STAT 240 Course Notes - Fall 2018

Max Zhu

September 26, 2018

Contents

1 Foundations

Andrey Kolmogorov *(1933, "Foundations of the theory of probability")* put probability on solid mathematical grounds using a model, probability space $(\Omega, \mathcal{F}, P)$. A probability space consists of:

1. Sample space Ω: set of all outcomes ω
2. σ-algebra $\mathcal{F}$: set of all events, i.e. subsets of Ω to which we can assign a probability
3. Probability measure $P : \mathcal{F} \to [0,1]$: function which assigns probabilities to events

We need measure theory to understand this.

1.1 σ-algebras and measures

A classical problem is to measure the volume $\lambda(A)$ of some $A \subseteq \mathbb{R}^d, d \geq 1$. Consider $d = 1$. Then, λ should:

(i) Assign to intervals its length: $\lambda([a,b]) = b - a$ for all $a, b \in \mathbb{R}, a \leq b$

(ii) Be invariant under translations, rotations, and reflections: $\lambda(A) = \lambda(B)$ for all congruent $A, B \subseteq \mathbb{R}$

(iii) Be σ-additive:
If $\{A_i\}_{i\in\mathbb{N}} \subseteq \mathbb{R}, A_i \cap A_j = \varnothing \ \forall i \neq j$, then $\lambda(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \lambda(A_i)$

In other words, the volume of the union of countably many subsets of $\mathbb{R}$ is equal to the sum of their volumes.

Can we take the power set $\mathcal{P}(\mathbb{R}) = \{A : A \subseteq \mathbb{R}\}$ as $\mathcal{F}$ and find such a λ?

Theorem 1.1: Vitali's Theorem

There exists no λ defined on $\mathcal{P}(\mathbb{R})$ which fulfils (i)-(iii). Furthermore, any measurable set $A \subseteq \mathbb{R} : \lambda(A) > 0$ contains a non-measurable set V (Vitali set).

What about weakening (iii) to finitely many sets? Still no!

Theorem 1.2: Banach-Tarski

Let $d \geq 3$ and $A, B \in \mathbb{R}^d$ be bounded with non-empty interior. Then, there exists $k \in \mathbb{N}$ and partitions:

$$A = \dot{\bigcup}_{i=1}^{k} A_i$$
$$B = \dot{\bigcup}_{i=1}^{k} B_i$$

such that A_i, B_i are congruent $\forall i \in \{1, ..., k\}$.

For countable Ω, one can define λ (or μ or P) on $\mathcal{P}(\Omega)$ but for uncountable Ω, $\mathcal{P}(\Omega)$ is too large. Therefore, we need to define λ, μ, or P on some proper subset of $\mathcal{P}(\Omega)$ which is closed under certain set operations.

Definition 1.3: σ-algebra

$\mathcal{F} \subseteq \mathcal{P}(\Omega)$ is a σ-algebra on Ω if:

(i) $\Omega \in \mathcal{F}$

(ii) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$

(iii) $\{A_i\}_{i\in\mathbb{N}} \subseteq \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

If (iii) holds for only finitely many sets, $\mathcal{F}$ is an algebra.

Remark 1.4

σ-algebras are closed w.r.t. countable intersection, since

$$\bigcap_{i=1}^{\infty} A_i = (\bigcup_{i=1}^{\infty} A_i^c)^c \in \mathcal{F}$$

by de Morgan.

Example 1.5

1. Trivial σ-algebra: $\{\varnothing, \Omega\} = \mathcal{F}$
2. $\mathcal{F} = \{\dot{\bigcup}_{i=1}^{n}(a_i, b_i] : 0 \leq a_i \leq b_i \leq 1\ \forall i, n \in \mathbb{N}\}$ is an algebra on $\Omega = (0, 1]$ but not a σ-algebra since $\dot{\bigcup}_{n=0}^{\infty}(\sum_{k=1}^{2n}(\frac{1}{2})^k, \sum_{k=1}^{2n+1}(\frac{1}{2})^k] \notin \mathcal{F}$, while $(\sum_{k=1}^{2n}(\frac{1}{2})^k, \sum_{k=1}^{2n+1}(\frac{1}{2})^k] \in \mathcal{F}$ for all k.

How can σ-algebras be constructed?

Proposition 1.6

Given $A \subseteq \mathcal{P}(\Omega)$, then there exists a unique minimal σ-algebra $\sigma(A)$ which contains all sets of A: a σ-algebra generated by A. $\sigma(A)$ is the intersection of all σ-algebras of which A is a subset.

Proof. Let $\mathcal{F}_A = \{\mathcal{F} : \mathcal{F}$ is a σ-algebra, $A \subseteq \mathcal{F}\}$. Then, $\sigma(A)$ is a σ-algebra since:

(i) $\Omega \in \mathcal{F}\ \forall \mathcal{F} \in \mathcal{F}_A$ since all $\mathcal{F} \in \mathcal{F}_a$ is a σ-algebra

(ii) $A \in \sigma(A) \Rightarrow A \in \mathcal{F}\ \forall \mathcal{F} \in \mathcal{F}_A \Rightarrow A^c \in \mathcal{F}\ \forall \mathcal{F} \in \mathcal{F}_A \Rightarrow A^c \in \sigma(A)$.

(iii) If $\{A_i\}_{i \in \mathbb{N}} \subseteq \sigma(A)$, then $\{A_i\}_{i \in \mathbb{N}} \subseteq \mathcal{F}\ \forall \mathcal{F} \in \mathcal{F}_A$.
So, $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}\ \forall \mathcal{F} \in \mathcal{F}_A$, so $\bigcup_{i=1}^{\infty} A_i \in \sigma(A)$.

So $\sigma(A)$ is a σ-algebra. Now, $\sigma(A) \supseteq A$ since $\mathcal{F} \supseteq A\ \forall \mathcal{F} \in \mathcal{F}_A$. Also, $\forall$ σ-algebra $\mathcal{F}' \supset A$, we have $\mathcal{F}' \in \mathcal{F}_A$, so $\mathcal{F}' \supseteq \sigma(A)$.

Therefore, $\sigma(A)$ is the minimal σ-algebra containing A. □

Remark 1.7

Unless $|\Omega| < \infty$, a construction of $\sigma(A)$ is typically hopeless.

Example 1.8

$\mathcal{B}(\Omega) := \sigma(\{O : O \subseteq \Omega, O \text{ is open}\})$ is the Borel σ-algebra on Ω. Its elements are called Borel sets. For $\Omega = \mathbb{R}^d$ one can show that

$$\begin{aligned}\mathcal{B}(\mathbb{R}^d) &= \sigma(\{(a, b] : a \leq b\}) \\ &= \sigma(\{(a, b) : a \leq b\}) \\ &= \sigma(\{[a, b] : a \leq b\}) \\ &= \sigma(\{(-\infty, b]\})\end{aligned}$$

and so on. Borel sets contain open sets, closed sets, and countable union and intersections of these sets.

Definition 1.9

Let $\mathcal{F}$ be a σ-algebra on Ω. Then, $(\Omega, \mathcal{F})$ is a measurable space, sets in $\mathcal{F}$ are measurable sets. A measure μ on $\mathcal{F}$ is a function such that:

(i) $\mu : \mathcal{F} \to [0, \infty]$

(ii) $\mu(\varnothing) = 0$

(iii) Let $\{A_i\}_{i\in\mathbb{N}} \subseteq \mathcal{F}, A_i \cap A_j = \varnothing \ \forall i \neq j$. Then, $\mu(\dot{\bigcup}_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$. ($\sigma$-additivity)

$(\Omega, \mathcal{F}, \mu)$ is then called a measure space.

If $\Omega = \bigcup_{i=1}^{\infty} A_i$ for $\{A_i\}_{i\in\mathbb{N}} \subseteq \mathcal{F} : \mu(A_i) < \infty \ \forall i$, then μ is σ-finite.

If $\mu(\Omega) < \infty$, μ is a finite measure.

A measure μ on $\mathcal{B}(\mathbb{R}^d)$ is a Borel measure on $\mathbb{R}^d$.

Remark 1.10

1. σ-additivity (in contrast to finite additivity) allows for limiting processes (pointwise limits of "measurable functions" are measurable). Many fundamental consequences follow, such as Central Limit Theorem and Law of Large Numbers.

2. Uncountable additivity is too strong, since for any $A \subseteq \mathbb{R}$:

$$\begin{aligned} \lambda(A) &= \lambda(\bigcup_{x\in A} \{x\}) \\ &= \sum_{x\in A} \lambda(\{x\}) \\ &= sup_{A'\subseteq A, |A'|<\infty} \sum_{x\in A} \lambda(\{x\}) \\ &= 0 \end{aligned}$$

Example 1.11

If Ω is countable, $\mathcal{F} = \mathcal{P}(\Omega)$, $\forall f : \Omega \to [0, \infty]$, $\mu(A) = \sum_{\omega\in A} f(\omega) \ \forall A \in \mathcal{F}$ defines a measure on $\mathcal{F}$.

If $f(x) = 1 \ \forall x \in \Omega$, μ is called a counting measure.

Suppose for some $\omega_0 \in \Omega$, $f(\omega) = 1$ if $\omega = \omega_0$, 0 otherwise. Then μ is point mass or Dirac measure.

Proposition 1.12

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Then:

1. $A, B \in \mathcal{F}, A \subseteq B \Rightarrow \mu(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mu(A_i)$ (monotonicity)
2. $\{A_i\}_{i\in\mathbb{N}} \subseteq \mathcal{F} \Rightarrow \mu(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mu(A_i)$ (sub additivity)

3.

$$\{A_i\}_{i\in\mathbb{N}} \subseteq \mathcal{F}, A_1 \subseteq ... \subseteq A_n \subseteq ... \Rightarrow \mu(\bigcup_{i=1}^{\infty} A_i) = \mu(\lim_{n\to\infty}\bigcup_{i=1}^{n} A_i)$$
$$= \lim_{n\to\infty}\mu(A_n)$$

(continuity from below)

4.

$$\{A_i\}_{i\in\mathbb{N}} \subseteq \mathcal{F}, A_1 \supseteq ... \supseteq A_n \supseteq ... \Rightarrow \mu(\bigcap_{i=1}^{\infty} A_i) = \mu(\lim_{n\to\infty}\bigcap_{i=1}^{n} A_i)$$
$$= \lim_{n\to\infty}\mu(A_n)$$

for $\mu(A_1) < \infty$. (continuity from above)

Proof.

1. $B = A\dot{\cup}(B\backslash A)$
   Therefore, $\mu(B) = \mu(A) + \mu(B\backslash A) \geq 0$
   $\mu(B) \geq \mu(A)$ and if $\mu(A) < \infty$,
   $\mu(B\backslash A) = \mu(B) - \mu(A)$.

2. Let $B_1 = A_1$, $B_n = A_n\backslash\bigcup_{i=1}^{n-1} A_i \subseteq A_n$ $\forall n \geq 2$.
   So, all B_n are pairwise disjoint and $\bigcup_{i=1}^{n} B_i = \bigcup_{i=1}^{n} A_i$ $\forall n \in \mathbb{N}$.
   So, $\mu(\bigcup_{i=1}^{\infty} A_i) = \mu(\bigcup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty}\mu(B_i) \leq \sum_{i=1}^{\infty}\mu(A_i)$.

3. Let $A_0 = \varnothing$. Then,

$$\begin{aligned}\mu(\bigcup_{i=1}^{\infty} A_i) &= \mu(\bigcup_{i=1}^{\infty}(A_i\backslash A_{i-1}) \\ &= \sum_{i=1}^{\infty}\mu(A_i\backslash A_{i-1}) \\ &= \lim_{n\to\infty}\sum_{i=1}^{n}\mu(A_i\backslash A_{i-1}) \\ &= \lim_{n\to\infty}\mu(A_n)\end{aligned}$$

4. Let $B_i = A_1 \backslash A_i = A_1 \cap A_i^c \; \forall i \in \mathbb{N}$. Then, $B_1 \subseteq B_2 \subseteq ...$

$$\begin{aligned}
\bigcup_{i=1}^{\infty} B_i &= \bigcup_{i=1}^{\infty} (A_1 \cap A_i^c) \\
&= A_1 \cap \bigcup_{i=1}^{\infty} A_i^c \\
&= A_1 \cap (\bigcap_{i=1}^{\infty} A_i)^c \\
&= A_1 \backslash \bigcap_{i=1}^{\infty} A_i \Rightarrow \\
\mu(A_i) - \mu(\bigcap_{i=1}^{\infty} &= \mu(A_1 \backslash \bigcap_{i=1}^{\infty} A_i) \\
&= \mu(\bigcup_{i=1}^{\infty} B_i) \\
&= \lim n \to \infty \mu(B_n) \\
&= \lim_{n \to \infty} (\mu(A_1) - \mu(A_n)) \\
&= \mu(A_1) - \lim_{n \to \infty} \mu(A_n) \\
\therefore \mu(\bigcap_{i=1}^{\infty} &= \lim_{n \to \infty} \mu(A_n)
\end{aligned}$$

$\square$

1.2 Probability Measures

Definition 1.13

Let $(\Omega, \mathcal{F})$ be a measure space. Then, a probability measure P on $\mathcal{F}$ is a function such that:

(i) $P : \mathcal{F} \to [0, 1]$

(ii) $P(\Omega) = 1$

(iii) $\{A_i\}_{i\in\mathbb{N}} \subseteq \mathcal{F}, A_i \cap A_j = \varnothing \; \forall i \neq j \Rightarrow P(\dot{\bigcup}_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty}(A_i)$
(σ-additivity)

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

Ω is a sample space.

$\omega \in \Omega$ is a sample point.

If Ω is countable/finite, then $(\Omega, \mathcal{F}, P)$ is discrete/finite.

Any $A \in \mathcal{F}$ is an event.

If $A = \{\omega\}$, A is a simple event.

Otherwise, A is a compound event.

Remark 1.14

If $(\Omega, \mathcal{F}, P)$ is discrete, $f(\omega) := P(\{\omega\}), \omega \in \Omega$ defines P via $P(A) = \sum_{\omega\in A} f(\omega), A \in \mathcal{F}$. Then, f is the probability mass function (pmf) on Ω.

Conversely, if Ω is countable, then in $(\Omega, \mathcal{P}(\Omega), P)$ with $P(A) := \sum_{\omega\in A} f(\omega)$, $A \in \mathcal{P}(\Omega)$, P defines a discrete probability measure for any $f : \Omega \to [0, 1]$ such that $\sum_{\omega\in A} f(\omega) = 1$.

Proposition 1.15

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Then,

1. $A \in \mathcal{F} \Rightarrow P(A^c) = 1 - P(A)$
2. $A, B \in \mathcal{F} \Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$
3. Let $\{A_i\}_{i\in\mathbb{N}} \subseteq \mathcal{F}$
   $S_{k,n} := \sum_{1\leq i_1<...<i_k\leq n} P(A_{i_1} \cap ... \cap A_{i_k})$, for $k = 1, ..., n$

   Then, $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n}(-1)^{k-1} S_{k,n}$

   (inclusion-exclusion principle)

Proof.

1. $1 = P(\Omega) = P(A \dot{\cup} A^c) = P(A) + P(A^c)$

This implies probability measures are measures (take $A = \varnothing$).

2. $A \cup B = (A \backslash (A \cap B)) \dot{\cup} (A \cap B) \dot{\cup} (B \backslash (A \cap B)) \Rightarrow$

$$\begin{aligned} P(A \cup B) &= P(A) - P(A \cap B) + P(A \cap B) + P(B) - P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

3. Induction on n. See Exercise 6 in Assignment 1.

$\square$

1.3 Null Sets

Definition 1.16: Null Set

If $(\Omega, \mathcal{F}, \mu)$ is a measure space, every $N \in \mathcal{F}$ such that $\mu(N) = 0$ is a (μ-)null set.

If some property holds $\forall \omega \in \Omega \backslash N$ for null set N, it holds (μ-)almost everywhere.

Or, if μ is a probability measure, (μ-)almost surely.

If $\mathcal{F}$ contains all subsets of null sets, μ is complete.

By sub additivity, any countable union of null sets from $\mathcal{F}$ is a null set of $\mathcal{F}$.

Theorem 1.17: Completion

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, $\mathcal{N}$ be the set of all null sets.

1. $\bar{\mathcal{F}} := \{F \cup A : F \in \mathcal{F}, A \subseteq N, N \in \mathcal{N}\}$ is a σ-algebra on Ω.
2. $\bar{\mu}(F \cup A) = \mu(F)$ uniquely extends μ to a complete measure on $\bar{\mathcal{F}}$.

1.4 Construction of Measures

Idea: Functions with properties as measures (premeasures) defined on a ring can be extended to complete measures on the σ-algebra generated by the ring.

Definition 1.18: Rings

$\mathcal{R} \subseteq \mathcal{P}(\Omega)$ is a ring on Ω if:

1. $\varnothing \in \mathcal{R}$
2. $A, B \in \mathcal{R} \Rightarrow A \backslash B \in \mathcal{R}$
3. $A, B \in \mathcal{R} \Rightarrow A \cup B \in \mathcal{R}$

A premeasure μ_0 on $\mathcal{R}$ is a function with:

(i) $\mu_0 : \mathcal{R} \to [0, \infty]$

(ii) $\mu_0(\varnothing) = 0$

(iii) $\{A_i\}_{i \in \mathbb{N}} \subseteq \mathcal{R}, A_i \cap A_j = \varnothing \ \forall i \neq j, \dot{\bigcup}_{i=1}^{\infty} A_i \in \mathcal{R} \Rightarrow \mu(\dot{\bigcup}_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu_0(A_i)$

Theorem 1.19: Caratheodory's extension theorem

If μ_0 is a premeasure on ring $\mathcal{R}$ on Ω, there exists a complete measure μ on $\mathcal{F} := \sigma(\mathcal{R})$ which coincides with μ_0 on $\mathcal{R}$. If μ_0 is σ-finite, then μ is unique.

Remark 1.20

The proof of 1.19 is constructive. The measure μ is constructed:

$$\begin{aligned} \mu(A) &= \inf_{A \subseteq \bigcup_{i=1}^{\infty} A_i, A_i \in \mathcal{R} \ \forall i} \sum_{i=1}^{\infty} \mu_0(A_i) \\ &= A_i \in \mathcal{R} \ \forall i \end{aligned}$$

Theorem 1.21

If $F : \mathbb{R} \to \mathbb{R}$ is right-continuous and increasing ($F(x) \leq F(y) \ \forall x < y$), there exists exactly one Borel measure μ_F such that:

$\mu_F((a, b]) = F(b) - F(a) \ \forall a \leq b$

Proof. $\mathcal{R} := \{\dot{\bigcup}_{k=1}^{n} (a_k, b_k] : -\infty < a_k \leq b_k < \infty \ \forall k, n \in \mathbb{N}\}$ is a ring on $\mathbb{R}$, and $\mu_0(\dot{\bigcup}_{k=1}^{n} (a_k, b_k]) := \sum_{k=1}^{n} (F(b_k) - F(a_k))$ is a premeasure on $\mathcal{R}$. By 1.19, there exists exactly one measure μ_F on $\sigma(\mathbb{R}) = \mathcal{B}(\mathbb{R})$ such that $\mu_F|_{\mathcal{R}} = \mu_0$ (i.e. $\mu_F(A) = \mu_0(A) \ \forall A \in \mathbb{R}$). ☐

Remark 1.22

1. By 1.19, μ_F is complete, and called the Lebesgue-Stietjes measures associated to F. Its domain, the completion $\bar{\mathcal{B}}(\mathbb{R})$, the Lebesgue σ-algebra, can be shown to strictly contain $\mathcal{B}(\mathbb{R})$. Sets in $\bar{\mathcal{B}}(\mathbb{R}$ are called Lebesgue measurable or Lebesgue sets. By our construction,

$$\mu_F(A) = \inf_{A \subseteq \bigcup_{i=1}^{\infty} (a_i, b_i]} \sum_{i=1}^{\infty} \mu_F((a, b])$$

2. If $F(x) = x$, $\lambda := \mu_F$ is a Lebesgue measure on $\mathbb{R}$. Sets $N \subseteq \bar{\mathcal{B}}(\mathbb{R})$ that are null sets are Lebesgue null sets, $\lambda(N) = 0$.

   By 1.17 (1), $B \in \bar{\mathcal{B}}(\mathbb{R}) \Leftrightarrow B = A \cup N$.

Example 1.23

1. $\{x\} \subseteq \mathbb{R}$ is a null set for all $x \in \mathbb{R}$, since

$$\begin{aligned}\lambda(\{x\}) &= \lambda\big(\bigcap_{n=1}^{\infty}(x - \frac{1}{n}, x]\big)\\ &= \lim_{n\to\infty} \lambda\big((x - \frac{1}{n}, x]\big)\\ &= 0\end{aligned}$$

2. $\mathbb{Q} \subseteq \mathbb{R}$ is a null set since

$$\begin{aligned}\lambda(\mathbb{Q}) &= \lambda(\dot{\bigcup}_{i=1}^{\infty}\{q_i\})\\ &= \sum_{i=1}^{\infty} 0\\ &= 0\end{aligned}$$

3. Cantor set: $C = \bigcap_{i=1}^{\infty} C_i$ where C_i is defined by:

$$\begin{aligned}C_0 &= [0,1]\\ C_1 &= [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]\\ C_2 &= [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]\\ &\dots\\ C_i &= \frac{C_{i-1}}{3} \cup (\frac{2}{3} + \frac{C_{i-1}}{3})\ \forall i \geq 1\end{aligned}$$

   By Cantor's diagonal argument, C is uncountable. Since $\lambda([0,1] \backslash C) = 2^0\frac{1}{3} + 2^1\frac{1}{9} + \ldots = \sum_{i=1}^{\infty} 2^{i-1}3^{-i} = 1$, therefore $\lambda(C) = 0$.

Remark 1.24

1.21 extends to $F : \mathbb{R}^d \to \mathbb{R}$ which is:

(i) right continuous: $F(\underline{x}) = \lim_{\underline{h}\downarrow\underline{0}} F(\underline{x} + \underline{h}) =: F(\underline{x}+)\ \forall x \in \mathbb{R}^d$

(ii) d-increasing: The F-volume $\Delta_{(\underline{a},\underline{b}]}F$ of $(a,b] \geq 0$ for $\underline{a} \leq \underline{b}$, where:

$$\begin{aligned}\Delta_{(\underline{a},\underline{b}]}F :&= \sum_{i \in \{0,1\}^d} (-1)^{\sum_{j=1}^{d} ij} F(a_1^{i_1} b_1^{1-i_1}, ..., a_d^{i_d} b_d^{1-i_d}) \\ &= \prod_{j=1}^{d} (b_j - a_j) \\ &= \lambda((a,b])\end{aligned}$$

(iii) If, additionally, $\lim_{x_j \downarrow -\infty} F(\underline{x}) = 0$ for some $j \in \{1, ..., d\}$ and $F(\underline{\infty}) = lim_{\underline{x} \uparrow \infty} F(x) = 1$, then μ_F is a probability measure on $\mathcal{B}(\mathbb{R}^d)$. Then, $\Delta_{(a,b]}F$ is the probability of $(\underline{a}, \underline{b}]$.

2 Geometric and Laplace probability spaces

Proposition 2.1

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space such that $0 < \mu(\Omega) < \infty$. Then, $(\Omega, \mathcal{F}, P)$ with $P(A) = \frac{\mu(A)}{\mu(\Omega}$ $\forall A \in \mathcal{F}$, is a probability space.

Proof.

(i) $0 \leq \mu(A) \leq \mu(\Omega) \leq \infty \; \forall A \in \mathcal{F} \Rightarrow P : \mathcal{F} \to [0, 1]$

(ii) $P(\Omega) = \frac{P(\Omega)}{P(\Omega)} = 1$

(iii) $\{A_i\}_{i \in \mathbb{N}} \subseteq \mathcal{F}, A_i \cap A_j = \varnothing \; \forall i \neq j \Rightarrow$

$$\begin{aligned} P(\bigcup_{i=1}^{\infty} A_i) &= \frac{\mu(\bigcup_{i=1}^{\infty} A_i)}{\mu(\Omega} \\ &= \sum_{i=1}^{\infty} \frac{\mu(A_i)}{\mu(\Omega} \\ &= \sum_{i=1}^{\infty} P(A_i) \end{aligned}$$

□

If $\mathcal{F}$ is a σ-algebra on Ω and $\Omega' \subseteq \Omega$, one can show that the restriction $\mathcal{F}|_{\Omega'} := \{A \cap \Omega' : A \in \mathcal{F}\}$ is a σ-algebra on Ω'. This is called the trace σ-algebra of Ω' in $\mathcal{F}$.

2.1 Geometric Probability Spaces

Definition 2.2

If:

$$\begin{aligned} \Omega \subseteq \mathbb{R}^d &: 0 < \lambda(\Omega) < \infty \\ \mathcal{F} &= \mathcal{B}(\Omega) \\ P(A) &= \frac{\lambda(A)}{\lambda(\Omega)} \ \forall A \in \mathcal{F} \end{aligned}$$

then the probability space $(\Omega, \mathcal{F}, P)$ is a geometric probability space.

Example 2.3

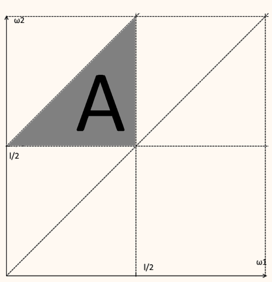
A stick length l is randomly marked and cut at 2 spots. Find the probability that the 3 pieces can form a triangle.

Solution. Let

$$\begin{aligned} \Omega &= \{(\omega_1, \omega_2) \in [0, l]^2, \omega_1 < \omega_2\} \\ \mathcal{F} &= \bar{\mathcal{B}}(\Omega) \\ P(A) &= \frac{\lambda(A)}{\lambda(\Omega)} \end{aligned}$$

We are interested in the set A: "each side $<$ sum of other 2", or

$$\begin{aligned} A =& \{(\omega_1, \omega_2) \in \Omega : \\ & \omega_1 < (\omega_2 - \omega_1) + (l - \omega_2), \\ & \omega_2 - \omega_1 < \omega_1 + l - \omega_2, \\ & l - \omega_2 < \omega_1 + \omega_2 - \omega_1\} \\ =& \{(\omega_1, \omega_2) \in \Omega : \omega_1 < \frac{l}{2}, \frac{l}{2} < \omega_2 < \omega_1 + \frac{l}{2}\} \end{aligned}$$



$P(A) = \frac{1}{4}$, from the picture.

2.2 Laplace Probability Spaces

Here is a similar construction, based on the number $|\Omega|$ of elements in Ω.

Proposition 2.4

Let $1 \leq |\Omega| < \infty$, $\mathcal{F} = \mathcal{P}(A), P(A) = \frac{|A|}{|\Omega|}$ $\forall A \in \mathcal{F}$. Then, $(\Omega, \mathcal{F}, P)$ is a finite probability space called a Laplace probability space.

P is discrete uniform distribution on Ω.

Proof. Apply Prop. 2.1 with $\mu(A) = |A|$ (counting measure). $\square$

Remark 2.5

For Laplace probability spaces, probability mass function on Ω is

$f(\omega) = P(\{\omega\}) = \frac{|\{\omega\}|}{|\Omega|} = \frac{1}{|\Omega|}$ $\forall \omega \in \Omega$
so the discrete uniform distribution assigns equal probability $\frac{1}{|\Omega|}$ to each $\omega \in \Omega$.

Example 2.6

1. Determine probability of obtaining 1 or 5 when rolling a fair, 6-sided die.

$$\begin{aligned} \Omega &= \{1, ..., 6\} \\ \mathcal{F} &= \mathcal{P}(\Omega) \\ P(A) &= \frac{|A|}{|\Omega|} \; \forall A \in \Omega \end{aligned}$$

   Let $A =$"rolling 1 or 5"$= \{1, 5\}$. Then, $P(A) = \frac{2}{6} = \frac{1}{3}$.

2. Determine probability of obtaining a sum of 2 and 7 when rolling the die twice.

$$\begin{aligned} \Omega &= \begin{matrix} \{(1,1), & \dots, & (1,6), \\ & \ddots & \\ (6,1), & \dots, & (6,6)\} \end{matrix} \\ \mathcal{F} &= \mathcal{P}(\Omega) \\ P(A) &= \frac{|A|}{|\Omega|} \; \forall A \in \mathcal{F} \end{aligned}$$

   So,

   $P(\text{“sum is 2”}) = \frac{1}{36}$

   $P(\text{“sum is 7”}) = \frac{6}{36} = \frac{1}{6}$

3. Determine probability of obtaining at least one 6 when rolling 3 times.

$$\begin{aligned} \Omega &= \{(\omega_1, \omega_2, \omega_3) | \omega_i \in \{1, ..., 6\} \; \forall i\} \\ \mathcal{F} &= \mathcal{P}(\Omega) \\ P(A) &= \frac{|A|}{|\Omega|} \; \forall A \in \mathcal{F} \end{aligned}$$

So, $P(\text{"at least one 6"}) = 1 - P(\text{"no 6s"}) = 1 - (\frac{5}{6})^3 = \frac{91}{216}$

3 Probability Counting Techniques

3.1 Basic Rules

Proposition 3.1

1. If $A_1, ..., A_n$ are pointwise disjoint finite sets, then $|\bigcup_{i=1}^{n} A_i| = \sum_{i=1}^{n} |A_i|$ (addition rule).
2. If $A_1, ..., A_n$ are finite sets, then $|\prod_{i=1}^{n} A_i| = \prod_{i=1}^{n} |A_i|$ (multiplication rule).

Proof. By induction. $\square$

Example 3.2

Consider an urn with 5 balls labelled 1, ..., 5. Determine the probability of obtaining precisely 1 even ball when drawing twice with replacement. Let $\mathbb{E}$ be the set of even integers.

$$\begin{aligned}
\Omega &= \begin{matrix} \{(1,1), & \dots, & (1,5), \\ & \ddots & \\ (5,1), & \dots, & (5,5)\} \end{matrix} \\
\mathcal{F} &= \mathcal{P}(\Omega) \\
A &= \{(\omega_1, \omega_2) \in \Omega : \omega_1 + \omega_2 \notin \mathbb{E}\} \\
&= A_1 \dot{\cup} A_2
\end{aligned}$$

where

$$\begin{aligned}
A_1 &= \{(\omega_1, \omega_2) : \omega_1 \in \mathbb{E}, \omega_2 \notin \mathbb{E}\} \\
A_2 &= \{(\omega_1, \omega_2) : \omega_1 \notin \mathbb{E}, \omega_2 \in \mathbb{E}\} \\
\therefore P(A) &= \frac{|A|}{|\Omega|} \\
&= \frac{|A_1| + |A_2|}{|\Omega|} \\
&= \frac{2 * 3 + 3 * 2}{25} \\
&= \frac{12}{25}
\end{aligned}$$

3.2 Urn Models

Many counting problems can be associated with drawing k balls from an urn with n balls. Classical models consider drawing:

(I) With order, with replacement

(II) With order, without replacement

(III) Without order, without replacement

(IV) Without order, with replacement

What are the number of possibilities in each of the four setups?

(I)

$$\begin{aligned}\Omega_I &= \{(\omega_1, ..., \omega_k) : \omega_i \in \{1, ..., n\}, i \in \{1, ..., k\}\} \\ &= \prod_{i=1}^{k} \{1, ..., n\} \\ &= \{1, ..., n\}^k \\ \Rightarrow |\Omega_I| &= n^k\end{aligned}$$

Example

(a) The number of 53 digit numbers containing only 0-1 is $2^{53} \approx 9 \cdot 10^{15}$.

(b) The number of functions from $A \to B$, where $|A| = k, |B| = n$ is n^k.

(II)

$$\begin{aligned}\Omega_{II} &= \{(\omega_1, ..., \omega_k) : \omega_i \in \{1, ..., n\}, \omega_i \neq \omega_j \forall i \neq j\} \\ \Rightarrow |\Omega_{II}| &= n(n-1)(n-2)...(n-k+1) \\ &=: (n)_k\end{aligned}$$

Where $(n)_k$ is the "falling factorial" and is read "n to k factors".

Example

(a) The number of 3 digit numbers with unique digits in $\{1, ..., 9\}$ is $(9)_3 = 9 \cdot 8 \cdot 7 = 504$.

(b) The number of injective functions from $A \to B$, where $|A| = k, |B| = n$ is $(n)_k$.

(c) If $k = n$, then $(n)_k = n!$, and $0! = 1! = 1$.

$$n! \sim \left(\frac{n}{e}\right)^n \sqrt{2\pi n}$$

(Stirling's formula)

(III)

$$\Omega_{III} = \{(\omega_1, ..., \omega_k) : \omega_i \in \{1, ..., n\} \forall i, \omega_1 < ... < \omega_k\}$$

Definition

An equivalence relation $\sim$ is a relation on some set S such that:

(i) $x \sim x$ for all $x \in S$. (reflexive)

(ii) $x \sim y \Leftrightarrow y \sim x$ for all $x, y \in S$. (symmetric)

(iii) $x \sim y$ and $y \sim z \Rightarrow x \sim z$ for all $x, y, z \in S$. (transitive)

An equivalence class of some $a \in S$ is $\{x \in S : a \sim x\}$.

Now, define an equivalence relation $\sim$ on Ω via $(\omega_1, ..., \omega_k) \sim (\omega'_1, ..., \omega'_k)$ iff there exists a permutation $\pi : \{1, ..., k\} \to \{1, ..., k\}$ such that
$(\omega_1, ..., \omega_k) \sim (\omega'_{\pi(1)}, ..., \omega'_{\pi(k)})$. Then, Ω_{II} consists of ordered representations of the equivalence classes of $\sim$, and each such class has n! elements. Thus, $|\Omega_{II}| = |\Omega_{III}|k!$, so $|\Omega_{III}| = \binom{n}{k}$.

Example

(a) Lotto 6/49 draws 6 from 49 without order and without replacement. Therefore, there are $\binom{49}{6}$ possible outcomes and the chance of some ticket winning is $\frac{1}{\binom{49}{6}} \approx 7.15 \cdot 10^{-8}$

(b) How many subsets of size k does a set of size n have? $\binom{n}{k}$. Therefore,

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n$$

Note that $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n!}{(n-k)!(n-(n-k))!} = \binom{n}{n-k}$.

(IV) Here, we can't simply use (I) and divide by k! (eg. If the 2^{nd} ball we draw equals (does not equal) the first, then the permutations don't need to be (do need to be) considered. Identify by distinguishable permutations of $n-1$ "1" and k "0".
$\Rightarrow |\Omega_{IV}| = \frac{number\ of\ n-1+k\ symbols}{(number\ of\ permutations\ of\ n-1\ \text{"1"}s)(number\ of\ permutations\ of\ k\ \text{"0"}s)} = \binom{n-1+k}{k}$

Formally, $\Omega_{IV} = \{(\omega_1, ..., \omega_k) : \omega_i \in \{1, ..., n\}\ \forall i, \omega_1 \leq ... \leq \omega_k\}$. Note that if $f(\omega_1, ..., \omega_k) = (\omega_1, \omega_2 + 1, ..., \omega_k + k - 1)$ is a bijection from Ω_{IV} to $\Omega'_{III} = \{(\omega_1, ..., \omega_k) : \omega_i \in \{1, ..., n+k-1\}\ \forall i, \omega_1 < ... < \omega_k\}$.
$\Rightarrow |\Omega_{IV}| = |\Omega'_{III}| = \binom{n+k-1}{k}$.

Example

(a) How many possible domino stones are there? A domino has two squares, each of which can be contain 0-6 dots.
$n = 7, k = 2 \Rightarrow \binom{n+k-1}{k} = 28$

(b) How many different partial derivatives $\frac{\partial^k}{\partial x_{jk} \ldots \partial x_{j1}} f$ of $f \in C^k(\mathbb{R}^n)$ exist?

By Scwartz' or Clairaut's theorem, order doesn't matter. Furthermore, we can differentiate with respect to same variable multiple times (so there is replacement).
$\Rightarrow \exists! \binom{n+k-1}{k}$ different partial derivatives.

4 Conditional Probability and Independence

Proposition 4.1

Let $(\Omega, \mathcal{F}, P)$ be a probability space, and $B \in \mathcal{F} : P(B) > 0$. Then, $P(A|B) := \frac{P(A \cap B)}{P(B)}, A \in \mathcal{F}$ is a probability measure on $(\Omega, \mathcal{F})$, called the ordinary conditional probability of A given B. (The vertical bar in the expression $P(A|B)$ means "given".)

Proof.

(i) Let $A \in \mathcal{F}$.

$$\begin{aligned} 0 &\leq P(A \cap B) \leq P(B) \\ &\Rightarrow \frac{P(A \cap B)}{P(B)} \leq 1 \\ &\Rightarrow P(.|B) : \mathcal{F} \to [0, 1] \end{aligned}$$

(ii) $P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$

(iii) If $\{A_i\}_{i \in \mathbb{N}} \in \mathcal{F}, A_i \cap A_j = \varnothing \; \forall i \neq j$, then

$$\begin{aligned} P(\bigcup_{i=1}^{\infty} A_i|B) &= \frac{P((\bigcup_{i=1}^{\infty} A_i) \cap B)}{P(B)} \\ &= \frac{\sum_{i=1}^{\infty} P(A_i \cap B)}{P(B)} \\ &= \sum_{i=1}^{\infty} \frac{P(A_i \cap B)}{P(B)} \\ &= \sum_{i=1}^{\infty} P(A_i|B) \end{aligned}$$

□

Although $P(A|B)$ is only defined if $P(B) > 0$, the convention $P(A|B)P(B) = P(A \cap B) \leq P(B)$ makes sense for any definition of $P(A|B) \in [0, 1]$ if $P(B) = 0$.

Theorem 4.2: Law of Total Probability

Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\{B_i\}_{i \in \mathbb{N}} \subseteq \mathcal{F}$ be a partition of Ω. That is,

$$\Omega = \dot{\bigcup}_{i=1}^{\infty} B_i, B_i \cap B_j = \varnothing \; \forall i \neq j$$

Then,

$$\begin{aligned} A \in \mathcal{F} \Rightarrow P(A) &= \sum_{i=1}^{\infty} P(A \cap B_i) \\ &= \sum_{i=1}^{\infty} P(A|B_i)P(B_i) \end{aligned}$$

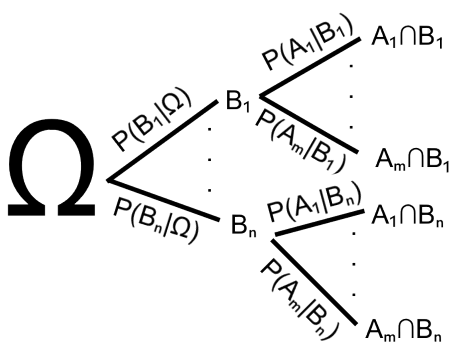
Proof.

$$
\begin{aligned}
A \in \mathcal{F} \Rightarrow P(A) &= P(A \cap \Omega) \\
&= P(A \cap \bigcup_{i=1}^{\infty} B_i) \\
&= P(\bigcup_{i=1}^{\infty} (A \cap B_i)) \\
&= \sum_{i=1}^{\infty} P(A \cap B_i) \\
&= \sum_{i=1}^{\infty} P(A|B_i)P(B_i)
\end{aligned}
$$

□

Remark 4.3

It is often helpful to visualize conditional probabilities in a tree diagram. For example, if $\{A_i\}_{i\in\mathbb{N}}^m, \{B_i\}_{i\in\mathbb{N}}^n \subseteq \mathcal{F}$ are permutations of $\mathcal{F}$,



$$
\begin{aligned}
P(A_1 \cap B_1) &= P(A_1|B_1)P(B_1) \\
P(A_m \cap B_1) &= P(A_m|B_1)P(B_1) \\
P(A_1 \cap B_n) &= P(A_1|B_n)P(B_n) \\
P(A_m \cap B_n) &= P(A_m|B_n)P(B_n)
\end{aligned}
$$

$$
\begin{aligned}
\sum_{i,j=1}^{m,n} P(A_i \cap B_j) &= \sum_{i,j=1}^{m,n} P(A_i|B_j)P(B_j) \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n} P(A_i|B_j)P(B_j) \\
&= P(\bigcup_{i=1}^{m} A) \\
&= P(\Omega) \\
&= 1
\end{aligned}
$$