An Analytical Introduction to Probability Theory

3. Probability Spaces

DASN, NTU https://personal.ntu.edu.sg/wptay/

3.1 Introduction

Recommended reference: "Probability: Theory and Examples" by Rick Durrett.

Let Ω be a sample space. An event is a subset of Ω. We are interested to define a "likelihood" or "chance" for each event to happen in the future. We call this the probability of the event.

Example 3.1. *Let $\Omega = [0,1]$, the probability of the event $(a,b]$, where $0 \le a \le b < 1$ can be defined by*

$$\mathbb{P}((a,b]) = F(b) - F(a),$$

where F is a non-decreasing and right-continuous (we will see later why this is needed) function with

$$\lim_{x\to 0} F(x) = 0,$$
$$\lim_{x\to 1} F(x) = 1.$$

However, there are many other events like $\bigcup_{i=1}^{\infty}(a_i, b_i]$ whose probabilities we are interested in. In particular, $\mathbb{P}$ should have the following properties:

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

(ii) If $A_1, A_2, \ldots$ are disjoint sets, then

$$\mathbb{P}\left(\bigcup_{i\ge 1} A_i\right) = \sum_{i\ge 1} \mathbb{P}(A_i).$$

(iii) If A is congruent to B (i.e., A is B transformed by translation, rotation or reflection), then $\mathbb{P}(A) = \mathbb{P}(B)$.

Unfortunately, for these conditions to hold for all *events would lead to inconsistency. To see why, define an equivalence $x \sim y$ iff $x - y$ is rational. Then Ω can be partitioned into equivalence classes. Let $N \subset \Omega$ be a subset that contains exactly one member of each* *equivalence class (we need the axiom of choice here). For each rational number $r \in \mathbb{Q} \cap [0,1)$, let*

$$N_r = \{x + r : x \in N \cap [0, 1-r)\} \cup \{x + r - 1 : x \in N \cap [1-r, 1]\},$$

i.e., N_r is N translated to the right by r with the part after $[0,1)$ shifted to the front (wrapped around) so that $N_r \subset \Omega = [0,1]$. From properties (ii) and (iii), we have for any rational $r \in \mathbb{Q} \cap [0,1)$,

$$\mathbb{P}(N) = \mathbb{P}(N \cap [0, 1-r)) + \mathbb{P}(N \cap [1-r, 1)) = \mathbb{P}(N_r). \tag{1}$$

We also have the following:

1. *Every $x \in \Omega$ belongs to a N_r because if $y \in N$ is an element of the equivalence class of x, then $x \in N_r$ where $r = x - y$ if $x \geq y$ or $r = x - y + 1$ if $x < y$.*
2. *Every $x \in \Omega$ belongs to exactly one N_r because if $x \in N_r \cap N_s$ for $r \neq s$, then $x - r$ or $x - r + 1$ and $x - s$ or $x - s + 1$ would be distinct elements of N belonging to the same equivalence class, contradicting how we chose N.*

Therefore, Ω is the disjoint union of N_r over all rational $r \in \mathbb{Q} \cap [0,1)$. From properties (i) and (ii), we also have $1 = \mathbb{P}(\Omega) = \sum_r \mathbb{P}(N_r)$. But $\mathbb{P}(N_r) = \mathbb{P}(N)$ from (1), so the sum is either 0 if $\mathbb{P}(N) = 0$ or ∞ if $\mathbb{P}(N) > 0$, a contradiction.

This example shows that it is impossible to define a suitable $\mathbb{P}$ for all possible events, some of which are very weird objects (Banach and Tarski (1924) showed that in $\mathbb{R}^n$ where $n \geq 3$, even stranger subsets can be constructed!). The solution that mathematicians have come up with is to restrict to a collection of subsets and *a $\mathbb{P}$ with "nice" properties, i.e., a σ-algebra and measure, respectively.*

3.2 σ-algebras and Measures

Let $\mathcal{A}$ be a collection of events (collection of subsets of Ω).

Definition 3.1. *$\mathcal{A}$ is an algebra if*

(i) $\Omega \in \mathcal{A}$.

(ii) $A \in \mathcal{A} \implies A^c = \Omega \setminus A \in \mathcal{A}$.

(iii) $A_1, A_2 \in \mathcal{A} \implies A_1 \cup A_2 \in \mathcal{A}$. *By induction,* $A_i \in \mathcal{A}, \forall\, i = 1, \ldots, n, \implies \bigcup_{i=1}^{n} A_i \in \mathcal{A}$.

Definition 3.2. *$\mathcal{A}$ is a σ-algebra or σ-field if* *(i)* $\Omega \in \mathcal{A}$.

(ii) $A \in \mathcal{A} \implies A^c = \Omega \setminus A \in \mathcal{A}$.

(iii) $A_i \in \mathcal{A}, \forall\, i = 1, 2, \dots \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

$(\Omega, \mathcal{A})$ is called a measurable space if $\mathcal{A}$ is a σ-algebra. A set $A \in \mathcal{A}$ is said to be measurable.

Definition 3.3. *For a measurable space $(\Omega, \mathcal{A})$, a function $\mathbb{P} : \mathcal{A} \mapsto [0, 1]$ is a probability measure if*

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

(ii) $A_1, A_2, \dots \in \mathcal{A}$ *with* $A_i \cap A_j = \emptyset$, $\forall\, i \neq j \implies \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ *(countably additive).*

Let $\mathcal{A}$ be a σ-algebra.

Lemma 3.1. *Suppose $B_i \in \mathcal{A}, B_i \subset B_{i+1}, \forall\, i \geq 1$, then*

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} B_i\right) = \lim_{i \to \infty} \mathbb{P}(B_i)$$

Proof. Let $C_1 = B_1, C_i = B_i \cap B_{i-1}^c, \forall\, i \geq 2$, then the C_i's are disjoint, and we have

$$B_n = \bigcup_{i=1}^{n} C_i,$$
$$\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} C_i.$$

Then we obtain

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} B_i\right) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} C_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(C_i) = \lim_{n \to \infty} \sum_{i=1}^{n} \mathbb{P}(C_i) = \lim_{n \to \infty} \mathbb{P}\left(\bigcup_{i=1}^{n} C_i\right) = \lim_{n \to \infty} \mathbb{P}(B_n).$$

□

Corollary 3.1. *For $A_i \in \mathcal{A}, \forall\ i = 1, 2, \dots$, we have*

$$\lim_{n \to \infty} \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right).$$

Proof. Let $B_n = \bigcup_{i=1}^{n} A_i$, which is an increasing sequence. We have $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right)$. From Lemma 3.1, we obtain $\mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty} \mathbb{P}(B_n) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right)$. □

Corollary 3.2. *For a decreasing sequence $B_i \supset B_{i+1}, \forall\ i \geq 1$, we have*

$$\mathbb{P}\left(\bigcap_{i=1}^{\infty} B_i\right) = \lim_{i\to\infty} \mathbb{P}(B_i).$$

Proof. Similar to the proof of Lemma 3.1. □

Lemma 3.2. *For $A, B \in \mathcal{A}, A \subset B$, we have $\mathbb{P}(A) \leq \mathbb{P}(B)$.*

Proof. $B = A \cup (B \setminus A) \implies \mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A)$. □

Lemma 3.3 (Union bound)**.** *For $A_1, A_2, \ldots \in \mathcal{A}$, we have*

$$\mathbb{P}\left(\bigcup_{i\geq 1} A_i\right) \leq \sum_{i\geq 1} \mathbb{P}(A_i).$$

Proof. Let $B_i = A_i \setminus \bigcup_{j<i} A_j$ for $i \geq 1$. Then the B_i's are disjoint, $B_i \subset A_i$, $\bigcup_{i\geq 1} A_i = \bigcup_{i\geq 1} B_i$ and

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{i\geq 1} A_i\right) &= \mathbb{P}\left(\bigcup_{i\geq 1} B_i\right) \\
&= \sum_{i\geq 1} \mathbb{P}(B_i) \\
&\leq \sum_{i\geq 1} \mathbb{P}(A_i).
\end{aligned}$$

□

In Example 3.1, let

$$\mathcal{A}' = \left\{ \bigcup_{i=1}^{n} (a_i, b_i] \,:\, n \geq 1, (a_i, b_i] \subset (0,1], (a_i, b_i] \cap (a_j, b_j] = \emptyset, \forall\ i \neq j \right\}.$$

Check that $\mathcal{A}'$ is an algebra. For each element of $\mathcal{A}'$, we define

$$\mathbb{P}\left(\bigcup_{i=1}^{n} (a_i, b_i]\right) = \sum_{i=1}^{n} \mathbb{P}((a_i, b_i])$$

for each $n \geq 1$. One can show that with this definition, $\mathbb{P}$ is countably additive on $\mathcal{A}'$, i.e., whenever $A_i \in \mathcal{A}'$, $i \geq 1$ and $\bigcup_{i\geq 1} A_i \in \mathcal{A}$ are finite unions of disjoint intervals, we have

$$\mathbb{P}\left(\bigcup_{i\geq 1}(a_i, b_i]\right) = \sum_{i\geq 1}\mathbb{P}((a_i, b_i]).$$

We also have

$$\begin{aligned}
\mathbb{P}((a,b]) &= \mathbb{P}\left(\bigcap_{n=1}^{\infty}(a, b+1/n]\right) \\
&= \lim_{n\to\infty}\mathbb{P}((a, b+1/n]) \\
&= \lim_{n\to\infty}(F(b+1/n) - F(a)) \\
&= F(b) - F(a),
\end{aligned}$$

where the second equality follows from Corollary 3.2 and the last equality requires the right-continuity of F.

Let $\mathcal{A} = \sigma(\mathcal{A}')$ be the σ-algebra generated by $\mathcal{A}'$, i.e., the intersection of all σ-algebras that contain $\mathcal{A}'$. Note that this is well-defined as a trivial σ-algebra containing $\mathcal{A}'$ is the power set of Ω. It is easy to show that $\mathcal{A}$ is the smallest σ-algebra containing $\mathcal{A}'$.

Theorem 3.1. *Carathéodory's Extension Theorem. If $\mathcal{A}'$ is an algebra, $\mathbb{P} : \mathcal{A}' \mapsto [0,1]$ is countably additive on $\mathcal{A}'$ and $\mathbb{P}(\emptyset) = 0$, then $\mathbb{P}$ has a unique extension to $\mathcal{A} = \sigma(\mathcal{A}')$.*

The proof of the existence of such an extension $\mathbb{P} : \mathcal{A} \mapsto [0,1]$ can be found in the book by Durrett. We focus on the proof of uniqueness here. We make use of the very useful Dynkin's π-λ Theorem.

Definition 3.4. *$\mathcal{P}$ is a π-system if $A, B \in \mathcal{P} \implies A \cap B \in \mathcal{P}$.*

Definition 3.5. *$\mathcal{L}$ is a λ-system if*

(i) $\Omega \in \mathcal{L}$.

(ii) $A \in \mathcal{L} \implies A^c = \Omega \setminus A \in \mathcal{L}$.

(iii) $A_i \in \mathcal{L}, \forall\ i \geq 1,\ A_i \cap A_j = \emptyset, \forall\ i \neq j \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{L}$.

Remark 3.1. *If $\mathcal{A}$ is both a π-system and λ-system, then $\mathcal{A}$ is σ-algebra.*

Theorem 3.2 (Dynkin's π-λ Theorem)**.** *If $\mathcal{P}$ is a π-system and $\mathcal{L}$ is a λ-system with $\mathcal{P} \subset \mathcal{L}$, then $\sigma(\mathcal{P}) \subset \mathcal{L}$.*

Proof. Let $\ell(\mathcal{P})$ be the smallest λ-system that contains $\mathcal{P}$. If $\ell(\mathcal{P})$ is a π-system, $\ell(\mathcal{P})$ is a σ-algebra. Then we have

$$\sigma(\mathcal{P}) \subset \ell(\mathcal{P}) \subset \mathcal{L}.$$

Therefore, it suffices to prove that $\ell(\mathcal{P})$ is a π-system. We prove that $\ell(\mathcal{P})$ is a π-system in the following three steps. For any $A \subset \Omega$, let $\mathcal{G}_A = \{B \subset \Omega : B \cap A \in \ell(\mathcal{P})\}$.

Step 1: We show that if $A \in \ell(\mathcal{P})$, then $\mathcal{G}_A$ is λ-system. This is done by checking the following conditions:

(i) $\Omega \cap A = A \in \ell(\mathcal{P}) \implies \Omega \in \mathcal{G}_A$.

(ii) Suppose $B \in \mathcal{G}_A$. We have $B^c \cap A = ((B \cap A) \cup A^c)^c$. Then we have

$$B \cap A,\ A^c \in \ell(\mathcal{P}) \implies (B \cap A) \cup A^c \in \ell(\mathcal{P}) \implies ((B \cap A) \cup A^c)^c \in \ell(\mathcal{P}),$$

which implies that $B^c \in \mathcal{G}_A$.

(iii) Let $B_i \in \mathcal{G}_A, \forall\ i = 1, 2, \ldots$ with $B_i \cap B_j = \emptyset, \forall\ i \neq j$. Therefore, $B_i \cap A \in \ell(\mathcal{P}), \forall\ i = 1, 2, \ldots$ are also disjoint. Then we have

$$\left(\bigcup_{i=1}^{\infty} B_i\right) \cap A = \bigcup_{i=1}^{\infty} (B_i \cap A) \in \ell(\mathcal{P}) \implies \bigcup_{i=1}^{\infty} B_i \in \mathcal{G}_A.$$

Step 2: We show that if $B \in \mathcal{P} \subset \ell(\mathcal{P})$, then $\ell(\mathcal{P}) \subset \mathcal{G}_B$. Since $\mathcal{P}$ is a π-system, we have

$$\forall\ C \in \mathcal{P}, C \cap B \in \mathcal{P} \subset \ell(\mathcal{P}),$$

which means that

$$\mathcal{P} \subset \mathcal{G}_B,$$

where $\mathcal{G}_B$ is a λ-system from Step 1 since $B \in \ell(\mathcal{P})$. Therefore,

$$\ell(\mathcal{P}) \subset \mathcal{G}_B. \tag{2}$$

Step 3: Consider any $B \in \mathcal{P}$ and an $A \in \ell(\mathcal{P})$. From Step 2, we have $A \in \mathcal{G}_B$. Therefore, $A \cap B \in \ell(\mathcal{P})$ and

$$\mathcal{P} \subset \mathcal{G}_A,$$

where $\mathcal{G}_A$ is a λ-system from Step 1. Thus $\ell(\mathcal{P}) \subset \mathcal{G}_A$. This means that for any $C \in \ell(\mathcal{P})$, we have

$$C \in \mathcal{G}_A \implies C \cap A \in \ell(\mathcal{P}).$$

Since $A, C \in \ell(\mathcal{P})$, the above result immediately shows that $\ell(\mathcal{P})$ is a π-system. $\square$

We now return to the uniqueness proof of Theorem 3.1. Suppose $\mathbb{P}_1$ and $\mathbb{P}_2$ are extensions of $\mathbb{P}$ with $\mathbb{P}_1(A) = \mathbb{P}_2(A), \forall\ A \in \mathcal{A}'$. Let

$$\mathcal{L} = \{A \in \mathcal{A} : \mathbb{P}_1(A) = \mathbb{P}_2(A)\}.$$

It is easy to see that $\mathcal{L}$ is a λ-system due to the properties of probability measures. $\mathcal{A}'$ is a π-system since it is an algebra, and $\mathcal{A}' \subset \mathcal{L}$ by definition. According to Theorem 3.2, $\mathcal{A} = \sigma(\mathcal{A}') \subset \mathcal{L}$. Thus, $\mathbb{P}_1 = \mathbb{P}_2$ on $\mathcal{A}$, meaning the extension is unique on $\mathcal{A}$.

3.3 Regularity

Definition 3.6. *Let (Ω, d) be a metric space. The Borel σ-algebra is a σ-algebra generated by the open sets of Ω (or equivalently, by the closed sets).*

Lemma 3.4. *Let $\mathcal{B}$ be the Borel σ-algebra. Then, $\forall\, A \in \mathcal{B}$,*

$$\mathbb{P}(A) = \sup\{\mathbb{P}(F)\ :\ F \subset A,\ F \textit{ is closed}\}. \tag{3}$$

Proof. Let

$$\mathcal{L} = \{A \in \mathcal{B}\ :\ \text{both } A \text{ and } A^c \text{ satisfy (3)}\}.$$

It can be checked that $\mathcal{L}$ is a λ-system (exercise). Let F be closed. It is obvious that F satisfies (3). Let $U = F^c$. We show that U also satisfies (3). To do this, since $\sup \mathbb{P}(C) \leq \mathbb{P}(U)$ for all closed $C \subset U$, it suffices to show that there is a sequence of closed subsets $F_n \subset U$ such that $\mathbb{P}(U) = \sup_n \mathbb{P}(F_n)$. To this end, for $n \geq 1$, let

$$F_n = \left\{\omega \in \Omega : \min_{x \in F} d(\omega, x) \geq 1/n\right\}.$$

Then we have $F_n \subset F_{n+1}$ and $U = \bigcup_{n=1}^{\infty} F_n$. From Lemma 3.1, we obtain

$$\mathbb{P}(U) = \lim_{n\to\infty} \mathbb{P}(F_n) = \sup_{n\geq 1} \mathbb{P}(F_n).$$

Therefore, U satisfies (3). As a consequence, $F \in \mathcal{L}$ for any closed F. Since $\mathcal{B}$ is generated by the closed sets, we have $\mathcal{B} \in \mathcal{L}$ and the proof is complete. □

A metric space is separable if it has a countable dense subset, i.e., $\exists\,\{x_n\}_{n=1}^{\infty}$ such that $\forall$ open $U \subset \Omega$, $x_i \in U$ for some x_i. Exercise: show that totally bounded implies separable. The converse is not true (e.g., consider the discrete metric space in Example 2.1).

Definition 3.7. *We say that a probability measure $\mathbb{P}$ for $(\Omega, \mathcal{B})$ is regular if*

$$\mathbb{P}(A) = \sup\{\mathbb{P}(K) : K \subset A, K \textit{ is compact}\}$$

for all $A \in \mathcal{B}$.

Theorem 3.3 (Ulam)**.** *If (Ω, d) is a complete separable space with Borel σ-algebra $\mathcal{B}$ and probability measure $\mathbb{P}$, then $\mathbb{P}$ is regular.*

Proof. Fix $\epsilon > 0$ and let $\{x_i\}_{i \geq 1}$ be a dense subset of Ω. Then for any $m \geq 1$, we have

$$\Omega = \bigcup_{i \geq 1} \overline{B}(x_i, 1/m),$$

where $\overline{B}(x_i, 1/m)$ is the closed ball of radius $1/m$ centered at x_i. Since $\mathbb{P}(\Omega) = 1$, there exists $n(m)$ sufficiently large so that

$$\mathbb{P}\left(\Omega \backslash \bigcup_{i=1}^{n(m)} \overline{B}(x_i, \frac{1}{m})\right) \leq \frac{\epsilon}{2^m}.$$

Let $K = \bigcap_{m \geq 1} \bigcup_{i=1}^{n(m)} \overline{B}(x_i, 1/m)$, which is closed and totally bounded. Since Ω is complete, K is also complete and hence compact. We then have

$$\begin{aligned}
\mathbb{P}(\Omega \backslash K) &= \mathbb{P}\left(\bigcap_{m \geq 1} \left(\Omega \backslash \bigcup_{i=1}^{n(m)} \overline{B}(x_i, 1/m)\right)\right) \\
&\leq \sum_{m \geq 1} \frac{\epsilon}{2^m} \\
&\leq \epsilon.
\end{aligned}$$

From Lemma 3.4, for any $A \in \mathcal{A}$, there exists a closed $F \subset A$ such that $\mathbb{P}(A \backslash F) \leq \epsilon$. Therefore,

$$\mathbb{P}(A \backslash (F \cap K)) \leq 2\epsilon,$$

and $F \cap K \subset A$ is a compact set. The theorem is now proved. □

3.4 Notes

(a) $([0,1], \mathcal{B}([0,1]), \lambda)$ is a probability space, where $\lambda((a,b)) = b - a$ is called the Lebesgue measure. This is the uniform distribution on $[0,1]$.

(b) Let $f : \mathbb{R} \mapsto \mathbb{R}_+$ be a function whose set of discontinuities has Lebesgue measure zero and $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$. Then $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P})$ where $\mathbb{P}(A) = \int_A f(x)\,\mathrm{d}x$, is a probability space. f is an example of a *probability density function* (pdf).

(c) Let $\mathcal{X}$ be a discrete set. Then $(\mathcal{X}, 2^{\mathcal{X}}, \mathbb{P})$ where $\mathbb{P}(\{x\}) = p(x)$ with $\sum_{x \in \mathcal{X}} p(x) = 1$, is a probability space. Here, $2^{\mathcal{X}}$ denotes the power set of $\mathcal{X}$, or the collection of all subsets of $\mathcal{X}$.

(d) There exist non-measureable sets, i.e., one cannot assign a measure to these sets without running into logical consistency issues (see Example 3.1 or Durrett). This is why the existence of probability spaces is non-trivial as we cannot simply define a measure over the power set 2^{Ω}.

(e) Exercise: If A is an algebra, then for any $B \in \sigma(A)$, $\exists\, B_n \in A$ such that

$$\lim_{n\to\infty} \mathbb{P}(B \triangle B_n) = 0,$$

where $B \triangle B_n = (B \cup B_n) \setminus (B \cap B_n)$ is the symmetric difference of two sets.