Part IA — Probability

Based on lectures by Dr D. Yeo

Lent 2022

Contents

0 Introduction 2

1 Formal Setup 3

1.1 Examples of Probability Spaces 5

2 Combinatorial Analysis 7

2.1 Subsets 7

2.2 Random Walks 7

2.3 Stirling's Formula 8

2.4 (Ordered) compositions 9

3 Properties of Probability measures 10

3.1 Countable sub-additivity 10

3.2 Continuity 11

3.3 Inclusion-Exclusion Principle 12

§0 Introduction

Probability theory is the mathematical formulation of randomness. Examples include the modelling of random experiments like flipping a coin, throwing a die, shuffle a deck, and so on. What we want to do is to develop a mathematical framework to study randomness.

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

□

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
$A^c = \{\text{all } n \text{ birthdays are different}\}$.

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

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

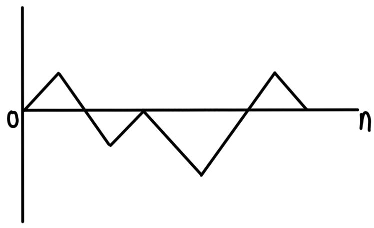
$$\begin{aligned}
M &= \underbrace{\binom{n}{n_1}}_{\substack{\text{Choose}\\\text{first part}}} \underbrace{\binom{n-n_1}{n_2}}_{\substack{\text{Then choose}\\\text{second part}}} \binom{n-n_1-n_2}{n_3} \dots \underbrace{\binom{n-(n_1+\dots+n_{k-1})}{n_k}}_{=1} \\
&= \frac{n!}{n_1!\cancel{(n-n_1)!}}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

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



$$\Omega = \{(X_0, X_1, \dots, X_n) : X_0 = 0, |X_n - X_{k-1}| = 1 \ \forall\ k = 1, \dots, n\}.$$
$$|\Omega| = 2^n \text{ (we can go either up or down at each } k)$$

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

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 $n \to \infty$.

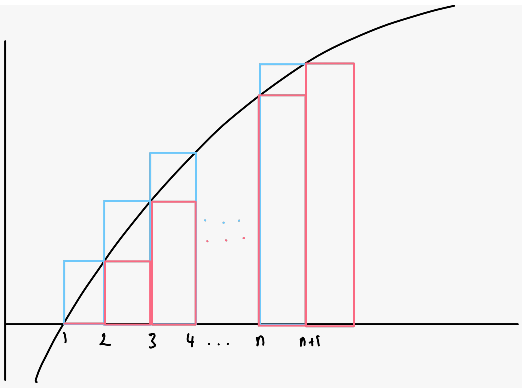
Example 2.1
$n^2 + 5n + \frac{6}{n} \sim n^2$

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



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

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

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 $(m_1, \ldots, m_k)$ 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.

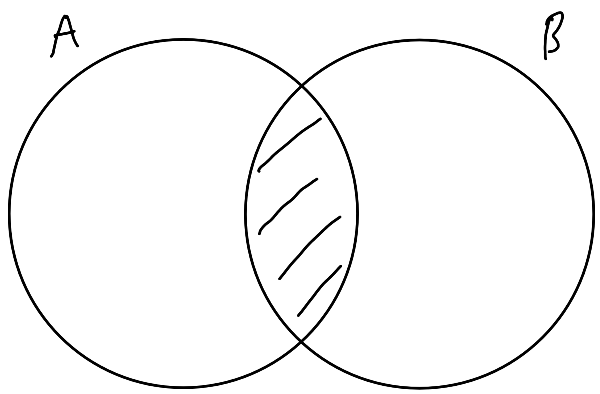
Question
What if the sets are not disjoint?

§3.1 Countable sub-additivity

Proposition 3.1 (Countable sub-additivity)
Let $(A_n)_{n\geq 1}$ be a sequence of events in $\mathcal{F}$. Then

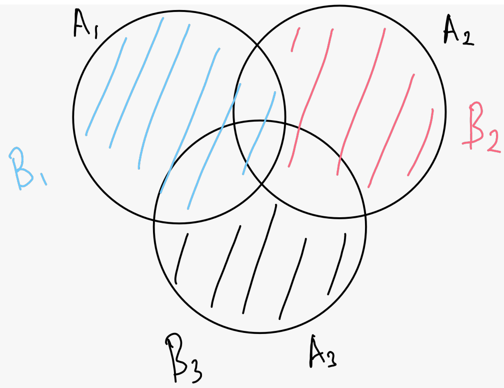
$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) \leq \sum_{n\in\mathbb{N}} \mathbb{P}(A_n).$$

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).
- $B_n \subseteq A_n \implies \underbrace{\mathbb{P}(B_n) \leq \mathbb{P}(A_n)}_{\text{Q4, Sheet 1}}$

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

□

§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 $\mathbb{P}(A_n) \leq \mathbb{P}(A_{n+1})$. So $\mathbb{P}(A_n)$ converges as $n \to \infty$.[a]

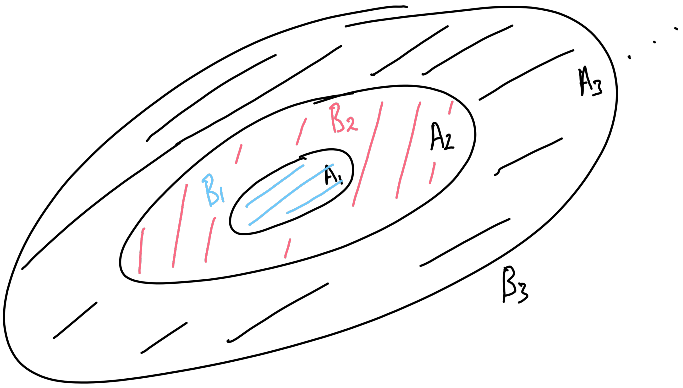
In fact: $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right)$.

[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: $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cup B)$.
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.