Part IA — Probability

Based on lectures by Dr D. Yeo

Lent 2022

Contents

§0 Introduction

Text in blue is usually less important.

Example 0.1

Dice: outcomes $1, 2, \ldots, 6$.

- $\mathbb{P}(2) = \frac{1}{6}$.
- $\mathbb{P}(\text{multiple of } 3) = \frac{2}{6} = \frac{1}{3}$.
- $\mathbb{P}(\text{not a multiple of } 3) = \frac{2}{3}$
- $\mathbb{P}(\text{prime}) = \frac{1}{2}$.
-

$$\begin{aligned}
\mathbb{P}(\text{prime or multiple of } 3) &= \cancel{\frac{1}{3} + \frac{1}{2} = \frac{5}{6}}. \\
&= \frac{4}{6} = \frac{2}{3}. \\
\mathbb{P}(\text{prime or multiple of } 3) &= \frac{1}{3} + \frac{1}{2} - \frac{1}{6} = \frac{2}{3}
\end{aligned}$$

§1 Formal Setup

Definition 1.1 (Sample Space)
The **sample space** Ω is a set of outcomes.

Definition 1.2 (σ-algebra)

- Let $\mathcal{F}$ a collection of subsets of Ω (called *events*).
- $\mathcal{F}$ is a **σ-alegbra** if

  F1. $\Omega \in \mathcal{F}$.

  F2. $A \in \mathcal{F}$ then $A^c = \Omega \setminus A \in \mathcal{F}$.

  F3. $\forall$ countable collections $(A_n)_{n\in\mathbb{N}} \in \mathcal{F}$, the union $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}$ also.

Remark 1. The motivation for F2 is so that $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ (the probability of not A is defined as expected).

Definition 1.3 (Probility Measure)
Given σ-algebra $\mathcal{F}$ on Ω, function $\mathbb{P} : \mathcal{F} \to [0, 1]$[a] is a **probability measure** if

P2. $\mathbb{P}(\Omega) = 1$.

P3. $\forall$ countable collections $(A_n)_{n\in\mathbb{N}}$ of *disjoint* events in $\mathcal{F}$:

$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mathbb{P}(A_n).$$

Then $(\Omega, \mathcal{F}, \mathbb{P})$ is a *probability space*.

[a] P1. $\mathbb{P}(A) \geq 0$

Example 1.1
Coming back to Example 0.1. $\Omega = \{1, 2, \ldots, 6\}$ so
$\mathbb{P}(\Omega) = \mathbb{P}(1 \text{ or } 2 \text{ or } 3 \text{ or } 4 \text{ or } 5 \text{ or } 6) = 1$ and $\mathcal{F}$ is all subsets of Ω.

Question

$$\text{Why } \mathbb{P} : \mathcal{F} \to [0, 1]$$

and not $\mathbb{P} : \Omega \to [0,1]$?

If Ω is countable:

- In general: $\mathcal{F} =$ all subsets of Ω, i.e. $\mathcal{P}(\Omega)$ (the power set).
- $\mathbb{P}(2)$ is shorthand for $\mathbb{P}(\{2\})$.
- $\mathbb{P}$ is determined by $(\mathbb{P}(\{w\}),\ \forall\, w \in \Omega)$ (e.g. unfair dice).

If Ω is uncountable:

- E.g. $\Omega = [0,1]$. Want to choose a real number, all equally likely.
- If $\mathbb{P}(\{0\}) = \alpha > 0$ then $\mathbb{P}\left(\left\{0, 1, \frac{1}{2}, \dots, \frac{1}{n}\right\}\right) = (n+1)\alpha$ ↯ if n large as $\mathbb{P} > 1$.
- So $\mathbb{P}(\{0\}) = 0$, or $\mathbb{P}(\{0\})$ is undefined.
- What about $\mathbb{P}\left(\left\{x : x \leq \frac{1}{3}\right\}\right)$?
  - ? "Add up" all $\mathbb{P}(\{x\})$ for $x \leq \frac{1}{3}$. However this range is uncountable and we can't take a sum of uncountably many terms.

Aside

Question

Can we choose uniformly from an infinite countable set? (E.g. $\Omega = \mathbb{N}$ or $\Omega = \mathbb{Q} \cap [0,1]$)

Answer

No it is not possible but that's ok there $\exists$ lots of interesting probability measures of $\mathbb{N}$!

Proof. Suppose possible

- $\mathbb{P}(\{0\}) = \alpha > 0 \quad \forall\, \omega \in \Omega$. Then $\mathbb{P}(\Omega) = \sum_{\omega\in\Omega} \mathbb{P}(\{\omega\}) = \sum_{\omega\in\Omega} \alpha = \infty$. ↯ of $\mathrm{P2} : \mathbb{P}(\Omega) = 1$.
- $\mathbb{P}(\{0\}) = 0 \quad \forall\, \omega \in \Omega$. Then $\mathbb{P}(\Omega) = \sum_{\omega\in\Omega} \mathbb{P}(\{\omega\}) = \sum_{\omega\in\Omega} 0 = 0$.

$\square$

Proposition 1.1 (From the axioms)

- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$

> *Proof.* A, A^c are disjoint. $A \cup A^c = \Omega$.
> $\implies \mathbb{P}(A) + \mathbb{P}(A^c) \underset{P3}{=} \mathbb{P}(\Omega) \underset{P2}{=} 1$ $\square$

- $\mathbb{P}(\emptyset) = 0$
- If $A \subseteq B$ then $\mathbb{P}(A) \leq \mathbb{P}(B)$
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

§1.1 Examples of Probability Spaces

Example 1.2 (Uniform Choice)

Ω finite, $\Omega = \{\omega_1, \ldots, \omega_n\}$, $\mathcal{F} =$ all subsets. *uniform* choice (equally likely)

$$\mathbb{P} : \mathcal{F} \to [0,1], \; \mathbb{P}(A) = \frac{|A|}{|\Omega|}.$$

In particular: $\mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|} \quad \forall\, \omega \in \Omega$.

Example 1.3 (Choosing without replacement)

n indistinguishable marbles labelled $\{1, \ldots, n\}$. Pick $k \leq n$ marbles uniformly at random. Here: $\Omega = \{A \subseteq \{1, \ldots, n\}, |A| = k\} \quad |\Omega| = \binom{n}{k}$

Example 1.4 (Well-shuffled deck of cards)

Uniformly chosen *permutation* of 52 cards.

$$\begin{aligned} \Omega &= \{\text{all permutations of 52 cards}\} \\ |\Omega| &= 52! \\ \mathbb{P}\left(\substack{\text{first three cards} \\ \text{have the same suit}}\right) &= \frac{52 \times 12 \times 11 \times 49!}{52!} = \frac{22}{425} \\ \text{Note: } &= \frac{12}{51} \times \frac{11}{50} \end{aligned}$$

Example 1.5 (Coincident Birthdays)

There are n people; what is the probability that at least two of them share a birthday?

Assumptions:

- No leap years! (365 days)
- All birthdays are equally likely

Let $\Omega = \{1, \ldots, 365\}^n$ and $\mathcal{F} = \mathcal{P}(\Omega)$.
Let $A = \{\text{at least two people share the same birthday}\}$ and so
$A^c = \{\text{all } n \text{ birthdays are different}\}$.

$$\begin{aligned} \mathbb{P}(A^c) &= \frac{|A^c|}{|\Omega|} = \frac{365 \times 364 \cdots \times (365 - n + 1)}{365^n} \\ \mathbb{P}(A) &= 1 - \mathbb{P}(A^c) \end{aligned}$$

Note that at $n = 22$, $\mathbb{P}(A) \approx 0.476$ and at $n = 23$, $\mathbb{P}(A) \approx 0.507$. So when there are at least 23 people in a room, the probability that two of them share a birthday is around 50%.

KEY IDEA: Calculating $\mathbb{P}(A^c)$ is easier than $\mathbb{P}(A)$.

§2 Combinatorial Analysis

§2.1 Subsets

Question

Let Ω be finite and $|\Omega| = n$. How many ways to *partition* Ω into k disjoint subsets $\Omega_1, \dots, \Omega_k$ with $|\Omega_i| = n_i$ (with $\sum_{i=1}^{k} n_i = n$)?

Answer

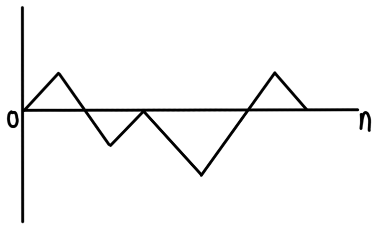
$$
\begin{aligned}
M &= \underbrace{\binom{n}{n_1}}_{\substack{\text{Choose}\\ \text{first part}}} \underbrace{\binom{n-n_1}{n_2}}_{\substack{\text{Then choose}\\ \text{second part}}} \binom{n-n_1-n_2}{n_3} \dots \underbrace{\binom{n-(n_1+\dots+n_{k-1})}{n_k}}_{=1} \\
&= \frac{n!}{n_1!\cancel{(n-n_1)!}} \times \frac{\cancel{(n-n_1)!}}{n_2!\cancel{(n-n_1-n_2)!}} \times \frac{\cancel{[n-(n+n_1+\dots+n_{k-1})]!}}{0!n_k!} \\
&= \frac{n!}{n_1!n_2!\dots n_k!} \\
&= \underbrace{\binom{n}{n_1, n_2, \dots, n_k}}_{\text{Multinomial coefficient}}
\end{aligned}
$$

Key sanity check

- Does ordering of the subsets matter?

E.g. Is $\Omega_2 = \{3,4,7\}, \Omega_3 = \{1,5,8\}$ equal to $\Omega_3 = \{3,4,7\}, \Omega_2 = \{1,5,8\}$? No, ordering does matter as we put elements first in the second subset then the third.

§2.2 Random Walks



$$\Omega = \{(X_0, X_1, \dots, X_n) : X_0 = 0, |X_n - X_{k-1}| = 1 \ \forall\ k = 1, \dots, n\}.$$

$$|\Omega| = 2^n \text{ (we can go either up or down at each } k\text{)}$$

$$\mathbb{P}(X_n = n) = \frac{1}{2^n}$$
$$\mathbb{P}(X_n = 0) = 0 \text{ if } n \text{ is odd}$$
$$\text{What about } \mathbb{P}(X_n = 0) \text{ when } n \text{ is even}$$

Idea - Choose $\frac{n}{2}$ ks for $X_k = X_{k-1} + 1$ and the rest $X_k = X_{k-1} - 1$ (i.e. go up half the time and down the other half).

$$\begin{aligned}\mathbb{P}(X_n = 0) &= 2^{-n}\binom{n}{\frac{n}{2}} \\ &= \frac{n!}{2^n \left(\frac{n}{2}!\right)^2}\end{aligned}$$

Question
What happens when n is large?

§2.3 Stirling's Formula

Notation. Let $(a_n), (b_n)$ be two sequences. Say $a_n \sim b_n$ as $n \to \infty$ if $\frac{a_n}{b_n} \to 1$ as $n \to \infty$.

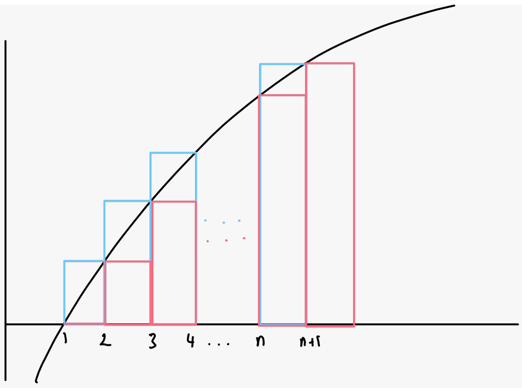
Example 2.1
$n^2 + 5n + \frac{6}{n} \sim n^2$

Example 2.2 (Non-Example)
$\exp\left(n^2 + 5n + \frac{6}{n}\right) \nsim \exp(n^2)$

Theorem 2.1 (Stirling)
$n! \sim \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n}$ as $n \to \infty$.

Theorem 2.2 (Weaker Version)
$\log n! \sim n \log n$.

Proof. $\log(n!) = \log 2 + \cdots + \log n$.



$$\overbrace{\int_1^n \log x\,dx}^{\text{“Upper Integral”}} \le \log n! \le \overbrace{\int_1^{n+1} \log x\,dx}^{\text{“Lower Integral”}}$$

$$\underbrace{n\log n - n + 1}_{\sim n\log n} \le \log n! \le \underbrace{(n+1)\log(n+1) - n}_{\sim n\log n}$$

Key idea: *Sandwiching* between lower/upper integrals. It was useful that

- $\log x$ is increasing
- $\log x$ has a nice integral!

$\square$

§2.4 (Ordered) compositions

Definition 2.1 (Composition)

A **composition** of m with k parts is a sequence $(m_1, \dots, m_k)$ of non-negative integers with $m_1 + \cdots + m_k = m$.

Example 2.3

$$3 + 0 + 1 + 2 = 6 \neq 1 + 2 + 0 + 3 = 6$$

$$\star\star\star\,||\,\star\,|\,\star\star$$

There is a bijection between compositions *and* sequences of m stars and $(k-1)$ dividers. So the number of compositions is $\binom{m+k-1}{m}$.

Comment: Easy to mistake k with $k-1$ in no. of dividers.

§3 Properties of Probability measures

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathbb{P} : \mathcal{F} \to [0, 1]$.

Definition 3.1 (Countable additivity)
P3 : $\mathbb{P}(\bigcup_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} \mathbb{P}(A_n)$ for $(A_n)_{n\in\mathbb{N}}$ disjoint.

Question
What if the sets are not disjoint?

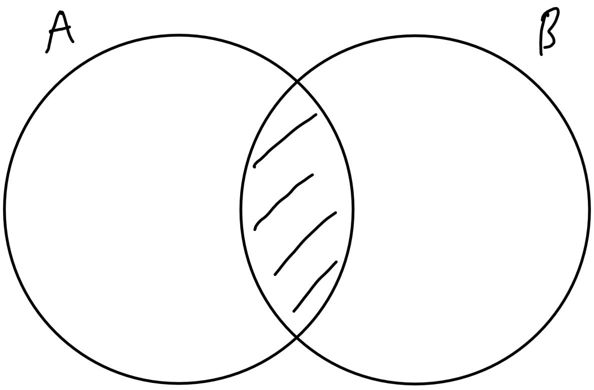
§3.1 Countable sub-additivity

Proposition 3.1 (Countable sub-additivity)
Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of events in $\mathcal{F}$. Then

$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) \leq \sum_{n\in\mathbb{N}} \mathbb{P}(A_n).$$

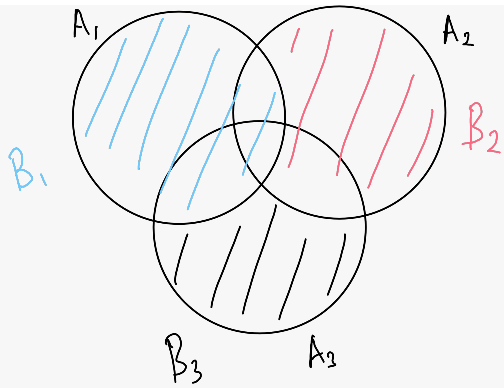
May also be called a *union bound*.

Intuition:



$\sum_{n\in\mathbb{N}} \mathbb{P}(A_n)$ "double counts" some sub-events.

Proof. Idea: Rewrite $\bigcup_{n\in\mathbb{N}} A_n$ as a *disjoint* union. Define $B_1 = A_1$ and $B_n = \underbrace{A_n \setminus (A_1 \cup \cdots \cup A_{n-1})}_{\in\mathcal{F} (\text{by Sheet 1})} \quad \forall\, n \geq 2$.



So

- $\bigcup_{n\in\mathbb{N}} B_n = \bigcup_{n\in\mathbb{N}} A_n$.
- $(B_n)_{n\in\mathbb{N}}$ is disjoint (by construction).
- $B_n \subseteq A_n \implies \underbrace{\mathbb{P}(B_n) \le \mathbb{P}(A_n)}_{\text{Q4, Sheet 1}}$

$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \mathbb{P}\left(\bigcup_{n\in\mathbb{N}} B_n\right) \underset{P3 \text{ on } (B_n)}{=} \sum_{n\in\mathbb{N}} \mathbb{P}(B_n) \le \sum_{n\in\mathbb{N}} \mathbb{P}(A_n)$$

□

§3.2 Continuity

Proposition 3.2 (Continuity)

Let $(A_n)_{n\in\mathbb{N}}$ be an increasing sequence of events in $\mathcal{F}$, i.e. $A_n \subseteq A_{n+1} \quad \forall\, n$. Then $\mathbb{P}(A_n) \le \mathbb{P}(A_{n+1})$. So $\mathbb{P}(A_n)$ converges as $n \to \infty$.[a]

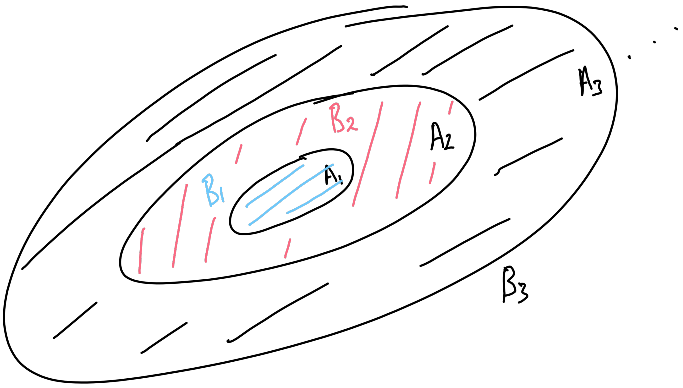
In fact: $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right)$.

[a] As probabilities are bounded above by 1 and increasing.

For motivation try Q6, Sheet 1.

Proof. Let us reuse the B_ns from the previous subsection.

- $\bigcup_{k=1}^{n} B_k = A_n$ (disjoint union).
- $\bigcup_{n\in\mathbb{N}} B_n = \bigcup_{n\in\mathbb{N}} A_n$



$$\mathbb{P}(A_n) = \sum_{k=1}^{n} \mathbb{P}(B_k) \overset{n\to\infty}{\to} \sum_{k\geq 1} \mathbb{P}(B_k) = \mathbb{P}\left(\bigcup_{n\in\mathbb{N}} B_n\right) = \mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right)$$

□

§3.3 Inclusion-Exclusion Principle

Background: $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
Similarly: $A, B, C \in \mathcal{F}$

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

Proposition 3.3 (Inclusion-Exclusion Principle)
Let $A_1, \ldots, A_n \in \mathcal{F}$, then:

$$\begin{aligned}\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) &= \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{1\leq i_1<i_2\leq n} \mathbb{P}(A_{i_1} \cap A_{i_2}) + \sum_{1\leq i_1<i_2<i_3\leq n} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap A_{i_3}) - \ldots \\ &\quad + (-1)^{n+1}\mathbb{P}(A_1 \cap \cdots \cap A_n) \\ &= \sum_{\substack{I\subset\{1,\ldots,n\} \\ I\neq\emptyset}} (-1)^{|I|+1}\mathbb{P}\left(\bigcap_{i\in I} A_i\right)\end{aligned}$$

Note: $\sum_{1\leq i_1<i_2\leq n}$ is the sum of all triples that are distinct and unordered.

Proof. By induction. For $n = 2$ it holds (Q4e, Sheet 1).

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \mathbb{P}\left(\left(\bigcup_{i=1}^{n-1} A_i\right) \cup A_n\right)$$

Using $n = 2$ case we get:

$$= \mathbb{P}\left(\bigcup_{i=1}^{n-1} A_i\right) + \mathbb{P}(A_n) - \mathbb{P}\left(\left(\bigcup_{i=1}^{n-1} A_i\right) \cap A_n\right)$$

We want to break down the final element on the RHS

$$\text{Idea: } \left(\bigcup_{i=1}^{n-1} A_i\right) \cap A_n = \bigcup_{i=1}^{n-1} (A_i \cap A_n)$$

If we apply IEP to $\bigcup_{i=1}^{n-1} (A_i \cap A_n)$ we need to calculate $\bigcap_{i \in J} (A_i \cap A_n)$

$$\bigcap_{i \in J} (A_i \cap A_n) = \bigcap_{i \in J \cup \{n\}} A_i, \qquad J \subset \{1, \dots, n-1\}$$

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) &= \underbrace{\sum_{\substack{J \subset \{1,\dots,n-1\} \\ J \neq \emptyset}} (-1)^{|J|+1} \mathbb{P}\left(\bigcap_{i \in J} A_i\right)}_{n-1 \text{ case}} + \mathbb{P}(A_n) \\
&\quad - \underbrace{\sum_{\substack{J \subset \{1,\dots,n-1\} \\ J \neq \emptyset}} (-1)^{|J|+1} \mathbb{P}\left(\bigcap_{i \in J \cup \{n\}} A_i\right)}_{n-1 \text{ case on } (A_i \cap A_n)} \\
J \cup \{n\} &\mapsto I. \\
-(-1)^{|J|+1} &\mapsto (-1)^{|I|+1} \\
&= \underbrace{\sum_{\substack{I \subset \{1,\dots,n-1\} \\ I \neq \emptyset}} (-1)^{|I|+1} \mathbb{P}\left(\bigcap_{i \in I} A_i\right)}_{\text{Just changed the labels}} + \mathbb{P}(A_n) \\
&\quad + \sum_{\substack{I \subset \{1,\dots,n\} \\ n \in I,\ |I| > 2}} (-1)^{|I|+1} \mathbb{P}\left(\bigcap_{i \in I} A_i\right) \\
&= \sum_{\substack{I \subset \{1,\dots,n\} \\ I \neq \emptyset}} (-1)^{|I|+1} \mathbb{P}\left(\bigcap_{i \in I} A_i\right).
\end{aligned}$$

Let us check that we have indeed counted all subsets I.

- $\sum_{\substack{I \subset \{1,\ldots,n-1\} \\ I \neq \emptyset}} (-1)^{|I|+1} \mathbb{P}\left(\bigcap_{i \in I} A_i\right)$ accounts for all subsets where $n \notin I$.
- $\mathbb{P}(A_n)$ accounts for $\{n\}$
- $\sum_{\substack{I \subset \{1,\ldots,n\} \\ n \in I,\ |I|>2}} (-1)^{|I|+1} \mathbb{P}\left(\bigcap_{i \in I} A_i\right)$ accounts for all subsets where $n \in I$ and $I \neq \{n\}$.

□

§3.4 Bonferroni Inequalities

Question

What if you *truncate* IEP (Inclusion-Exclusion Principle)?

Proposition 3.4 (Bonferroni Inequality)

Recall: Countable sub-additivity - $\mathbb{P}(\cup A_i) \leq \sum \mathbb{P}(A_i)$.

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{k=1}^{r} (-1)^{k+1} \sum_{i_1 < i_2 < \cdots < i_k} \mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) \quad \text{if } r \text{ is odd}$$

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \geq \sum_{k=1}^{r} (-1)^{k+1} \sum_{i_1 < i_2 < \cdots < i_k} \mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) \quad \text{if } r \text{ is even}$$

Proof. By induction on r and n. Let r be odd

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \mathbb{P}\left(\bigcup_{i=1}^{n-1} A_i\right) + \mathbb{P}(A_n) - \mathbb{P}\left(\bigcup_{i=1}^{n-1} (A_i \cap A_n)\right)$$

$$\mathbb{P}\left(\bigcup_{i=1}^{n-1} A_i\right) \leq \underbrace{\sum_{\substack{J \subset \{1,\ldots,n-1\} \\ 1 \leq |J| \leq r}} (-1)^{|J|+1} \mathbb{P}\left(\bigcap_{i \in J} A_i\right)}_{n-1 \text{ case.}}$$

$$\mathbb{P}\left(\bigcup_{i=1}^{n-1} (A_i \cap A_n)\right) \geq \underbrace{\sum_{\substack{J \subset \{1,\ldots,n-1\} \\ 1 \leq |J| \leq r-1^{a}}} (-1)^{|J|+1} \mathbb{P}\left(\bigcap_{i \in J \cup \{n\}} A_i\right)}_{r-1 \text{ case on } (A_i \cap A_n)}$$

Using the same rearranging as in the proof of Inclusion-Exclusion Principle

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{\substack{I \subset \{1,\dots,n\} \\ 1 \leq |I| \leq r}} (-1)^{|I|+1} \mathbb{P}\left(\bigcap_{i \in I} A_i\right).$$

The case of r being even is similar, simply note all three inequalities are reversed. □

[a] J doesn't include n and we only want r elements in the intersection

[a] J doesn't include n and we only want r elements in the intersection

Question

When is it good to truncate at e.g. r = 2?

§3.5 Counting with IEP

Uniform probability measure on Ω, $|\Omega| < \infty$. $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$ $\forall\ A \subseteq \Omega$. Then $\forall\ A_1, \dots, A_n \subseteq \Omega$

$$|A_1 \cup \dots \cup A_n| = \sum_{k=1}^{n} (-1)^{k+1} \sum_{i_1 < \dots < i_k} |A_i \cap \dots \cap A_{i_k}|$$

(and similarly for Bonferroni Inequalities).

Example 3.1 (Surjections)

What is the probability that a function $f : \{1, \dots, n\} \to \{1, \dots, m\}$, $n \geq m$ is a surjection? Let $\Omega = \{f : \{1, \dots, n\} \to \{1, \dots, m\}\}$ and $A = \{f \in \Omega : \text{Image}(f) = \{1, \dots, m\}\}$.
$\forall\ i \in \{1, \dots, m\}$ define $B_i = \{f \in \Omega : i \notin \text{Image}(f)\}$.

Key observations:

- $$\begin{aligned} A &= B_1^c \cap \dots B_m^c \\ &= (B_1 \cup \dots \cup B_m)^c \end{aligned}$$
- $|B_{i_1} \cap \dots \cap B_{i_k}|$ is nice to calculate.

$$\begin{aligned} |B_{i_1} \cap \dots \cap B_{i_k}| &= |\{f \in \Omega : i_1, \dots, i_k \notin \text{Image}(f)| \\ &= (m-k)^n \end{aligned}$$

$$
\begin{aligned}
IEP \to |B_1 \cup \cdots \cup B_m| &= \sum_{k=1}^{m} (-1)^{k+1} \sum_{i_1 < \cdots < i_k} \underbrace{|B_{i_1} \cap \cdots \cap B_{i_k}|}_{\text{same for all } i_1, \ldots, i_k} \\
&= \sum_{k=1}^{m} (-1)^{k+1} \binom{m}{k} (m-k)^n \\
|A| &= m^n - |B_{i_1} \cap \cdots \cap B_{i_k}| \\
&= \sum_{k=0}^{m} (-1)^{k} \binom{m}{k} (m-k)^n
\end{aligned}
$$

Example 3.2 (Derangements)

What is the probability that a permutation has no fixed points? Derangements can be useful in a Secret Santa.
$\Omega = \{\text{permuations of } \{1, \ldots, n\}\}$ and the derangements, D, are $\{\sigma \in \Omega : \sigma(i) \neq i \quad \forall\, i = 1, \ldots, n\}$.

Question

Is $\mathbb{P}(D) = \frac{|D|}{|\Omega|}$ large or small (e.g. when $n \to \infty$?)

$\forall\, i \in \{1, \ldots, n\} : A_i = \{\sigma \in \Omega : \sigma(i) = i\}$.
Key observations:

- $D = A_1^c \cap \ldots A_n^c = \left(\bigcup_{i=1}^n A_i\right)^c$.
-

$$
\begin{aligned}
\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) &= \frac{(n-k)!}{n!} \\
IEP \to \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) &= \sum_{k=1}^{n} (-1)^{k+1} \sum_{i_1 < \cdots < i_k} \mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) \\
&= \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} \frac{(n-k)!}{n!} \\
&= \sum_{k=1}^{n} (-1)^{k+1} \frac{1}{k!} \\
\mathbb{P}(D) &= 1 - \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \\
&= 1 - \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k!}
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{k=0}^{n} \frac{(-1)^k}{k!} \\
\lim_{n\to\infty} \mathbb{P}(D) &= \sum_{k=0}^{\infty} \frac{(-1)^{k+1}}{k!} \\
&= e^{-1} \approx 0.37.
\end{aligned}
$$

Remark 2.

- What if instead $\Omega' = \{\text{all functions } f : \{1,\dots,n\} \text{ to itself}\}$?

$$
\begin{aligned}
D = \{f \in \Omega' : f(i) \neq i \quad \forall\, i = 1,\dots,n\}. \\
\mathbb{P}(D) &= \frac{(n-1)^n}{n^n} \\
&= \left(1 - \frac{1}{n}^n\right) \\
\lim_{n\to\infty} D &= e^{-1}.
\end{aligned}
$$

- We would liked to have calculated $\mathbb{P}(D)$ by doing $\left(\frac{n-1}{n}\right)^n$ as we have n choices each with probability $\frac{n-1}{n}$. We will be allowed to do this soon! See Example 3.6
- $f(i)$ is a random quantity associated to Ω. We will be allowed to study $f(i)$ as a *random variable* soon.
- We are allowed to toss a fair coin n times, $\Omega = \{H,T\}^n$. But we have not yet studied tossing an unfair coin n times.

§3.6 Independence

$(\Omega, \mathcal{F}, \mathbb{P})$ as before.

Definition 3.2 (Indepence)

Events $A, B \in \mathcal{F}$ are **independent** ($A \perp\!\!\!\perp B$) if

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

A countable[a] collection of events (A_n) are **independent** if $\forall$ distinct $i_1,\dots,i_k$[b] we have:

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \prod_{j=1}^{k} \mathbb{P}(A_{i_j}).$$

[a] including finite
[b] k is finite

Remark 3 (Caution). "Pairwise independence" does not imply independence.

Example 3.3

$$\begin{aligned}
\Omega &= \{(H,H),(H,T),(T,H),(T,T)\} \\
\mathbb{P}(\{\omega\}) &= \frac{1}{4} \quad \forall\, \omega \in \Omega. \\
A &= \text{first coin is } H = \{(H,H),(H,T)\}. \\
B &= \text{second coin is } H = \{(T,H),(H,H)\}. \\
C &= \text{both coins have the same outcome} = \{(T,T),(H,H)\}. \\
\mathbb{P}(A) = \mathbb{P}(B) &= P(C) = \frac{1}{2}. \\
A \cap B = A \cap C &= B \cap C = \{(H,H)\}. \\
\mathbb{P}(A \cap B) = \mathbb{P}(A \cap C) &= \mathbb{P}(B \cap C) = \frac{1}{4}. \quad \text{Pairwise independence } \checkmark \\
\mathbb{P}(A \cap B \cap C) &= \frac{1}{4} \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C). \quad \text{Independence } \times
\end{aligned}$$

Example 3.4 (Independence)

- $$\begin{aligned}
\Omega' &= \{\text{all functions } f : \{1,\ldots,n\} \text{ to itself} \\
A_i &= \{f \in \Omega' : f(i) = i\}. \\
\mathbb{P}(A_i) &= \frac{n^{n-1}}{n^n} = \frac{1}{n} \\
\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) &= \frac{n^{n-k}}{n^n} \\
&= \frac{1}{n^k} \\
&= \prod_{j=1}^{k} \mathbb{P}(A_{i_j})
\end{aligned}$$

  Here: (A_i) are independent events.

- $$\begin{aligned}\Omega &= \{\sigma : \text{ permutation of } \{1,\dots,n\}\} \\ A_i &= \{\sigma \in \Omega : \sigma(i) = i\} \\ \mathbb{P}(A_i) &= \frac{(n-1)!}{n!} = \frac{1}{n}. \\ i \neq j \quad \mathbb{P}(A_i \cap A_j) &= \frac{(n-2)!}{n!} \\ &= \frac{1}{n(n-1)} \\ &\neq \mathbb{P}(A_i)\mathbb{P}(A_j)\end{aligned}$$

  Here: (A_i) are not independent events.

§3.6.1 Properties

Claim

If A is independent of B, then A is also independent of B^c.

Proof.

$$\begin{aligned}\mathbb{P}(A \cap B^c) &= \mathbb{P}(A) - \mathbb{P}(A \cap B) \\ &= \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) \\ &= \mathbb{P}(A)[1 - \mathbb{P}(B)] \\ &= \mathbb{P}(A)\mathbb{P}(B^C)\end{aligned}$$

□

Claim

A is independent of $B = \Omega$ and of $C = \emptyset$

Proof. $\mathbb{P}(A \cap \Omega) = P(A) = \mathbb{P}(A)\underbrace{\mathbb{P}(\Omega)}_{=1}$ and so $A \perp\!\!\!\perp \emptyset$ by Claim 1. □

§3.7 Conditional Probability

$(\Omega, \mathcal{F}, \mathbb{P})$ as before.

Consider $B \in \mathcal{F}$ with $\mathbb{P}(B) > 0,\ A \in \mathcal{F}$

Definition 3.3 (Conditional Probability)

The **conditional probability** of A given B is $\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$

"The probability of A is we know B happened". (e.g. revealing information in succession)

Example 3.5

A, B independent.

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A)$$

"Knowing whether B happened doesn't affect the probability of A".

§3.7.1 Properties

- P1 - $\mathbb{P}(A \mid B) \geq 0$
- P2 - $\mathbb{P}(B \mid B) = 1 = \mathbb{P}(\Omega \mid B)$
- P3 - (A_n) disjoint events $\in \mathcal{F}$:

Claim

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n \mid B\right) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n \mid B)$$

Proof.

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n \mid B\right) &= \frac{\mathbb{P}\left(\left(\bigcup_n A_n\right) \cap B\right)}{\mathbb{P}(B)} \\
&= \frac{\mathbb{P}\left(\bigcup_n (A_n \cap B)\right)}{\mathbb{P}(B)}, \quad \cup(A_n \cap B) \text{ is a disjoint union} \\
&= \frac{\sum_n \mathbb{P}(A_n \cap B)}{\mathbb{P}(B)} \\
&= \sum_{n \in \mathbb{N}} \mathbb{P}(A_n \mid B)
\end{aligned}$$

Summary: Use definition and apply P1, P2, P3 to the numerator. □

$\mathbb{P}(\bullet \mid B)$ is a function from $\mathcal{F} \to [0,1]$ that satisfies the rules to be a probability measure on Ω.

Aside?

Consider $\Omega' = B$ (especially in finite or countable setting). Let $\mathcal{F}' = \mathcal{P}(B)$. Then $(\Omega', \mathcal{F}', \mathbb{P}(\bullet \mid B))$ also satisfies rules to be a probability measure on Ω'.

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B \mid A) \tag{1}$$

$$\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1)\mathbb{P}(A_3 \mid A_1 \cap A_2)\dots\mathbb{P}(A_n \mid A_1 \cap \dots \cap A_{n-1})$$

Example 3.6

Uniform choice of a permutation $(\sigma(1), \sigma(2), \dots, \sigma(n)) \in \Sigma_n$.

Claim

$$\begin{aligned}\mathbb{P}\left(\sigma(k) = i_k \mid \sigma(1) = i, \dots, \sigma(k-1) = i_{k-1}\right), &\quad i_1, \dots, i_{k-1} \text{ distinct.} \\ = \begin{cases} 0 & \text{if } i_k \in \{i_1, \dots, i_{k-1}\}. \\ \frac{1}{n-k+1}{}^{a} & \text{else} \end{cases}\end{aligned}$$

[a] This is an example of (Ordered) compositions

[a] This is an example of (Ordered) compositions

Proof.

$$\begin{aligned}\mathbb{P}\left(\sigma(k) = i_k \mid \sigma(1) = i, \dots, \sigma(k-1) = i_{k-1}\right) &= \frac{\mathbb{P}(\sigma(1) = i, \dots, \sigma(k) = i_k)}{\mathbb{P}(\sigma(1) = i, \dots, \sigma(k-1) = i_{k-1})} \\ &= \frac{\frac{(n-k)!}{n!}}{\frac{(n-k+1)!}{n!}} \\ &= \frac{(n-k)!}{(n-k+1)!} \\ &= \frac{1}{n-k+1} \\ \mathbb{P}(\sigma(1) = i, \dots, \sigma(k) = i_k) &= 0 \quad \text{if } i_k \in \{i_1, \dots, i_{k-1}\}.\end{aligned}$$

□

§3.7.1 Law of Total Probability and Bayes' Formula

Definition 3.4 (Partition)
$(B_1, B_2, \dots)^a \in \Omega$ is a **partition** of Ω if:

- $\Omega = \bigcup_n B_n$
- (B_n) are disjoint

[a] finite or countable

Theorem 3.1 (Law of Total Probability)
(B_n) a finite or countable partition of Ω with $B_n \in \mathcal{F}\ \forall\ n$ s.t. $\mathbb{P}(B_n) > 0$. Then $\forall\ A \in \mathcal{F}$:

$$\mathbb{P}(A) = \sum_n \mathbb{P}(A \mid B_n)\mathbb{P}(B_n).$$

Also know as "Partition Theorem".

Proof. Note that $\bigcup_n (A \cap B_n) = A$.

$$\begin{aligned}\mathbb{P}(A) &= \sum_{n\in\mathbb{N}} \mathbb{P}(A \cap B_n) \\ &= \sum_n \mathbb{P}(A \mid B_n)\mathbb{P}(B_n) \quad \text{by Equation (1)}\end{aligned}$$

□

Theorem 3.2 (Bayes' Formula)
Same setup as above

$$\begin{aligned}\mathbb{P}(B_n \mid A) &= \frac{\mathbb{P}(A \cap B_n)}{\mathbb{P}(A)} \\ &= \frac{\mathbb{P}(A \mid B_n)\mathbb{P}(B_n)}{\sum_m \mathbb{P}(A \mid B_m)\mathbb{P}(B_m)}.\end{aligned}$$

Let $n = 2$: $\mathbb{P}(B \mid A)\mathbb{P}(A) = \mathbb{P}(A \mid B)\mathbb{P}(B) = \mathbb{P}(A \cap B)$

Example 3.7 (Lecture course)
Consider a Lecture course which has 2/3 of the lectures on weekdays and 1/3 on

weekends. Let

$$\mathbb{P}(\text{forget notes} \mid \text{weekday}) = \frac{1}{8}$$
$$\mathbb{P}(\text{forget notes} \mid \text{weekend}) = \frac{1}{2}$$

What is $\mathbb{P}(\text{weekend} \mid \text{forget notes})$? Let $B_1 = \{\text{weekday}\}, B_2 = \{\text{weekend}\}$ and $A = \{\text{forget notes}\}$.
By LTP (Law of Total Probability): $\mathbb{P}(A) = \frac{2}{3} \times \frac{1}{8} + \frac{1}{3} \times \frac{1}{2} = \frac{1}{4}$.
By Bayes' Formula: $\mathbb{P}(B_2 \mid A) = \frac{1}{3} \times \frac{1}{2}/\frac{1}{4} = \frac{2}{3}$.

Example 3.8 (Disease Testing)

Suppose p are infected and $(1-p)$ are not. $\mathbb{P}(\text{tests positive} \mid \text{infected}) = 1 - \alpha$ and $\mathbb{P}(\text{tests positive} \mid \text{not infected}) = \beta$ where $\alpha, \beta \in (0, 1)$.

We want to work out $\mathbb{P}(\text{infected} \mid \text{test positive})$.

By LTP: $\mathbb{P}(\text{test positive}) = p(1-\alpha) + (1-p)\beta$.
By Bayes' Formula: $\mathbb{P}(\text{infected} \mid \text{test positive}) = \frac{p(1-\alpha)}{p(1-\alpha)+(1-p)\beta}$.

Suppose $p \ll \beta$ then $p(1-\alpha) \ll (1-p)\beta$ so $\mathbb{P}(\text{infected} \mid \text{test positive}) \sim \frac{p(1-\alpha)}{(1-p)\beta} \sim \frac{p}{\beta}$ which is small.

Example 3.9 (Simpson's Paradox)

Scientists ask: do jelly beans make you tongue change colour?

Oxford	Change	No change	% change	
Blue	15	22	41%	$\Delta = 3\%$
Green	5	8	38%	

Cambridge	Change	No change	% change	
Blue	10	3	77%	$\Delta = 15\%$
Green	23	14	62%	

Total	Change	No change	% change	
Blue	25	25	50%	$\Delta = -6\%$
Green	28	22	56%	

The conclusion from this example should be that the Cambridge methodology is different to the Oxford one rather than anything about blue/ green jelly beans.[a]

Let $A = \{\text{change colour}\}$, $B = \{\text{blue}\}$, $B^C = \{\text{green}\}$, $C = \{\text{Cambridge}\}$, $C^c = \{\text{Oxford}\}$.

$$\begin{aligned}
&\mathbb{P}(A \mid B \cap C) > \mathbb{P}(A \mid B^c \cap C) \\
&\mathbb{P}(A \mid B \cap C^c) > \mathbb{P}(A \mid B^c \cap C^c) \\
&\not\Longrightarrow \ \mathbb{P}(A \mid B) > \mathbb{P}(A \mid B^c)
\end{aligned}$$

[a] Obviously this is a frivolous example however if we changed Oxford to November 2021, Cambridge to January 2021 and we were measuring vaccine efficacy for different vaccines we would get similar results. And it would be reasonable to conclude that the main underlying factor was a change in the viral landscape rather than waning efficacy.

Theorem 3.3 (Law of Total Probability for Conditional Probabilities)
Suppose $C_1, C_2, \ldots$ a partition of B.

$$\mathbb{P}(A \mid B) = \sum_n \mathbb{P}(A \mid C_n)\mathbb{P}(C_n \mid B)$$

Proof.

$$\begin{aligned}
\mathbb{P}(A \mid B) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \\
&= \frac{\mathbb{P}\left(A \cap \left(\bigcup_n C_n\right)\right)}{\mathbb{P}(B)} \\
&= \frac{\mathbb{P}\left(\bigcup_n (A \cap C_n)\right)}{\mathbb{P}(B)} \\
&= \frac{\sum_n \mathbb{P}(A \cap C_n)}{\mathbb{P}(B)} \\
&= \frac{\sum_n \mathbb{P}(A \mid C_n)\mathbb{P}(C_n)}{\mathbb{P}(B)} \\
&= \sum_n \mathbb{P}(A \mid C_n)\frac{\mathbb{P}(B \cap C_n)}{\mathbb{P}(B)} \qquad C_n \subset B \implies B \cap C_n = C_n \\
&= \sum_n \mathbb{P}(A \mid C_n)\mathbb{P}(C_n \mid B)
\end{aligned}$$

$\square$

Non Examinable

Special Case:

- If all $\mathbb{P}(C_n)$ are equal then so are $\mathbb{P}(C_n \mid B)$. Note $\sum_n \mathbb{P}(C_n \mid B) = 1$.
- If $\mathbb{P}(A \mid C_n)$ are all equal.

Then $\mathbb{P}(A \mid B) = \mathbb{P}(A \mid C_n)$.

Example 3.10 (Well-shuffled deck of cards)

Uniformly chosen *permutation*, $\sigma \in \Sigma_{52}$, of 52 cards. $\{1, 2, 3, 4\}$ are *aces*. Let $A = \{\sigma(1), \sigma(2) \text{ are aces}\}$, $B = \{\sigma(1) \text{ is an ace}\} = \{\sigma(1) \le 4\}$, $C_1 = \{\sigma(1) = 1\} \dots C_4 = \{\sigma(1) = 4\}$.

Note:

- $$\begin{aligned} \mathbb{P}(A \mid C_i) &= \mathbb{P}(\sigma(2) \in \{1, 2, 3, 4\} \mid \sigma(1) = i) \quad i \le 4 \\ &= \frac{3}{51} \text{ by } \textbf{??} \end{aligned}$$
- $\mathbb{P}(C_1) = \cdots = \mathbb{P}(C_4) = \frac{1}{52}$.

So $\mathbb{P}(A \mid B) = \frac{3}{51}$ and $\mathbb{P}(A) = \mathbb{P}(B)\mathbb{P}(A \mid B) = \frac{4}{52} \times \frac{3}{51}$

§4 Discrete Random Variables

Motivation: Roll two dice. $\Omega = \{1, \dots, 6\}^2 = \{(i, j) : 1 \le i, j \le 6\}$. If we restrict our attention to:

- the first dice e.g. $\{(i, j) : i = 3\}$.
- the sum of the dice e.g. $\{(i, j) : i + j = 8\}$.
- the max of the dice e.g. $\{(i, j) : i, j \le 4, i \text{ or } j = 4\}$.

This is annoying and we want to move on from sets.

Goal: "Random real-valued measurements", we want the value of the first dice to be X and the sum to be $X + Y$ ...

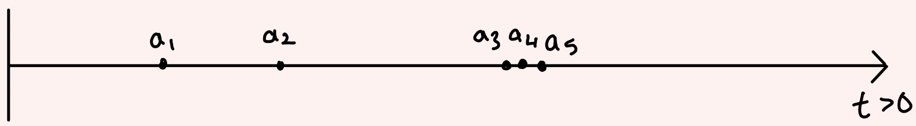
Definition 4.1 (Discrete Random Variable)

A **discrete random variable** X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a function $X : \Omega \to \mathbb{R}$ s.t.

- $\{\omega \in \Omega : X(\omega) = x\} \in \mathcal{F}$
- $\text{Image}(X)$ if finite or countable (subset of $\mathbb{R}$).
- We abbreviate $\{\omega \in \Omega : X(\omega) = x\}$ as $\{X = x\}$. So $\mathbb{P}(X = x)$ is valid.
- Often $\text{Image}(X) = \mathbb{Z}$ or $\mathbb{N}_0$ or $\{0, 1\}$ etc. *not* {Heads or Tails}.

If Ω is finite or countable and $\mathcal{F} = \mathcal{P}(\Omega)$ both blue bullet points hold automatically.

Example 4.1 (Part II Applied Probability)



"random arrival process". Let $\Omega = \{\text{countable subsets } (a_1, a_2, \dots) \text{ of } (0, \infty)\}$ and N_t = number of arrivals by time $t = |\{a_i : a_i \le t\}| \in \mathbb{N}_0$ is a discrete RV (random variable) for each time t.

Definition 4.2 (Probability Mass Function)

The **probability mass function** of discrete RV X is the function $p_X : \mathbb{R} \to [0, 1]$ given by $p_X(x) = \mathbb{P}(X = x) \quad \forall\, x \in \mathbb{R}$.

Note.

- if $x \notin \text{Image}(X)$ then $p_X(x) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\}) = \mathbb{P}(\emptyset) = 0$.
-

$$\begin{aligned}
\sum_{x \in \text{Im}(X)} p_X(x) &= \sum_{x \in \text{Im}(X)} \mathbb{P}(\underbrace{\{\omega \in \Omega : X(\omega) = x\}}_{\text{disjoint}}) \\
&= \mathbb{P}\left(\bigcup_{x \in \text{Im}(X)} \{\omega \in \Omega : X(\omega) = x\} \right) \\
&= \mathbb{P}(\Omega) \\
&= 1
\end{aligned}$$

Example 4.2 (Indicator function)
For event $A \in \mathcal{F}$, define $1_A : \Omega \to \mathbb{R}$ by

$$1_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{else} \end{cases}$$

1_A is a discrete RV with Image $= \{0, 1\}$. $p_{1_A}(1) = \mathbb{P}(1_A = 1) = \mathbb{P}(A)$, $p_{1_A}(0) = \mathbb{P}(1_A = 0) = \mathbb{P}(A^c)$ and $p_{1_A}(x) = 0 \quad \forall\, x \notin \{0, 1\}$.

This encodes "did A happen" as a real number.

Remark 4. Given a pmf p_X (probability mass function), we can always construct a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a RV defined on it with this pmf.

- $\Omega = \text{Im}(X)$ i.e. $\{x \in \mathbb{R} : p_X(x) > 0\}$
- $\mathcal{F} = \mathcal{P}(\Omega)$
- $\mathbb{P}(\{x\}) = p_X(x)$ and extend to all $A \in \mathcal{F}$

§4.1 Discrete Probability Distributions

Ω is finite.

§4.1.1 Bernoulli distribution - "(biased) coin toss"

If $X \sim \text{Bern}(p)$ where $p \in [0, 1]$ then $\text{Im}(X) = \{0, 1\}$ and $p_X(1) = \mathbb{P}(X = 1) = p$ and $p_X(0) = 1 - p$.

Example 4.3

$1_A \sim \mathrm{Bern}(p)$ with $p = \mathbb{P}(A)$.

§4.1.2 Binomial distribution

This can be used to model the number of heads when a coin is tossed n times.

If $X \sim \mathrm{Bin}(n,p)$ where $n \in \mathbb{Z}^+$ and $p \in [0,1]$ then $\mathrm{Im}(X) = \{0,1,\dots,n\}$, $p_X(k) = \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$. $\sum_{k=0}^n p_X(k) = (p + (1-p))^n = 1$.

§4.1.3 More than one RV

Motivation: Roll a dice with outcome $X \in \{1,2,\dots,6\}$. Events: $A = \{1 \text{ or } 2\}$, $B = \{1 \text{ or } 2 \text{ or } 3\}$, $B = \{1 \text{ or } 3 \text{ or } 5\}$. $1_A \sim \mathrm{Bern}\left(\frac{1}{3}\right)$, $1_B \sim \mathrm{Bern}\left(\frac{1}{2}\right)$, $1_C \sim \mathrm{Bern}\left(\frac{1}{2}\right)$.
Note: $1_A \le 1_B$ for all outcomes
but $1_A \le 1_C$ for all outcomes *is false*.

Definition 4.3 (Independent RVs)

Let $X_1, \dots, X_n$ be discrete RVs. We say $X_1, \dots, X_n$ are *independent* if:

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}(X_1 = x) \dots \mathbb{P}(X_n = x_n) \quad \forall\, x_1, \dots, x_n \in \mathbb{R}.$$

(suffices to check $\forall\, x_i \in \mathrm{Im}(X_i)$)

Example 4.4

$X_1, \dots, X_n$ independent RVs each with the Bernoulli(p) distribution. Study $S_n = X_1 + \dots + X_n$. Then

$$\begin{aligned}
\mathbb{P}(S_n = k) &= \sum_{\substack{x_1 + \dots + x_n = k \\ x_i \in \{0,1\}}} \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) \\
&= \sum_{x_1 + \dots + x_n = k} \mathbb{P}(X_1 = x_1) \dots \mathbb{P}(X_n = x_n) \\
&= \sum_{x_1 + \dots + x_n = k} p^{|\{i : x_i = 1\}|} (1-p)^{|\{i : x_i = 0\}|} \\
&= \sum_{x_1 + \dots + x_n = k} p^k (1-p)^{n-k} \\
&= \binom{n}{k} p^k (1-p)^{n-k}.
\end{aligned}$$

So $S_n \sim \text{Bin}(n, p)$.

Example 4.5 (Non-example)

$(\sigma(1), \sigma(2), \dots, \sigma(n))$ a uniform permutation.

Claim

$\sigma(1)$ and $\sigma(2)$ are *not* independent.

Suffices to find i_1, i_2 s.t. $\mathbb{P}(\sigma(1) = i_1, \sigma(2) = i_2) \neq \mathbb{P}(\sigma(1) = i_1)\mathbb{P}(\sigma(2) = i_2)$. E.g. $\mathbb{P}(\sigma(1) = 1, \sigma(2) = 1) = 0 \neq \underbrace{\mathbb{P}(\sigma(1) = 1)\mathbb{P}(\sigma(2) = 1)}_{=1/n \times 1/n}$

Consequence of definition

Let $X_1, \dots, X_n$ be independent. Then $\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \mathbb{P}(X \in A_1) \dots \mathbb{P}(X_n \in A_n) \quad \forall\, A_1, \dots, A_n \subset \mathbb{R}$ countable.

Let $\Omega = \mathbb{N}$, "Ways of choosing a random integer"

§4.1.4 Geometric distribution ("Waiting for success")

This can be used to model the number of coin tosses until we get a head.

If $X \sim \text{Geo}(p)$ were $p \in (0, 1)$. $\text{Im}(X) = \{1, 2, \dots\}$,
$p_X(k) = \mathbb{P}(\text{(k-1) failures, then success on the kth trial}) = (1-p)^{k-1}p$. Check: $\sum_{k \geq 1}(1-p)^{k-1}p = p\sum_{t \geq 0}(1-p)^t = \frac{p}{1-(1-p)} = 1$.

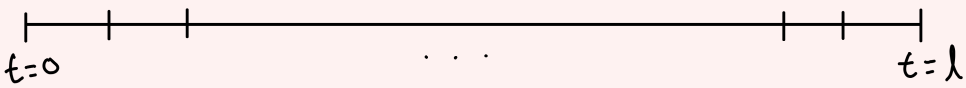
Alternatively: "Count how many failures before a success"
$\text{Im}(Y) = \{0, 1, 2, \dots\}$, $p_Y(k) = \mathbb{P}(\text{k failures, then success on the (k+1)th trial})$. Check: $\sum_{k \geq 0}(1-p)^k p = 1$.

§4.1.5 Poisson Distribution

If $X \sim \text{Po}(\lambda)$ (or $\text{Poi}(\lambda)$) with $\lambda \in (0, \infty)$. $\text{Im}(X) = \{0, 1, 2, \dots\}$ and $\mathbb{P}(X = k) = e^{-\lambda}\lambda^k / k! \quad \forall\, k \geq 0$. Check: $\sum_{k \geq 0} \mathbb{P}(X = k) = e^{\lambda} \sum_{k \geq 0} \frac{\lambda^k}{k!} = e^{\lambda} e^{\lambda}$.

Motivation: Consider $X_n \sim \text{Bin}(n, \frac{\lambda}{n})$

Example 4.6 ("Arrival proccess")



Split time interval $[0, \lambda]$ into n small intervals.

- Probability of an arrival in each interval is p, independently across intervals.
- Total no. of arrivals is X_n.

$$\mathbb{P}(X_n = k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

Fix k and let $n \to \infty$

$$= \frac{n!}{n^k (n-k)!} \times \underbrace{\frac{\lambda^k}{k!}}_{\text{no n}} \times \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\to e^{-\lambda}} \times \underbrace{\left(1 - \frac{1}{n}\right)^{-k}}_{\to 1}$$

$$\begin{aligned}
\frac{n!}{n^k (n-k)!} &= \frac{n(n-1)\dots(n-k+1)}{n^k} \\
&= 1 \times \left(1 - \frac{1}{n}\right) \times \left(1 - \frac{2}{n}\right) \times \cdots \times \left(1 - \frac{k-1}{n}\right) \\
&\to 1 \quad \text{There are a fixed number of terms all converging to 1}
\end{aligned}$$

$$\mathbb{P}(X_n = k) \underset{n \to \infty}{\to} e^{-\lambda} \frac{\lambda^k}{k!}.$$