Probability and Stochastic Processes I
Lecture 1

Michael Evans
University of Toronto
https://utstat.utoronto.ca/mikevans/stac62/staC622023.html

2023

I.1 What is probability?

- let Ω be a set, called the *sample space*, and $\omega \in \Omega$, (ω is an element of Ω) called the *outcome* or *response*, is not known

- let $A \subseteq \Omega$ (A is a *subset* of Ω) called an *event* and it is desired to assess whether or not $\omega \in A$

- how?

- let 2^{Ω} be the *power set* of $\Omega =$ the set which consists of all subsets of Ω

- so an element of 2^{Ω} is a subset of Ω

- somehow we come up with a function $P : 2^{\Omega} \rightarrow [0.1]$ s.t. (such that) $P(A)$ measures our **belief** that $\omega \in A$ is true

- $P(A) = 0$ means it is known categorically that $\omega \in A$ is false and the closer $P(A)$ is to 0 the stronger is our belief that $\omega \in A$ is false

- $P(A) = 1$ means it is known categorically that $\omega \in A$ is true and the closer $P(A)$ is to 1 the stronger is our belief that $\omega \in A$ is true

- $P(A) = 1/2$ means there is no belief one way or the other as to the truth that $\omega \in A$, sometimes referred to as ignorance

Example I.1.1 - rolling a labelled symmetrical cube

- suppose we have a symmetric cube such that two sides are labelled 1, three sides are labelled 2 and one side is labelled 3

- the cube is rolled and the label ω on the face up is concealed and our concern is whether or not ω is odd

- so $\Omega = \{1, 2, 3\}$ and $A = \{1, 3\}$

- here $2^\Omega = \{\phi, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \Omega\}$ and $A \in 2^\Omega$

- ϕ is the set with no elements (the *null set*) and $\phi \subseteq \Omega$ always

- **note** - the cardinality (number of elements) of 2^{Ω} is $\#(2^{\Omega}) = 8 = 2^3 = 2^{\#(\Omega)}$ and the formula

$$\#(2^{\Omega}) = 2^{\#(\Omega)}$$

holds generally

- since the cube is symmetrical it seems reasonable to say that each face has the same weight in our belief about which face will be up

- as such it then seems reasonable that we assign

$$\begin{aligned} P(\{1\}) &= 2/6, = 1/3 \\ P(\{2\}) &= 3/6 = 1/2 \\ P(\{3\}) &= 1/6 \end{aligned}$$

- what about $P(A) = P(\{1,3\})$?

- a reasonable assignment is clearly

$$\begin{aligned} P(\{1,3\}) &= P(\{1\}) + P(\{3\}) = 1/3 + 1/6 = 1/2 \\ P(\{1,2\}) &= P(\{1\}) + P(\{2\}) = 1/3 + 1/2 = 5/6 \\ P(\{2,3\}) &= P(\{2\}) + P(\{3\}) = 1/2 + 1/6 = 2/3 \end{aligned}$$

and together with

$$\begin{aligned} P(\phi) &= 0 \\ P(\Omega) &= 1 \end{aligned}$$

this completes the definition of $P : 2^{\Omega} \rightarrow [0,1]$
- $P(\{1,3\}) = 1/2$ indicates we are ignorant as to whether or not the face up is odd
■ (end of example, proof or definition)

- the assignment of probability in the example was based on symmetry and counting and this works quite often to give a reasonable assignment

- in general suppose that Ω is a finite set and the context in question possesses a symmetry that leads to the assignment $P(\{\omega\}) = 1/\#(\Omega)$ for each element $\omega \in \Omega$

- then for $A \subseteq \Omega$ symmetry also suggests that $P(A) = \#(A)/\#(\Omega)$

- this counting definition implies that for $A, B \in 2^{\Omega}$ such that $A \cap B = \phi$

$$
\begin{aligned}
(i) \text{ (additive) } P(A \cup B) &= \frac{\#(A \cup B)}{\#(\Omega)} = \frac{\#(A) + \#(B)}{\#(\Omega)} \\
&= \frac{\#(A)}{\#(\Omega)} + \frac{\#(B)}{\#(\Omega)} = P(A) + P(B) \\
(ii) \text{ (normed) } P(\Omega) &= \frac{\#(\Omega)}{\#(\Omega)} = 1
\end{aligned}
$$

- any $P : 2^{\Omega} \rightarrow [0,1]$ satisfying (i) and (ii) is called a *probability measure* on Ω and when Ω is finite with $P(\{\omega\}) = 1/\#(\Omega)$ for each element $\omega \in \Omega$, then P is called the *uniform probability measure* on Ω

- **note** - the P defined in Example I.1.1 is not the uniform probability measure on $\Omega = \{1, 2, 3\}$ although it is derived from a uniform probability measure on the six faces of a symmetrical cube

- so one probability measure can be derived from another

- in this course it does not matter where the probability measure P comes from only that it is a function defined on a set of events into $[0, 1]$ that is additive and normed and we study the mathematical properties of such functions

- we want to give a definition of P for much more complicated sets Ω than just finite ones and for this to work we need to restrict the domain of P

Assume throughout these exercises that P is a probability measure defined on a finite Ω.

Exercise I.1.1 Give an argument that shows how P in Example I.1.1 is derived from a uniform probability measure.

Exercise I.1.2 Use induction to prove that if $A_1, \ldots, A_n \in 2^\Omega$ are mutually disjoint, then $P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$.

Exercise I.1.3 Prove that for $A \in 2^\Omega$, then $P(A^c) = 1 - P(A)$.

Exercise I.1.4 For $A, B \in 2^\Omega$ prove that
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Exercise I.1.5 Suppose that a roulette wheel is divided into 4 equal sectors labelled as 1,2,3 and 4 respectively. The wheel is spun and the sector where the wheel comes stops under the pointer is recorded. Identify $\omega, \Omega, 2^\Omega$ and a relevant P. What is the relevant P if the sector formerly labeled 4 is now labeled 3?

- consider sample spaces like

$$\begin{aligned}
\Omega &= \mathbb{R}^1 = \{\omega : -\infty < \omega < \infty\} \\
\Omega &= [0,1] = \{\omega : 0 \le \omega \le 1\} \\
\Omega &= \mathbb{R}^k = \mathbb{R}^1 \times \mathbb{R}^1 \times \cdots \times \mathbb{R}^1 \\
&= \{(\omega_1, \ldots, \omega_k) : \omega_i \in R^1, i = 1, \ldots, k\} \\
\Omega &= [0,1]^k = [0,1] \times [0,1] \times \cdots \times [0,1] \\
&= \{(\omega_1, \ldots, \omega_k) : \omega_i \in [0,1], i = 1, \ldots, k\}
\end{aligned}$$

which are all infinite sets, namely, $\#(\Omega) = \infty$

- to get "nice" probability measures on such sets we often have to restrict the domain of P to some subset of 2^Ω

Example I.2.1 *Uniform probability on* $[0,1]$

- would like such a P to satisfy $P([a,b]) = b - a$ for any $[a,b] \subseteq [0,1]$

- also would like P to be *countably additive:* if $A_1, A_2, \ldots$ are mutually disjoint subsets of $[0,1]$, then $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

- fact:there is no such P defined for every element of $2^{[0,1]}$ ■

- one general solution to this problem is to require only that the domain of P be a subset $\mathcal{A} \subseteq 2^{\Omega}$

- we want $\mathcal{A}$ closed under countable Boolean operations (intersection, union and complementation) so, for example if

$$\text{if } A_1, A_2, \ldots \in \mathcal{A} \text{ then } \cup_{i=1}^{\infty} A_i = \{\omega : \omega \in A_i \text{ for some } i\} \in \mathcal{A}$$
$$\text{if } A_1, A_2, \ldots \in \mathcal{A} \text{ then } \cap_{i=1}^{\infty} A_i = \{\omega : \omega \in A_i \text{ for all } i\} \in \mathcal{A}$$
$$\text{if } A \in \mathcal{A} \text{ then } A^c = \{\omega : \omega \notin A\} \in \mathcal{A}$$

Proposition I.2.1. (i) $(\cup_{i=1}^{\infty} A_i)^c = \cap_{i=1}^{\infty} A_i^c$ and (ii) $(\cap_{i=1}^{\infty} A_i)^c = \cup_{i=1}^{\infty} A_i^c$

Proof: (i) Let $\omega \in (\cup_{i=1}^{\infty} A_i)^c$. Then $\omega \notin \cup_{i=1}^{\infty} A_i$ and $\omega \notin A_i$ for all i and so $\omega \in A_i^c$ for all i, which implies $\omega \in \cap_{i=1}^{\infty} A_i^c$. Therefore $(\cup_{i=1}^{\infty} A_i)^c \subseteq \cap_{i=1}^{\infty} A_i^c$.
Now let $\omega \in \cap_{i=1}^{\infty} A_i^c$. Then $\omega \in A_i^c$ for all i, which implies $\omega \notin A_i$ for all i, which implies $\omega \notin \cup_{i=1}^{\infty} A_i$, which implies $\omega \in (\cup_{i=1}^{\infty} A_i)^c$. Therefore $\cap_{i=1}^{\infty} A_i^c \subseteq (\cup_{i=1}^{\infty} A_i)^c$ and conclude that (i) holds. ■

Exercise I.2.1 Prove Proposition I.2.1(ii).

Definition The set $\mathcal{A} \subseteq 2^{\Omega}$ is a *σ-algebra* (*σ-field*) on the set Ω if

(i) $\phi \in \mathcal{A}$,
(ii) if $A_1, A_2, \ldots \in \mathcal{A}$ *then* $\cup_{i=1}^{\infty} A_i \in \mathcal{A}$,
(iii) if $A \in \mathcal{A}$ *then* $A^c \in \mathcal{A}$. ■

Exercise I.2.2 Prove: if $A_1, A_2, \ldots \in \mathcal{A}$ where $\mathcal{A}$ is a *σ-algebra* then $\cap_{i=1}^{\infty} A_i \in \mathcal{A}$. Also prove that $\Omega \in \mathcal{A}$.

Exercise I.2.3 Prove: if $A_1, A_2, \ldots, A_n \in \mathcal{A}$ where $\mathcal{A}$ is a *σ-algebra* then $\cup_{i=1}^{n} A_i \in \mathcal{A}$ and $\cap_{i=1}^{n} A_i \in \mathcal{A}$.

Example 1.2.2

- clearly for any set Ω then 2^{Ω} is a σ-algebra on Ω called the *finest* σ-algebra on Ω

- also $\{\phi, \Omega\}$ is a σ-algebra on Ω called the *coarsest* σ-algebra on Ω

- also if $\mathcal{A}$ is a σ-algebra on Ω, then $\{\phi, \Omega\} \subseteq \mathcal{A} \subseteq 2^{\Omega}$ ■

Example 1.2.3

- suppose $\Omega = \{1, 2, 3, 4\}$

- then $\mathcal{A} = \{\phi, \{1, 2\}, \{3, 4\}, \Omega\}$ is a σ-algebra on Ω

- but $\mathcal{A} = \{\phi, \{1, 2\}, \{1, 3, 4\}, \Omega\}$ is not a σ-algebra on Ω since $\{1, 3, 4\}^c = \{2\} \notin \mathcal{A}$ and this violates condition (iii) ■

I.3 Probability Measures and Probability Models

- we can now give the formal definition of a probability measure P

Definition. A *probability measure* P defined on a set Ω with σ-algebra $\mathcal{A}$ is a function $P : \mathcal{A} \to [0, 1]$ that satisfies

(i) (normed) $P(\Omega) = 1$,
(ii) (countably additive) if $A_1, A_2, \ldots$ *are mutually disjoint elements of* $\mathcal{A}$, *then* $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

The triple $(\Omega, \mathcal{A}, P)$ is called a probability model. ■

Proposition I.3.1. If $(\Omega, \mathcal{A}, P)$ is a probability model, then $P(\phi) = 0$.
Proof: Let $A_i = \phi$ for $i = 1, 2, \ldots$ so $\phi = \cup_{i=1}^{\infty} A_i$ and the A_i are mutually disjoint. Suppose now that $P(\phi) > 0$ and we will obtain a contradiction. By countable additivity of P we have
$P(\phi) = \sum_{i=1}^{\infty} P(\phi) = \infty \cdot P(\phi) = \infty$. This contradicts $P(\phi) \in [0, 1]$ and so we must have $P(\phi) = 0$. ■

Example 1.3.1 *Uniform probability on a finite set Ω.*

- $(\Omega, 2^{\Omega}, P)$ where $P(A) = \#(A)/\#(\Omega)$ is additive

- now 2^{Ω} is a σ-algebra on Ω

- the only way for there to be infinitely many mutually disjoint $A_i \in 2^{\Omega}$ is for all but finitely many of the A_i to be equal to ϕ (2^{Ω} is a finite set)

- so since $\cup_{i=1}^{\infty} A_i = \cup_{\{i:A_i \neq \phi\}} A_i$ is a finite union, P is finitely additive and $P(\phi) = 0$, then

$$P(\cup_{i=1}^{\infty} A_i) = P(\cup_{\{i:A_i \neq \phi\}} A_i) = \sum_{\{i:A_i \neq \phi\}} P(A_i) = \sum_{i=1}^{\infty} P(A_i)$$

so P is countably additive and $P(\Omega) = \#(\Omega)/\#(\Omega) = 1$

- therefore P is a probability measure

Exercise I.3.1 For probability model $(\Omega, \mathcal{A}, P)$ and $A_1, A_2, \ldots, A_n \in \mathcal{A}$ mutually disjoint, prove that $P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$.

Exercise I.3.2 For probability model $(\Omega, \mathcal{A}, P)$ and $A, B \in \mathcal{A}$ s.t. $A \subseteq B$ prove that $P(A) \leq P(B)$.

Exercise I.3.3 For probability model $(\Omega, \mathcal{A}, P)$ and $A \in \mathcal{A}$ prove that $P(A^c) = 1 - P(A)$.

Exercise I.3.4 Let $\Omega = \{1, 2, 3, 4\}$ with $\mathcal{A} = \{\phi, \{1, 2\}, \{3, 4\}, \Omega\}$ and P defined by $P(\phi) =0, P(\{1, 2\}) = 1/3, P(\{3, 4\}) = 2/3$ and $P(\Omega) = 1$. Is $(\Omega, \mathcal{A}, P)$ a probability model? Why or why not?