.$$

We want to show $\mathbb{P}(X > Y) = 0$. Let $A_n = \{X - Y > 1/n\}$. Then $\{X > Y\} = \bigcup_n A_n$. From sub additivity we have

$$\mathbb{P}(X > Y) \leq \sum_n \mathbb{P}(A_n).$$

On the other hand, observe that

$$0 \leq \mathbb{E}\Big[(Y - X)\mathbb{1}_{A_n}\Big] \leq \frac{-1}{n}\mathbb{P}(A_n),$$

which implies $\mathbb{P}(A_n) = 0$ for all $n \in \mathbb{N}$. Thus we conclude that

$$\mathbb{P}(X > Y) = 0 \Leftrightarrow \mathbb{P}(X \leq Y) = 1.$$

□

Summary 8.1 — A very important note. The theme of the proof above is the proof by contradiction (or contrapositive. I don't know!). To show $\mathbb{P}(X \leq Y) = 1$ we instead prove $\mathbb{P}(X > Y) = 0$. This theme shows up a lot in probability. For instance, to show that a sequence of random variables $\{X_n\}$ converges almost surely to some other random variable X, we instead usually show that the set of points where this does not happen is small (has measure zero). In other words we show $\mathbb{P}(\limsup_n\{|X_n - X| \geq \epsilon\}) = 0$.

■ **Problem 8.7** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the uniform distribution on $\Omega = \{1, 2, 3\}$. Find random variables X, Y, and Z on this space such that

$$\mathbb{P}(X > Y)\mathbb{P}(Y > Z)\mathbb{P}(Z > X) > 0,$$

and

$$\mathbb{E}\Big[X\Big] = \mathbb{E}\Big[Y\Big] = \mathbb{E}\Big[Z\Big].$$

Solution Define

$$X(\omega) = \begin{cases} 1 & \omega \in \{1\}, \\ 0 & \omega \in \{2, 3\}. \end{cases} \quad , \quad Y(\omega) = \begin{cases} 1 & \omega \in \{2\}, \\ 0 & \omega \in \{1, 3\}. \end{cases} \quad , \quad Z(\omega) = \begin{cases} 1 & \omega \in \{3\}, \\ 0 & \omega \in \{1, 2\}. \end{cases} \quad .$$

Then

$$\mathbb{E}\Big[X\Big] = \mathbb{E}\Big[Y\Big] = \mathbb{E}\Big[Z\Big] = 1/3,$$

and

$$\mathbb{P}(X > Y) = \mathbb{P}(Y > Z) = \mathbb{P}(Z > X) = 1/3.$$

■ **Problem 8.8** Suppose $\mathbb{P}(Z=0)=\mathbb{P}(Z=1)=1/2$, that $Y\sim\mathcal{N}(0,1)$, and that Y and Z are independent. Set $X=YZ$. What is the law of X?

Solution Let $A\in\mathcal{B}$. Then

$$\begin{aligned}\mu_X(A)=\mathbb{P}(X\in A)&=\mathbb{P}(X\in A|Z=1)\mathbb{P}(Z=1)+\mathbb{P}(X\in A|Z=0)\mathbb{P}(Z=0)\\&=1/2\mathbb{P}(Y\in A)+1/2\mathbb{P}(0\in A).\end{aligned}$$

So we can write
$$\mu_X=\frac{1}{2}\mu_Y+\frac{1}{2}\delta_0.$$

■ **Problem 8.9** Prove Cantelli's inequality, which states that if X is a random variable with finite mean m and finite variance v, then for $\alpha>0$,

$$\mathbb{P}(X-m\geq\alpha)\leq\frac{v}{v+\alpha^2}.$$

Hint: First show $\mathbb{P}(X-m\geq\alpha)\leq\mathbb{P}((X-m+y)^2\geq(\alpha+y)^2)$ for all $y>0$. Then use Markov's inequality, and minimize the resulting bound over choice of $y>0$.

Solution Observe that $\{X-m\geq\alpha\}\subseteq\{(X-m+y)^2\geq(\alpha+y)^2\}$. By monotonicity of probability it follows that $\mathbb{P}(X-m\geq\alpha)\leq\mathbb{P}((X-m+y)^2\geq(\alpha+y)^2)$. Using the Markov's inequality we will have

$$\mathbb{P}(X-m\geq\alpha)\leq\mathbb{P}((X-m+y)^2\geq(\alpha+y)^2)\leq\frac{\mathbb{E}\Big[(X-m+y)^2\Big]}{(\alpha+y)^2}=\frac{\mathbb{E}\Big[(X-m)^2\Big]+y^2}{(\alpha+y)^2}.$$

By setting the derivative of RHS equal to zero we will get $y=v/\alpha$. Substituting this value in the RHS we will get
$$\mathbb{P}(X-m\geq\alpha)\leq\frac{v}{v+\alpha^2}.$$

■ **Problem 8.10** For $p\geq1$ let $\|X\|_p=\mathbb{E}\Big[|X|^p\Big]^{1/p}$. Let $\|X\|_\infty=\inf\{M:\mathbb{P}(|X|>M)=0\}$.

(a) Prove that $\|XY\|_1\leq\|X\|_1\|Y\|_\infty$.

(b) Prove that $\|X\|_\infty=\lim_{p\to\infty}\|X\|_p$.

Solution (a) First, observe that we have $|Y|\leq\|Y\|_\infty$ a.s. This implies $|XY|\leq|X|\|Y\|_\infty$ a.s. So we will have $\mathbb{E}\Big[|XY|\Big]\leq\mathbb{E}\Big[|X|\Big]\|Y\|_\infty$. In other words

$$\|XY\|_1\leq\|X\|_1\|Y\|_\infty.$$

(b) First, observe that since $|X|\leq\|X\|_\infty$ *a.s.*, we will have $\|X\|_p\leq\|X\|_\infty$ for all $p\geq1$. So we can write
$$\limsup_{p\to\infty}\leq\|X\|. \qquad (\text{♩})$$

On the other hand, because $\|X\|_\infty$ is the essential suprimum of X, for any $\epsilon>0$, if we define $A_\epsilon=\{|X|>\|X\|_\infty-\epsilon\}$ we will have $\mathbb{P}(A_\epsilon)>0$. So we can write

$$\|X\|_p=\mathbb{E}\Big[|X|^p\Big]^{1/p}=\Big(\mathbb{E}\Big[|X|^p\mathbb{1}_{A_\epsilon}\Big]+\mathbb{E}\Big[|X|^p\mathbb{1}_{A_\epsilon^c}\Big]\Big)^{1/p}\geq\Big(\mathbb{E}\Big[|X|^p\mathbb{1}_{A_\epsilon}\Big]\Big)^{1/p}\geq(\|X\|_\infty-\epsilon)\mathbb{P}(A_\epsilon)^{1/p}.$$

This implies that

$$\liminf_{p\to\infty} \|X\|_p \geq (\|X\|_\infty - \epsilon) \quad \forall \epsilon > 0.$$

Since this is true for all $\epsilon > 0$, we can conclude $\liminf_{p\to\infty} \|X\|_p \geq \|X\|_\infty$. Combining with (♩) and using the fact that lim inf of any sequence of always less than or equal to lim sup that sequence, we conclude that

$$\|X\|_\infty = \limsup_{p\to\infty} \|X\|_p = \liminf_{p\to\infty} \|X\|_p = \lim_{p\to\infty} \|X\|_p.$$

■ **Problem 8.11** Let $X_0 = (1, 0) \in \mathbb{R}^2$ and define $X_n \in \mathbb{R}^2$ inductively by declaring that X_{n+1} is chosen at random from the ball of radius $|X_n|$ centered at the origin, i.e., $X_{n+1}/|X_n|$ is uniformly distributed on the ball of radius 1 and independent of $X_1, \cdots, X_n$. Prove that $n^{-1} \log |X_n| \to c$ a.s. and compute c.

Solution Let $X_{n+1}/|X_n| = U_{n+1}$, where U_{n+1} is uniformly distributed on the unit ball. So we can write $|X_n| = |U_n \cdots U_1||X_0|$. Taking log from both sides we will have

$$\log |X_n| = \sum_{i=1}^{n} \log(U_n).$$

Using the SLLN's we have

$$n^{-1} \log |X_n| = n^{-1} \sum_{i=1}^{n} \log(U_n) \to \mathbb{E}\Big[\log(U_1)\Big].$$

Note that we have used the fact that $\log(U_n)$ are i.i.d. (because X_n and U_n are) and has finite mean (we did not check the latter statement here). Furthermore observe that $\mathbb{P}(U_n \leq r) = r^2$, thus U_n has density $2r$. So

$$\mathbb{E}\Big[\log(U_1)\Big] = \int_0^1 2r \log r dr = \frac{-1}{2}.$$

Thus we conclude

$$n^{-1} \log X_n \to \frac{-1}{2}.$$

■ **Problem 8.12** Use Jensen's inequality to prove that $\|X\|_p \leq \|X\|_q$ for $1 \leq p \leq q \leq \infty$.

Solution Let $\varphi(x) = |x|^{q/p}$ that is a convex function when $q \geq p$. Observe that

$$\varphi(\mathbb{E}\Big[|X|^p\Big]) \leq \mathbb{E}\Big[\varphi(|X|^p)\Big] = \mathbb{E}\Big[|X|^q\Big].$$

By using the fact that $\varphi(\mathbb{E}\Big[|X|^p\Big]) = \mathbb{E}\Big[|X|^p\Big]^{q/p}$ and raising the both side to power $1/q$ we will get

$$(\mathbb{E}\Big[|X|^p\Big])^{1/p} \leq (\mathbb{E}\Big[|X|^q\Big])^{1/q}.$$

■ **Remark** Another way to prove the inequality above is to use Hölder's inequality.

■ **Problem 8.13** Let $X, X_1, X_2, \cdots$ be random variables on the same probability space. Let $p \in [1, \infty)$. We say that $X_n \to X$ in L^p if $\lim_{n\to\infty} \|X_n - X\|_p = 0$.

(a) Show that if $X_n \to X$ in L^p then $X_n \to X$ in probability.

(b) Give a counterexample to the converse of (a).

(c) Let $1 \le p \le q < \infty$. Show that $\|Y\|_p \le \|Y\|_q$ for any random variable Y and hence if $X_n \to X$ in L^q then $X_n \to X$ i L^p.

Solution (a) We can prove this very easily with Markov's inequality

$$\mathbb{P}(|X_n - X| \ge \epsilon) = \mathbb{P}(|X_n - X|^p \ge \epsilon^p) \le \frac{\mathbb{E}\Big[|X_n - X|^p\Big]}{\epsilon^p} \to 0 \quad \text{as } n \to \infty.$$

(a') In one of my attempts to solve this problem, my brain followed the following pathway!

$$\begin{aligned}\mathbb{E}\Big[|X_n - X|^p\Big] &= \mathbb{E}\Big[|X_n - X|^p \mathbb{1}_{|X_n - X| \ge \epsilon}\Big] + \mathbb{E}\Big[|X_n - X|^p \mathbb{1}_{|X_n - X| < \epsilon}\Big] \\ &\ge \mathbb{E}\Big[|X_n - X|^p \mathbb{1}_{|X_n - X| \ge \epsilon}\Big] \\ &\ge \epsilon \mathbb{P}(|X_n - X| \ge \epsilon).\end{aligned}$$

Since $\mathbb{E}\Big[|X_n - X|^p\Big] \to 0$ as $n \to \infty$, it follows that $\mathbb{P}(|X_n - X| \ge \epsilon) \to 0$ as $n \to \infty$ and for any choice of $\epsilon > 0$.

(b) Take $X_n : (0,1) \to \mathbb{R}$ and $X_n = \mathbb{1}_{[0,1/n]}$. Then $X_n \to 0$ in probability but $1 = \mathbb{E}\Big[X_n\Big] \not\to \mathbb{E}\Big[X\Big] = 0$.

(c) Let $\varphi(x) = |x|^{q/p}$ that is a convex function when $q \ge p$. Observe that

$$\varphi(\mathbb{E}\Big[|X|^p\Big]) \le \mathbb{E}\Big[\varphi(|X|^p)\Big] = \mathbb{E}\Big[|X|^q\Big].$$

By using the fact that $\varphi(\mathbb{E}\Big[|X|^p\Big]) = \mathbb{E}\Big[|X|^p\Big]^{q/p}$ and raising the both side to power $1/q$ we will get

$$(\mathbb{E}\Big[|X|^p\Big])^{1/p} \le (\mathbb{E}\Big[|X|^q\Big])^{1/q}.$$

■ **Problem 8.14** Suppose $\mathbb{E}\Big[|X|\Big] < \infty$. Prove that, for any $\epsilon > 0$, there exists a $\delta > 0$ such that $\mathbb{E}\Big[|X|\mathbb{1}_A\Big] < \epsilon$ for all $A \in \mathcal{F}$ with $\mathbb{P}(A) < \delta$.

Solution First, we want to show if $\mathbb{E}\Big[|X|\Big] < \infty$ then $\mathbb{E}\Big[|X|\mathbb{1}_{|X| \ge n}\Big] \to 0$ as $n \to \infty$. To see why, observe that

$$|X|\mathbb{1}_{|X| \ge n} \to 0 \qquad \text{as } n \to \infty.$$

Furthermore, observe that

$$|X|\mathbb{1}_{|X| \ge n} \le |X|.$$

Since $\mathbb{E}\Big[|X|\Big] < \infty$, using the Dominated Convergence Theorem we can conclude that $\mathbb{E}\Big[|X|\mathbb{1}_{|X| \ge n}\Big] \to 0$ as $n \to \infty$.

Let $\epsilon > 0$ given. Choose N large enough such that $\mathbb{E}\Big[X\mathbb{1}_{|X| \ge n}\Big] < \epsilon/2$ for all $n \ge N$. Choose $\delta = \epsilon/(2N)$. Let $A \in \mathcal{A}$ such that $\mathbb{P}(A) < \delta$. Then we can write

$$\begin{aligned}\mathbb{E}\Big[|X|\mathbb{1}_A\Big] &= \mathbb{E}\Big[|X|\mathbb{1}_{A \cap \{|X| \ge N\}}\Big] + \mathbb{E}\Big[|X|\mathbb{1}_{A \cap \{|X| < N\}}\Big] \\ &\le \frac{\epsilon}{2} + \frac{N\epsilon}{2N} = \epsilon.\end{aligned}$$

■ **Problem 8.15** Calculate the variance and mean of Poisson random variable. *Hint: Probability generating function will be useful!*

Solution Recall that for $X \sim \text{Poisson}(\lambda)$ we have

$$\mathbb{P}(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}.$$

So we can calculate the probability generating function

$$G_X(z) = \mathbb{E}\Big[z^X\Big] = \sum_{k=0}^{\infty} z^k \mathbb{P}(X = k) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda z)^k}{k!} = e^{\lambda(z-1)},$$

where we have used the fact the the series above is absolutely convergent and converges to $e^{\lambda z}$. Using the fact that $G_X^{(n)}(z) = \lambda^n e^{\lambda(z-1)}$.

$$\mathbb{E}\Big[X\Big] = G'_X(1) = \lambda, \qquad \mathbb{E}\Big[X(X-1)\Big] = G''_X(1) = \lambda.$$

Using the fact that $\text{Var}(X) = \mathbb{E}\Big[X^2\Big] - \mathbb{E}\Big[X\Big]^2$ we will get

$$\text{Var}\, X = \lambda.$$

■ **Problem 8.16** Let $\{X_n\}, X$ be random variables with law μ_n, μ. Suppose $\mu_n \Rightarrow \delta_c$ for some $c \in \mathbb{R}$. Prove that $\{X_n\}$ converges to c in probability.

Solution For a given $\epsilon > 0$ consider

$$\mathbb{P}(|X_n - c| \geq \epsilon) = \mathbb{P}(X_n \leq c-\epsilon) + \mathbb{P}(X_n \geq c+\epsilon) = 1 + \mathbb{P}(X_n \leq c-\epsilon) - \mathbb{P}(X_n < c+\epsilon) = 1 - \mu_n((c-\epsilon, c+\epsilon)).$$

Observe that since $\mu(\partial(c-\epsilon, c+\epsilon)) = 0$ and $\mu_n \Rightarrow \mu$ then $\mu_n((c-\epsilon, c+\epsilon)) \to \mu((c-\epsilon, c+\epsilon)) = 1$, hence for any choice of ϵ we have

$$\mathbb{P}(|X_n - X| \geq \epsilon) \to 0 \quad \text{as } n \to \infty.$$

■ **Problem 8.17** Let $0 < M < \infty$, and let $f, f_1, f_2, \cdots : [0,1] \to [0, M]$ be Borel-measurable functions with $\int_0^1 f\, d\lambda = \int_0^1 f_n\, d\lambda = 1$. Suppose $\lim_n f_n(x) = f(x)$ for each fixed $x \in [0,1]$. Define probability measures $\mu, \mu_1, \mu_2, \cdots$ by $\mu(A) = \int_A f\, d\lambda$, and $\mu_n(A) = \int_A f_n\, d\lambda$, for Borel $A \subset [0,1]$. Prove that $\mu_n \Rightarrow \mu$.

Solution Let $g : [0,1] \to \mathbb{R}$ be any bounded continuous function. Then by change of variable formula we have

$$\int g d\mu_n = \int_0^1 g f_n d\lambda, \qquad \int g d\mu = \int_0^1 g f d\lambda.$$

Since $f_n \to f$ pointwise, then $gf_n \to gf$ pointwise as well. Observe that $gf_n \leq gM$ and since g is bounded and continuous then $\int gMd\lambda < \infty$. Thus by monotone convergence theorem we have

$$\int g f_n d\lambda \to \int g f d\lambda \quad \text{as } n \to \infty.$$

In other words $\int g d\mu_n \to \int g d\mu$ as $n \to \infty$. This proves that $\mu_n \Rightarrow \mu$.

■ **Problem 8.18** Let $\varphi(t)$ be the characteristic function of the random variable X. Suppose that $|\varphi(t_0)| = 1$ for some $t_0 \neq 0$. Prove there exists $a, b \in \mathbb{R}$ such that $\mathbb{P}(X \in a + b\mathbb{Z}) = 1$.

Solution Since $|\varphi(t_0)| = 1$, it is on the unit circle in complex plane. I.e. $\exists a \in \mathbb{R}$ such that $\varphi(t_0) = e^{iat_0}$. On the other hand, from definition we have

$$\varphi(t_0) = \int e^{it_0 X} d\mathbb{P}.$$

This implies $1 = \int e^{it_0(X-a)} d\mathbb{P}$. So $\int (1 - e^{it_0(X-a)}) d\mathbb{P} = 0$. This implies that

$$1 - e^{it_0(X-a)} = 0 \ a.s.$$

Taking the real part we will get $\cos\big(t_0(X-a)\big) = 1 a.s.$, hence $t_0(X-a) \in 2\pi\mathbb{Z}$ almost surely. This implies $X \in a\mathbb{Z} + b$ almost surely, for some $a, b \in \mathbb{R}$.

■ **Problem 8.19** Recall that for a Geometric distribution with parameter p we have $\mathbb{P}(X = k) = (1-p)^{k-1}p$. Consider the following variation. X_n has p.m.f given as $\mathbb{P}(X_n = k/n) = (1-\lambda/n)^{k-1}(\lambda/n)$ for $k \in \mathbb{N}$, with $\lambda > 0$. Apply the Continuity theorem to prove that X_n converges weakly to an $\text{Exp}(\lambda)$ random variable.

Solution First, we need to calculate the characteristic function of X_n.

$$\varphi_{X_n}(\xi) = \mathbb{E}\Big[e^{i\xi X}\Big] = \frac{\lambda}{n}\sum_{k=1}^{\infty} e^{i\xi k/n}(1-\lambda/n)^{k-1} = \frac{\lambda}{n} \cdot \frac{e^{i\xi/n}}{1 - e^{i\xi/n}(1-\lambda/n)},$$

where we have used the geometric progression sum formula. The denominator is a asymptotically

$$1 - e^{i\xi/n}(1-\lambda/n) = 1 - (1 + i\xi/n + O(1/n^2))(1-\lambda/n) = \frac{\lambda - i\xi}{n} + O(1/n^2), \quad \text{as } n \to \infty.$$

So we will have

$$\varphi_{X_n}(\xi) \to \frac{\lambda}{\lambda - i\xi} \quad \text{as } n \to \infty,$$

which is the characteristic function of an exponential random variable.

■ **Problem 8.20 — Convergence in distribution to uniform distribution.** Let μ be Lebesgue measure on $[0,1]$, and let μ_n be defined by $\mu_n(i/n) = 1/n$ for $i = 1, 2, \cdots, n$. Show to $\mu_n \Rightarrow \mu$ as $n \to \infty$.

Solution First, observe that μ is the distribution of a uniform random variable, call it U. The c.d.f of this random variable is $F_U(x) = x$ on $[0,1]$, 0 on $(-\infty, 0)$ and 1 on $(1, \infty)$. We want to show that c.d.f of μ_n converges to μ on (0,1) (i.e. all continuity points of $F_U(x)$). Observe that

$$\mu_n((-\infty, x]) = \frac{\lfloor nx \rfloor}{n}$$

Then for $x \in (0,1)$ we have

$$|\mu((-\infty, x]) - \mu_n((-\infty, x])| = |x - \frac{\lfloor nx \rfloor}{n}| \le 1/n.$$

So $\mu_n((-\infty, x]) \to \mu((-\infty, x])$ as $n \to \infty$.

■ **Problem 8.21 — Convergence in distribution to uniform distribution.** Let $Y_1, Y_2, \cdots$ be i.i.d. uniform with values in $\{0, 1, 2, \cdots, 9\}$. Define

$$X_n = \sum_{i=1}^{n} \frac{Y_i}{10^i}.$$

Show that $X_n \Rightarrow U$ where U is a random variable with uniform distribution.

Solution First, we calculate the characteristic function of Y_1:

$$\varphi_{Y_1}(\xi) = \mathbb{E}\Big[e^{i\xi Y}\Big] = \frac{1}{10}\sum_{k=0}^{9} e^{i\xi k} = \frac{1}{10}\cdot\frac{1-e^{10i\xi}}{1-e^{i\xi}}.$$

So for X_n we will have

$$\varphi_{X_n}(\xi) = \varphi_{Y_1}(\xi/10)\varphi(Y_1)(\xi/100)\cdots\varphi_{Y_1}(\xi/10^n) = \frac{1}{10^n}\frac{1-e^{i\xi}}{1-e^{i\xi/10^n}} \sim \frac{1-e^{i\xi}}{10^n(-i\xi/10^n)} \to \frac{e^{i\xi}-1}{i\xi},$$

as $n\to\infty$, which is the characteristic function of a uniform random variable on $[0,1]$.

■ **Problem 8.22** Using the definition of the convergence in distribution, prove that if $\mu_n \Rightarrow \mu$ then $\{\mu_n\}$ is tight.

Solution Let $\epsilon > 0$ be given. Choose $a, b \in \mathbb{R}$ such that $\mu[a,b] > 1-\epsilon$, and a, b are not atoms of μ. So by definition $\mu_n[a,b] \to \mu[a,b]$. I.e. $\exists N \in \mathbb{N}$ such that for all $n > N$ we have $\mu_n[a,b] > 1-\epsilon$. Enlarge $[a,b]$ to $[a_n,b_n]$ such that for all $n \le N$ we have $\mu_n[a_n,b_n] \ge 1-\epsilon$. Let $[A,B] = \bigcup_{n=1}^{N}[a_n,b_n]$. So by construction $\mu_n[A,B] \ge 1-\epsilon$ for all $n\in\mathbb{N}$. So the collection $\{\mu_n\}$ is tight.

■ **Problem 8.23** Using the fact that if $\{\mu_n\}$ is tight, then $\{\varphi_n\}$ is equicontinuous, prove that if $\mu_n \Rightarrow \mu$ then $\varphi_n \to \varphi$ uniformly on compact sets. Show that the convergence should not be uniform on all real line.

Solution We use the result of Problem 8.22. Since $\mu_n \Rightarrow \mu$, $\{\mu_n\}$ is tight. Also, since $\{\mu_n\}$ is tight, $\{\varphi_n\}$ is equicontinuous. We know that $\mu_n \Rightarrow \mu$ implies $\varphi_n(\xi)\to\varphi(\xi)$ pointwise. Pointwise convergence of equicontinuous family of functions is uniform convergence on compact sets. This proves the first part. For the second part, consider $\mu_n = \mathcal{N}(0,1/n^2)$. Then $\varphi_n(\xi) = e^{-\xi^2 n^2/2} \to 1$ pointwise on $\mathbb{R}$. However, this convergence is not uniform all all$\mathbb{R}$ as $\varphi_n(\xi)$ decays to zero as $\xi\to\infty$ or $\xi\to-\infty$.

■ **Problem 8.24** Using the identity $\sin t = 2\sin\big(t/2\big)\cos\big(t/2\big)$ repeatedly leads to $\sin(t)/t = \prod_{m=1}^{\infty}\cos\big(t/2^m\big)$. Prove the last identity by interpreting each side as a characteristic function.

■ **Problem 8.25** Observer that by applying the trigonometric identity for 3 times we will get

$$\sin(t) = \frac{1}{2^3}\sin\Big(\frac{t}{8}\Big)\cos\Big(\frac{t}{8}\Big)\cos\Big(\frac{t}{4}\Big)\cos\Big(\frac{t}{2}\Big).$$

For any n large we can write

$$\sin(t) = \frac{1}{2^n}\sin\Big(\frac{t}{2^n}\Big)\prod_{i=1}^{n}\cos\Big(\frac{t}{2^n}\Big) \sim t\prod_{i=1}^{n}\cos\Big(\frac{t}{2^n}\Big).$$

So we can write

$$\frac{\sin t}{t} = \prod_{n=1}^{\infty}\cos\Big(\frac{t}{2^n}\Big).$$

The RHS can be thought as the characteristic function of $Y_1 + Y_2 + \cdots$ where $Y_i = B_i/2^i$ where B_i is Bernoulli random variable (with values in $\{-1,1\}$ and parameter $p = 1/2$). The LHS can be thought as the characteristic function of uniform distribution on $[-1,1]$. So this identity shows that

$$\sum_{i=1}^{n}\frac{B_i}{2^i} \Rightarrow \mu_U \quad \text{as } n\to\infty.$$

where U is the uniform distribution on $[-1,1]$.

■ **Remark** Notice the similarity between the problem above and Problem 8.21.

■ **Problem 8.26** Show that if φ is a characteristic function, then $\mathrm{Re}(\varphi)$ and $|\varphi|^2$ are also characteristic functions. Also, show that if $\varphi_1, \varphi_2, \cdots, \varphi_n$ are characteristic function, then any convex linear combination is also a characteristic function.

Solution (a) For $\mathrm{Re}(\varphi)$: Let X be a random variable with characteristic function φ. Let Y be a random variable such that $\mathbb{P}(Y=1)=\mathbb{P}(Y=0)=1/2$ (i.e. Y is a Bernoulli random variable). Let Z be a random variable that is X when $Y=1$ and X' when $Y=0$, where X' is an independent copy of $-X$ (hence $\varphi_{X'}=\overline{\varphi_X}$). So the c.d.f of Z will be

$$F_Z(x)=\mathbb{P}(Z\le x)=\frac{1}{2}\mathbb{P}(X\le x)+\frac{1}{2}\mathbb{P}(X'\le x)=\frac{1}{2}F_X(x)+\frac{1}{2}F_{X'}(x).$$

And the law of Z will be

$$\mu_Z=\frac{1}{2}\mu_X+\frac{1}{2}\mu_{X'}.$$

So the characteristic function of Z will be

$$\varphi_Z(\xi)=\frac{1}{2}\int e^{i\xi x}\mu_X(dx)+\frac{1}{2}\int e^{i\xi x}\mu_{X'}(dx)=\frac{\varphi_X(\xi)+\overline{\varphi_X(\xi)}}{2}=\mathrm{Re}(\varphi_X(\xi)).$$

(b) For $|\varphi|^2$: Let X' be an independent copy of $-X$ (hence $\varphi_{X'}=\overline{\varphi_X}$). Define the random variable $Z=X+X'$. Then the characteristic function of Z is

$$\varphi_Z(\xi)=\varphi_X(\xi)\overline{\varphi_X(\xi)}=|\varphi_X(\xi)|^2.$$

(c) For $\sum_{i=1}^n \lambda_i\varphi_i$ (i.e. convex linear combination): This generalizes the item (a) above. Let Y be a random variable with values in $\{1,2,\cdots,n\}$ such that $\mathbb{P}(Y=i)=\lambda_i$. Define Z to be a random variable that is equal to X_i (that has characteristic function φ_i) when $Y=i$. Then the law of Z will be

$$\mu_Z=\sum_{i=1}^n \lambda_i\varphi_i.$$

So the characteristic function of Z will be

$$\varphi_Z=\sum_{i=1}^n \lambda_i\varphi_i.$$

■ **Problem 8.27** Show that if $\lim_{t\to 0}(\varphi(t)-1)/t^2=c>-\infty$, then $\mathbb{E}\Big[X\Big]=0$ and $\mathbb{E}\Big[X^2\Big]=-2c<\infty$. In particular, if $\varphi(t)=1+o(t^2)$, then $\varphi(t)\equiv 1$.

Solution First observe that

$$\lim_{t\to 0}\overline{\frac{\varphi(t)-1}{t^2}}=\lim_{t\to 0}\frac{\overline{\varphi(t)}-1}{t^2}=\lim_{t\to 0}\frac{\varphi(-t)-1}{t^2}=c.$$

So by adding these two limits and noting that $\varphi(0)=1$ we will have

$$\lim_{t\to 0}\frac{\varphi(t)+\varphi(-t)-2}{t^2}=2c.$$

For this part of the reasoning I will be using a weak arguments, however you can see 3.3.21 Durrent for a more precise argument. The limit about "looks like" the second derivative of φ. So we conclude that $\mathbb{E}\Big[X^2\Big] < \infty$. So we can have the following expansion for φ

$$\varphi(\xi) = 1 + i\xi\mathbb{E}\Big[X\Big] - \xi^2/2\mathbb{E}\Big[X^2\Big] + o(t^2).$$

Using this expansion with $\lim_{t\to 0}(\varphi(t)-1)/t^2 = c$ implies that

$$\mathbb{E}\Big[X\Big] = 0, \qquad \mathbb{E}\Big[X^2\Big] = -2c < \infty.$$

In particular, if we know $\varphi(\xi) = 1 + o(\xi)$, then

$$\lim_{t\to 0} \frac{\varphi(t)-1}{t^2} = \lim_{t\to 0} \frac{o(t^2)}{t^2} = 0.$$

So $\mathbb{E}\Big[X\Big] = 0$, and $\mathbb{E}\Big[|X|^2\Big] = 0$. Thus $|X| = 0$ almost surely. This implies that $X \equiv 0$ almost surely. So $\varphi(t) = 1$.

■ **Problem 8.28** If Y_n are random variables with characteristic functions φ_n, then $Y_n \Rightarrow 0$ if and only if there is a $\delta > 0$ such that $\varphi_n(t) \to 1$ for $|t| \le \delta$.

Solution Using Lemma 11.1.13 we see that $\{\mu_n\}$ is tight. So there is a convergent subsequence $\{\mu_{n_k}\}$ that $\mu_{n_k} \Rightarrow \nu$ as $k \to \infty$. So $\varphi_{n_k} \to \varphi$. Thus every subsequence of $\{\mu_n\}$ converges to ν. Since $\varphi_n \to 1$ on $[-\delta, \delta]$, so $\varphi = 1$ on $[-\delta, \delta]$. This implies that

$$\varphi(t) = 1 + o(t^2) \quad \text{as } t \to 0.$$

(To see this observe that $\lim_{t\to 0}(1-\varphi(t)) = 0$ as $\varphi = 1$ on $[-\delta, \delta]$). From the problem above we see that $\varphi(t) \equiv 1$ on. This implies that $X = 0$ almost surely, thus $Y_n \Rightarrow X$. This completes the proof.

■ **Remark** Most of the work above is done to show that $\varphi_n \to \varphi$, where φ is some characteristic function with $\varphi(\xi) = 1$ on $[-\delta, \delta]$. If we know that φ_n converges pointwise to some characteristic function that is equal to 1 on some closed neighborhood of 0, then we can immediately write $\varphi(\xi) = 1 + o(\xi^2)$ and conclude that $\varphi(\xi) \equiv 1$.

■ **Problem 8.29** Let $X_1, X_2, \cdots$ be independent. If $S_n = \sum_{m\le n} X_m$ converges in distribution then it converges in probability. *Hint: The last exercise implies that if $m, n \to \infty$, then $S_m - S_n \to 0$ in probability.*

Solution First observe that $\varphi_{S_n}(\xi) = (\varphi_X(\xi))^n$. Because φ_X is a characteristic function (hence uniformly continuous), then $\exists \delta > 0$ such that $\varphi_X(\xi) > 0$ when $\xi \in [-\delta, \delta]$. Let $m, n \in \mathbb{N}$ with $n > m$. Then

$$(S_n - S_m)(\xi) = \sum_{l=m}^{n} X_l = (\varphi_X(\xi))^{m-n} = \frac{\varphi_X(\xi)^m}{\varphi_X(\xi)^n}.$$

Let $\xi \in [-\delta, \delta]$. Then

$$\lim_{m,n\to\infty}(S_n(\xi) - S_m(\xi)) = \lim_{m,n\to\infty} \frac{\varphi_X(\xi)^m}{\varphi_X(\xi)^n} = \frac{\lim_{m\to\infty} \varphi_X(\xi)^m}{\lim_{n\to\infty} \varphi_X(\xi)^n} = 1.$$

So $(S_m - S_n)(\xi) \to 1$ on $[\delta, \delta]$. This implies that $S_m - S_n \Rightarrow 0$ as $m, n \to \infty$. Since the limiting measure is constant, then the convergence in distribution can be upgraded to converges in probability. So $S_m - S_n \to 0$ as $m, n \to \infty$.

■ **Problem 8.30** The distribution of a random variable X is called *infinitely divisible* if, for all $n \in \mathbb{N}$, there exists a sequence of independent and identically distributed random variables $Y_1^{(n)}, Y_2^{(n)}, \ldots, Y_n^{(n)}$ such that X and $Y_1^{(n)} + Y_2^{(n)} + \cdots + Y_n^{(n)}$ have the same distribution. Prove that the characteristic function φ of an infinitely divisible distribution is nonzero for all real t, i.e., that $\varphi(t) \neq 0$ for all $t \in \mathbb{R}$.

Hint: Let φ_n be the characteristic function of $Y_i^{(n)}$. Let $\psi_n(t) = |\varphi_n(t)|^2$ and $\psi(t) = |\varphi(t)|^2$. Explain why these are also characteristic functions. Study the limit of $\psi_n(t)$.

Solution Let $\varphi_n(\xi)$ denote the characteristic function of $Y_1^{(n)}$ and $\varphi(\xi)$ denote the characteristic function of X. Because X is infinitely divisible

$$\varphi(\xi) = (\varphi_n(\xi))^n \qquad \forall n \in \mathbb{N}.$$

Let $\psi(\xi) = |\varphi(\xi)|^2$ and $\psi_n(\xi) = |\varphi_n(\xi)|^2$. Then we can write

$$\psi_n(\xi) = |\varphi(\xi)^{1/n}|^2 = (|\varphi(\xi)|^2)^{1/n} = \psi_n(\xi)^{1/n}.$$

Let $g(\xi) = \lim_{n\to\infty} \psi_n(\xi)$. Then if $\psi_n(\xi) > 0$ then $g(\xi) = 1$ and if $\psi_n(\xi) = 0$ then $g(\xi) = 0$. Since ψ_n is a characteristic function then $\psi_n(0) = 1$ and exists $\delta > 0$ such that $\psi_n(\xi) > 0$ on $[-\delta, \delta]$. This implies that $g(\xi) = 1$ on $[-\delta, \delta]$. We can show (see the remark below) that g is a characteristic function and on $[-\delta, \delta]$ we can write $g(\xi) = 1 + o(\xi^2)$. This implies that $g(\xi) \equiv 1$.

■ **Remark** Since $\psi_n \to g$ and g is continuous then the collection of corresponding distributions is dense. Thus there exists a subsequence of distributions that converge to some distribution. Thus the corresponding subsequence of the characteristic functions should also converge to the corresponding characteristic function of that distribution which is g. So g is a characteristic function of some distribution.

■ **Problem 8.31 — Application of Central Limit Theorem.** An astronomer is interested in measuring, in light years, the distance from her observatory to a distant star. Although she has a measuring technique, she knows that because of changing atmospheric conditions and experimental error, each time a measurement is made it will not yield the exact value but rather an approximate value. As a result the astronomer plans to make a series of measurements and then use the average of these as the estimated value of the actual distance. She believes that the values of the measurement errors are not systematic, so that the measurements are described by a random variable with mean d (the true distance) and a variance of 4 (light years)^2. Use the central limit theorem to determine approximately the number of measurements that should be made to be 95% sure that the estimated distance is accurate to within ± 0.5 light years.

Solution Let S_n/n denote the average measured value. Then we want to have

$$\mathbb{P}(|S_n/n - d| \leq 1/2) = 0.95.$$

To find the value of n we can write

$$0.95 = \mathbb{P}(|S_n/n - d| \leq 1/2) = \mathbb{P}(|\frac{S_n - nd}{\sqrt{n\sigma^2}}| \leq \frac{\sqrt{n}}{2\sigma}) \approx \mathbb{P}(|Z| \leq \frac{\sqrt{n}}{2\sigma}) = 1 - 2\Phi(\frac{\sqrt{n}}{2\sigma}),$$

where Z has standard normal distribution. Using the standard normal distribution table we get $\sqrt{n}/(2\sigma) = 1.96$. This implies

$$n \approx 62.$$

■ **Problem 8.32** Let $X_1, X_2, \ldots$ be i.i.d. random variables with known mean m and unknown variance σ^2, both finite and with $\sigma^2 \neq 0$. Prove that

$$\frac{S_n - nm}{\sqrt{\sum_{k=1}^{n}(X_k - m)^2}} \Rightarrow N(0,1).$$

To avoid $\frac{0}{0}$ issues, assume that $P(X_k = m) = 0$.

Solution Replace $X_1, X_2, \cdots$ with $Y_1 = X_1 - m, Y_2 = X_2 - m, \cdots$. So $\mathbb{E}\left[Y_i\right] = 0$. Let $S_n = Y_1 + Y_2 + \cdots + Y_n$. Thus we want to show

$$\frac{S_n}{\sum_{i=1}^{n} Y_i^2} \Rightarrow \mathcal{N}(0,1).$$

First observe that by CLT

$$\frac{S_n}{\sqrt{n\sigma^2}} \Rightarrow \mathcal{N}(0,1).$$

Also, observe that by weak law of large numbers

$$\frac{\sum_{i=1}^{n} Y_i^2}{n} \to \mathbb{E}\left[Y_i^2\right] = \sigma^2 \quad \text{in probability.}$$

So $\frac{\sum_{i=1}^{n} Y_i^2}{\sigma^2 n} \to 1$ in probability. The multiplication of these two will also converge in distribution to $1 \cdot \mathcal{N}(0,1)$. So

$$\frac{S_n}{\sum_{i=1}^{n} Y_i^2} \Rightarrow \mathcal{N}(0,1)$$

■ **Problem 8.33** Let $X_1, X_2, \ldots$ be i.i.d. random variables with c.d.f. F. The *empirical c.d.f.* $\hat{F}_n : \mathbb{R} \to [0,1]$ is defined by

$$\hat{F}_n(x) = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{\{X_k \leq x\}}.$$

Thus, for each $x \in \mathbb{R}$, $\hat{F}_n(x)$ is a random variable.

(a) For each $x \in \mathbb{R}$, prove that $\hat{F}_n(x) \to F(x)$ a.s.

(b) If x satisfies $0 < F(x) < 1$, prove that

$$\frac{\sqrt{n}\left(\hat{F}_n(x) - F(x)\right)}{\sqrt{F(x)[1-F(x)]}} \Rightarrow N(0,1).$$

Solution (a) For the first part we use the SLLN. First observe that since X_i are i.i.d. then $\mathbb{1}_{X_i \leq x}$ are also i.i.d. Furthermore, observe that the average of these random variables is finite. Because

$$\mathbb{E}\left[\mathbb{1}_{X_i \leq x}\right] = \mathbb{P}(\{X_i \leq x\}) = F_i(x).$$

So using the strong law of large numbers for i.i.d. random variables with finite mean we will have

$$\hat{F}_n(x) \to \mathbb{E}\left[\mathbb{1}_{X_i \leq x}\right] = F_1(x)\ a.s.$$

(b) For the second part, notice that

$$\mathbb{E}\Big[\mathbb{1}^2_{X_i \le x}\Big] = F_i(x).$$

So the variance of $\mathbb{1}_{X_i \le x}$ will be

$$\mathrm{Var}(\mathbb{1}_{X_i \le x}) = F(x) - F(x)^2 = F(x)(1 - F(x)).$$

So using the central limit theorem we have

$$\frac{n\hat{F}_n(x) - nF(x)}{\sqrt{nF(x)(1-F(x))}} = \frac{\sqrt{n}\left(\hat{F}_n(x) - F(x)\right)}{\sqrt{F(x)[1-F(x)]}} \Rightarrow N(0,1)$$

■ **Problem 8.34** This problem concerns the method of Monte Carlo integration, which is a method for the approximate evaluation of an integral $I = \int_0^1 f(x)\,dx$.

(a) Let $U_1, \dots, U_N$ be i.i.d. uniform random variables on the interval $(0,1)$, and let

$$I_N = \frac{1}{N}\left[f(U_1) + \cdots + f(U_N)\right].$$

Suppose that $\int_0^1 f(x)^2\,dx < \infty$, and let $\sigma^2 = \mathrm{Var} f(U_1) = \int_0^1 f(x)^2\,dx - I^2$. Apply the central limit theorem to show that I_N converges to I as $N \to \infty$, in the sense that

$$P\left(|I_N - I| \le \frac{\sigma x}{\sqrt{N}}\right) \to P(|Z| \le x),$$

where Z is a standard normal random variable.

(b) Assuming that $\sigma \le 1$, how large should N be taken to be 95% confident that I_N is within 0.01 of I?

Solution First notice that $f(U_i)$ has finite mean I and finite variance $\mathrm{Var}(f(U_1)) = \int_0^1 f(x)^2\ dx - I_2 = \sigma^2$. So by central limit theorem we have

$$\mathbb{P}(\frac{NI_n - NI}{\sqrt{N\sigma^2}} \le x) \to \mathbb{P}(Z \le x) = 1 - 2\Phi(x).$$

So using the fact that

$$\mathbb{P}(|\frac{NI_n - NI}{\sqrt{N\sigma^2}}| \le x) = \mathbb{P}(|I_n - I| \le \frac{x\sigma}{\sqrt{N}})$$

we have

$$\mathbb{P}(|I_n - I| \le \frac{x\sigma}{\sqrt{N}}) \to 1 - \Phi(x).$$

By change of variable $y = \frac{x\sigma}{\sqrt{N}}$ we can write

$$\mathbb{P}(|I_n - I| \le y) \approx 1 - 2\Phi(\frac{\sqrt{N}y}{\sigma}).$$

We have $y = 0.01$ and we assume $\sigma = 1$. Then we need to have

$$1 - 2\Phi(\sqrt{N}/100) = 0.95.$$

This implies $\Phi(\sqrt{N}/100) = 0.025$. From the standard normal distribution we read $\sqrt{N}/100 = 1.96$. Thus we need to have

$$N \approx (196)^2.$$

■ **Problem 8.35** Let $X_1, X_2, \ldots$ be i.i.d. with $\mathbb{E}\Big[X_i\Big] = 0$ and $0 < \mathrm{Var}(X_i) < \infty$, and let $S_n = X_1 + \cdots + X_n$.Use the central limit theorem and Kolomogrov's zero-one law to conclude that $\limsup S_n/\sqrt{n} = \infty$ almost surely.

Solution Observe that

$$\limsup_n \mathbb{P}(\frac{S_n}{\sqrt{n}} > M) = \limsup_n \mathbb{P}(\frac{S_n}{\sigma\sqrt{n}} > \frac{M}{\sigma}) = \mathbb{P}(Z > M/\sigma) > 0.$$

Also observe that

$$\mathbb{P}(\frac{S_n}{\sqrt{n}} > M \ i.o.) \geq \limsup_n \mathbb{P}(\frac{S_n}{\sqrt{n}} > M) > 0.$$

Since $\{\frac{S_n}{\sqrt{n}} > M \ i.o.\}$ is a tail event and its probability is positive then it should be 1. Thus

$$\mathbb{P}(\frac{S_n}{\sqrt{n}} > M \ i.o.) = 1.$$

So

$$\limsup_n \frac{S_n}{\sqrt{n}} = \infty \ a.s.$$

■ **Problem 8.36** Let $X_1, X_2, \ldots$ be i.i.d. with $X_i \geq 0$ and $\mathbb{E}\Big[X_i\Big] = 1$, and $\mathrm{Var}(X_i) = \sigma^2 \in (0, \infty)$. Show that $2(\sqrt{S_n} - \sqrt{n}) \Rightarrow \sigma\chi$.

Solution By CLT we have

$$\frac{S_n - n}{\sqrt{n}} \Rightarrow \sigma\chi.$$

We can write the LHS as

$$\frac{S_n - n}{\sqrt{n}} = \frac{(\sqrt{S_n} - \sqrt{n})(\sqrt{S_n} + \sqrt{n})}{\sqrt{n}} = 2(\sqrt{S_n} - \sqrt{n}) \cdot \frac{(\sqrt{S_n} + \sqrt{n})}{2\sqrt{n}} = 2(\sqrt{S_n} - \sqrt{n}) \cdot (\sqrt{S_n/n} + 1)/2.$$

By WLLN we have

$$S_n/n \to \mathbb{E}\Big[X_1\Big] = 1 \quad \text{in probability.}$$

So by continuous mapping theorem (Durrent) we have

$$(\sqrt{S_n/n} + 1)/2 \to 1 \quad \text{in probability.}$$

By converging together theorem (Durrent) we have

$$2(\sqrt{S_n} - \sqrt{n}) \to \sigma\chi.$$