LESSON 4

Probability: Sample Space, Events, Classical And Axiomatic Definitions of Probability, Properties of Probability Measures

What is probability?

It is a mathematical theory of "probabilistic" uncertainty.

Remark: There are other types of uncertainty that

Cannot be modeled using probability.

Example: Make a random number generator select $x \in [0,1]$: $\star$ $\boxed{}$

what is the probability of selecting $\star$? 0

Is it probable to select ☆? no

Is it possible to select ☆? yes

probable $\longrightarrow$ possible

possible $\not\longrightarrow$ probable

because $\mathbb{P}$(Something possible) may be 0.

Possibility is modeled by Set Theory and extensions such as fuzzy set theory

It is NOT AN EASY THEORY.

Example: You have three decks of 52 cards

blue	blue	red
blue	red	red

You pick a card and observe that one side of the card is blue. What is the probability that the other side is also blue? $= 1/2$ $< 1/2$ $> 1/2$

Classical Definition of probability:

Ratio of favorable outcomes and the total number of possible outcomes, if all outcomes are equally likely

$$\mathbb{P}(A) = \frac{n_A}{N}$$

Example: we have five balls, one is red, the rest are blue.

$$\mathbb{P}(\text{blue}) = \frac{4}{5}$$

Relative Frequency Definition:

$$\mathbb{P}(A) = \lim_{N \to \infty} \frac{n_A}{N}$$

Example: We flip a coin 10,000,000 times.

$$\mathbb{P}(\text{Heads}) \approx \frac{\#\ \text{Heads}}{10^7}$$

Axiomatic Definition:

A countably additive function defined on the set of events with a range $[0,1]$.

Sample Space:

Set of all possible outcomes of an experiment is called the sample Space, Ω

Example: Coin toss

(a) Interested in knowing whether the toss produces a head or a tail $\Omega = \{H, T\}$

(b) Interested in the number of tumbles before the coin hits the ground $\Omega = \{0, 1, 2, \cdots\}$

$= \mathbb{Z}^{\geq 0}$

- Interested in the speed with which the coin hits the ground $\Omega = \mathbb{R}$ or $[0, c]$

Thus,
For the same experiment, Ω can be different, depending on what the observer is interested in.

Example: (a) Roll of a die

$$\Omega = \{1,2,3,4,5,6\}$$

(b) Tossing two coins

$$\Omega = \{H,T\} \times \{H,T\} =$$

$$\{(H,H),(H,T),(T,H),(T,T)\}$$

Example: What is Ω when a coin is tossed until heads appear?

$\Omega = \{ H, (T,H), (T,T,H), \dots$

all sequences of H,T that end with H $\}$

Example: Life expectancy of a random person

$\Omega = [0, C]$

Def (Informal): An event is a subset of the sample space Ω, to which probabilities will be assigned

Def (Informal): An event is said to have occured if the outcome of the experiment (ω) belongs to it.

Example: Coin toss $\Omega = \{H, T\}$

$E_1 = \phi \quad E_2 = \{H\}, \; E_3 = \{T\}, \; E_4 = \{H, T\} = \Omega$

Roll of a die : $\Omega = \{1, 2, 3, 4, 5, 6\}$

$E = \{2, 4, 6\} \doteq$ seeing an even number

Life expectancy $\Omega = [0, 120]$

$E = [50, 120] \doteq$ person lives more than 50 years

Assignment of probability to events:

nothing occured, impossible event

$\mathbb{P}(\emptyset) = 0$

$\mathbb{P}(H) = p \qquad \mathbb{P}(T) = 1 - p$

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

something occured = sure event

$\Omega = \{1, 2, \ldots, 6\}$

Roll of a die

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{6} \qquad \forall A \subseteq \Omega$$

Question: Can we always assign probabilities to all subsets of Ω?

- Yes, if Ω is countable

- No, if Ω is uncountable
  - Handout on Vitali sets (non-measurable)
  - " Banach-Tarski Paradox

Relation between set theory and probability

We use the algebra of sets to model events

A or B occured $\boxed{A \cup B}$

A and B occured $\boxed{A \cap B}$

A occured, but B didn't occur

$$\boxed{A \cap B^c = A \setminus B}$$

To model events, we need some structure that includes all possible events

σ-field of events

Def: A (non-empty) subset of 2^{Ω}, $\mathcal{F} \subseteq 2^{\Omega}$

is called a σ-field (or σ-algebra)

iff

(1) $\phi \in \mathcal{F}$ (redundant)

$\phi \in \mathcal{F}, \Omega \in \mathcal{F}$

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

closure under complement

(3) $A_1, A_2, \cdots \in \mathcal{F} \Rightarrow \bigcup_{i \in I} A_i \in \mathcal{F}$

closure under countable union

This structure allows us to model events like

$A^c \cup (B^c \cap C^c)$

Any $A \in \mathcal{F}$ is called event or an $\mathcal{F}$-measurable set

Example: The smallest σ-field for any Ω

$\mathcal{F} = \{\phi, \Omega\}$

Example: $A \subseteq \Omega$ $\quad \mathcal{F} = \{\phi, A, A^c, \Omega\}$

smallest sigma field generated by $A \subseteq \Omega$

Example: is 2^Ω a σ-field?

yes

Exercise: Prove that if $A_1, A_2, \cdots \in \mathcal{F}$ then the countable intersection of A_i's $\bigcap_{i \in I} A_i \in \mathcal{F}$

$$A_i \in \mathcal{F} \Rightarrow A_i^c \in \mathcal{F}$$
$$\Rightarrow \bigcup_{i \in I} A_i^c \in \mathcal{F} \Rightarrow \left(\bigcup_{i \in I} A_i^c\right)^c \in \mathcal{F}$$
$$\Rightarrow \bigcap_{i \in I} A_i \in \mathcal{F}$$

Exercise: using axioms (2), (3) show that ϕ, Ω also in $\mathcal{F}$

$$A_i = \begin{cases} A & i \in \text{odd} \\ A^c & i \in \# \end{cases}$$

$$A \in \mathcal{F}, A^c \in \mathcal{F} \Rightarrow A_i \in \mathcal{F}, \forall i \in \mathbb{N}$$
$$\Rightarrow \bigcup_{i \in N} A_i \in \mathcal{F} = A \cup A^c \cup A \cup A^c \cdots = \Omega$$
$$\Rightarrow \Omega \in \mathcal{F} \Rightarrow \Omega^c \in \mathcal{F} \Rightarrow \phi \in \mathcal{F}$$

Lemma: if $A_1 A_2 \cdots A_n$ is a finite sequence of event $\bigcup_{i=1}^{n} A_i \in \mathcal{F}$

define $B_i = \begin{cases} A_i & 1 \le i \le n \\ \phi & i > n \end{cases}$

$$\bigcup_{i \in N} B_i \in \mathcal{F} \quad \because B_i \in \mathcal{F}$$
$$\Rightarrow \bigcup_{i=1}^{n} A_i \in \mathcal{F}$$

Probability Measure (Kolmogorov)

Assume that Ω is a sample space and $\mathcal{F}$ is a σ-field of subsets of Ω, i.e. $\mathcal{F} \subseteq 2^{\Omega}$.

A probability measure on measurable space $(\Omega, \mathcal{F})$ is a function

$$\mathbb{P} : \underset{\text{set}}{\mathcal{F}} \longrightarrow [0,1]$$

s.t.

(1) Probability of Ω

$P(\Omega)=1$

(2) Probability of countable unions of disjoint events (Countable additivity property)

$A_1, A_2, \cdots \in \mathcal{F}$. $\{A_i\}_{i\in\mathbb{N}}$

$A_i \cap A_j = \phi \quad i \neq j$ mutually exclusive : disjoint

$$P\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Alternatively $\mathbb{P} : \mathcal{F} \to \mathbb{R}$

(1) Probability of $\emptyset$

$P(\phi) = 0$

(2) Probability of Ω

$P(\Omega) = 1$

(3) Countable additivity

We call $(\Omega, \mathcal{F}, \mathbb{P})$

a probability space or a

probability model or a probability triplet

Aside: $\mathbb{P}$ is a special case

of an additive measure:

$$\mu : \mathcal{F} \rightarrow [0, +\infty)$$

null preserving $\mu(\emptyset) = 0$

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \boxed{\sum_{i=1}^{\infty} \mu(A_i)}$$

when A_i's are mutually exclusive

Examples:

Coin toss: $\Omega = \{H, T\}$

$\mathcal{F} = \{\phi, \{H\}, \{T\}, \Omega\}$

$\mathbb{P}(\emptyset) = 0$ $\mathbb{P}(\{H\}) = p \in (0,1)$

$\mathbb{P}(\{T\}) = 1-p$ $\mathbb{P}(\Omega) = 1$

Roll of a die

$\Omega = \{1, 2, \cdots, 6\}$

$\mathcal{F} = 2^{\Omega}$

$\mathbb{P}(\{i\}) = p_i$, $\sum_{i=1}^{6} p_i = 1$

$\mathbb{P}(A) = \sum_{i \in A} p_i$ e.g.) $P(\{2, 3, 6\})$

$\forall A \subseteq \Omega$ $= p_2 + p_3 + p_6$

Lemma

(1) Additivity for a finite number of events

$A_1, \ldots, A_n \in \mathcal{F}$, $A_i \cap A_j = \emptyset$ $\forall i \neq j$

$\Rightarrow \mathbb{P}(\bigcup_{i=1}^{n} A_i) = \boxed{\sum_{i=1}^{n} P(A_i)}$

(2) Probability of complement

$A \in \mathcal{F} \Rightarrow \mathbb{P}(A^c) = 1 - P(A)$

(3) Preservation of order
monetonocity

$A \subseteq B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$

(4) Boole's Inequality (Union Bound) subadditivity of probability measure

$A_1, A_2, \cdots \in \mathcal{F}$

$$\Rightarrow \quad \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

(5) Inclusion-Exclusion

$A_1, A_2 \in \mathcal{F}$

$$\mathbb{P}(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

(6) Inclusion-Exclusion

(General form)

$A_1, A_2, \cdots, A_n \in \mathcal{F}$ not necessarly disjoint

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) + \cdots + (-1)^{n+1} P\left(\bigcap_{i=1}^{n} A_i\right)$$

for $n=3$

$$\mathbb{P}(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - \{P(A_1 \cap A_2) + P(A_1 \cap A_3) + P(A_2 \cap A_3)\} + P(A_1 \cap A_2 \cap A_3)$$

for $n=4$

$$\mathbb{P}(A_1 \cup A_2 \cup A_3 \cup A_4) =$$

Proof.

(1) Finite sequence of disjoint events $A_1, \dots, A_n \in F$

To use axioms of probability, we create an infinite sequence of disjoint events

$$B_i = \begin{cases} A_i & 0 \le i \le n \\ \phi & i > n \end{cases}$$

, B_i's are mutually exclusive

$$\bigcup_{i \in \mathbb{N}} B_i = \bigcup_{i=1}^{n} A_i$$

$$\Rightarrow P\left(\bigcup_{i=1}^{n} A_i\right) = P\left(\bigcup_{i \in \mathbb{N}} B_i\right) = \sum_{i=1}^{\infty} P(B_i) = \sum_{i=1}^{n} P(A_i) + \sum_{i=n+1}^{\infty} P(\phi)$$

$$= \sum_{i=1}^{n} P(A_i) \quad \text{Q.E.D.}$$

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

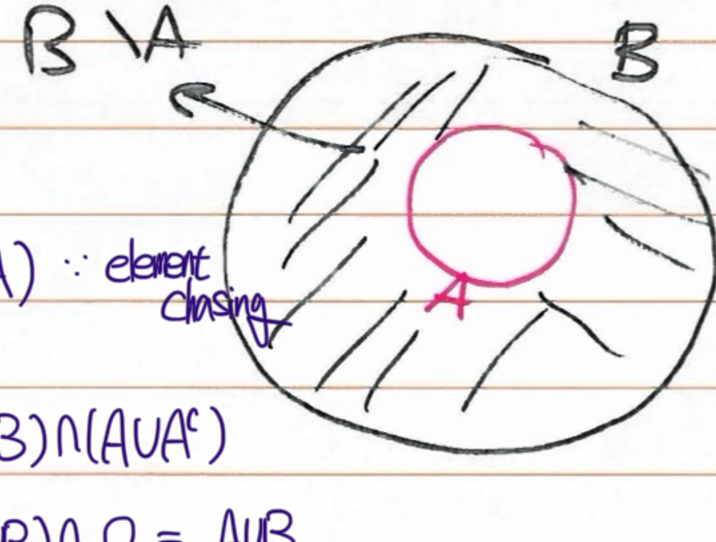
Note that

$A \cup A^c = \Omega \qquad A \cap A^c = \phi$

$\mathbb{P}(A \cup A^c) = P(A) + P(A^c) - P(A \cap A^c) = P(\Omega) = 1$

$\therefore \; P(A^c) = 1 - P(A)$

(3) $A \subseteq B \qquad \mathbb{P}(A) \leq \mathbb{P}(B)$



$B = A \cup (B \setminus A)$ $\because$ element chasing

$A \cup (B \cap A^c) = (A \cup B) \cap (A \cup A^c)$

$= (A \cup B) \cap \Omega = A \cup B$

$A \cap (B \setminus A) = \phi \qquad \therefore$ A and $B \setminus A$ are disjoint

$$P(B) = P(A \cup (B \setminus A)) = P(A) + P(B \setminus A)$$

$$\Rightarrow P(A) \leq P(B) \quad \because P(B \setminus A) \geq 0$$

also $\quad P(B \setminus A) = P(B) - P(A)$

Important Corollary:

$$A \subseteq B \Rightarrow \mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$$

(4) Boole's Inequality

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \le \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

Union bound / Subadditivity of the probability measure

define a sequence of event

$B_1 = A_1$

$B_2 = A_2 \setminus A_1$

$B_3 = A_3 \setminus (A_1 \cup A_2)$

$B_4 = A_4 \setminus (A_1 \cup A_2 \cup A_3)$

$\vdots$

$B_j = A_j \setminus \left(\bigcup_{i=1}^{j-1} A_i\right)$

Claim B_i's are mutually exclusive (disjoint)

$B_1 \cup B_2 = A_1 \cup A_2$

$B_1 \cup B_2 \cup B_3 = A_1 \cup A_2 \cup A_3$

$\Rightarrow \bigcup_{i=1}^{n} B_i = \bigcup_{i=1}^{n} A_i$ ← why?

$\bigcup_{i \in \mathbb{N}} B_i = \bigcup_{i \in \mathbb{N}} A_i$

$$\mathbb{P}\left(\bigcup_{i \in \mathbb{N}} A_i\right) = P\left(\bigcup_{i \in \mathbb{N}} B_i\right) = \sum_{i=1}^{\infty} P(B_i)$$

Claim $\quad B_i \subseteq A_i \Rightarrow P(B_i) \le P(A_i)$

$$\Rightarrow \sum_{i=1}^{\infty} P(B_i) \le \sum_{i=1}^{\infty} P(A_i)$$

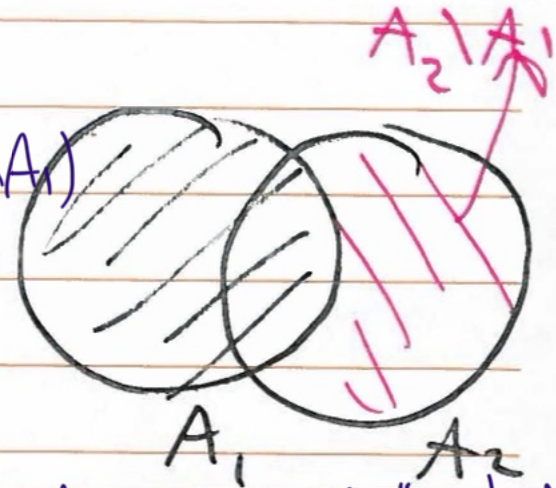
$$\Rightarrow P\left(\bigcup_{i \in \mathbb{N}} A_i\right) \le \sum_{i=1}^{\infty} P(A_i)$$

Q.E.D.

(5) Inclusion Exclusion

$$\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2)$$

$A_1 \cup A_2 = A_1 \cup (A_2 \setminus A_1)$



Prove that A_1 and $A_2 \setminus A_1$ are mutually disjoint

$$P(A_1 \cup A_2) = P(A_1 \cup (A_2 \setminus A_1))$$
$$= P(A_1) + P(A_2 \setminus A_1)$$

claim $A_2 \setminus A_1 = A_2 \setminus (A_1 \cap A_2)$

$$\therefore \; P(A_1 \cup A_2) = P(A_1) + P(A_2 \setminus (A_1 \cap A_2))$$

because $A_1 \cap A_2 \subseteq A_2$

$$= P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

$$\Rightarrow \; P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \quad \text{Q.E.D.}$$

- How can we prove the general case?

By mathematical induction

Let us review the concept of mathematical induction

Principle of Mathematical Induction

Theorem: Let $p(n)$ be a predicate, $n \in \mathbb{N}$. Assume that:

(a) $p(m_0)$ is true for some $m_0 \in \mathbb{N}$

(b) If $p(k)$ is true for some $k \geq m_0$ then $p(k+1)$ is also true

Then $p(n)$ is true for all $n \geq m_0$, $n \in \mathbb{N}$

Example: Prove that

$$S(n): 1 + 3 + 5 + \cdots + (2n-1) = n^2 \quad \forall n \in \mathbb{N}$$

Induction on n

$n = 1 \Rightarrow 1 = 1$ true ✓

Assume for $n=k$

$S(k): 1+3+5+\cdots+(2k-1)=k^2$. is true

To use induction, we must

show that $S(k+1)$ is also true, i.e.

$$S(k+1): 1+3+5+\cdots+(2k-1)+(2k+1)$$
$$= k^2+2k+1 = (k+1)^2$$

$\Rightarrow \quad S(k+1): 1+3+5+\cdots+(2k-1)+(2k+1)=(k+1)^2$

$\Rightarrow \quad S(n)$ is true for $\forall n \in \mathbb{N}$

Back to proving inclusion-exclusion

$n=2$ $\quad \mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2)$

$- \mathbb{P}(A_1 \cap A_2)$

Assume that

$n=l$: $\mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_l)$

$$= \sum_{i=1}^{l} \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j)$$

$$+ \sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k) - \cdots$$

$$+ (-1)^{l+1} \mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_l)$$

We must prove that for

$n = l+1$:

$n = l+1$ $\quad \mathbb{P}\left(\bigcup_{i=1}^{l+1} A_i\right) = P\left(\bigcup_{i=1}^{l} A_i \cup A_{l+1}\right)$

(It is a good exercise in applying probability properties to solving problems)

$$P\left(\bigcup_{i=1}^{l+1} A_i\right) = P\left(\bigcup_{i=1}^{l} A_i \cup A_{l+1}\right)$$

by the basic inclusion exclusion

$$= P\left(\bigcup_{i=1}^{l} A_i\right) + P(A_{l+1}) - P\left(\left[\bigcup_{i=1}^{l} A_i\right] \cap A_{l+1}\right) \star$$

$$\star = P\left(\bigcup_{i=1}^{l} [A_i \cap A_{l+1}]\right)$$

$$= \sum_{i=1}^{l} P(A_i \cap A_{l+1}) - \sum_{i<j} P(A_i \cap A_j \cap A_{l+1}) + \cdots + (-1)^{l+1} P\left(\bigcap_{i=1}^{l} A_l \cap A_{l+1}\right)$$

If you add $\star$ to all term $P(A_1 \cup A_2 \cdots \cup A_{l+1})$, you will get inclusive exclusion for $l+1$

Continuity of Probability Measure

Lemma: Assume that

$\forall i \in \mathbb{N}$, $A_i \in \mathcal{F}$ (A_i is an event, or is $\mathcal{F}$-measurable) then

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

This lemma defines a sequence of numbers $p_n = \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right)$ and states that the limit of the sequence p_n is equivalent to the probability $\mathbb{P}\left(\bigcup_{i \in \mathbb{N}} A_i\right)$.

Aside: Real Sequences and their limits

A sequence $\{a_n\}_{n=1}^{\infty}$ is a function from $\mathbb{N}$ to another set, therefore a real sequence is $a_n : \mathbb{N} \to \mathbb{R}$

Def: limit of a sequence

We say a_n converges to L as $n \to \infty$ and write

$$\lim_{n \to \infty} a_n = L$$

iff: $\forall \varepsilon > 0, \exists N_0 \in \mathbb{N}$ s.t.

$n > N_0 \Rightarrow |a_n - L| < \varepsilon$

Example: $\lim_{n \to \infty} 1/n = 0$

Should prove that

$\forall \varepsilon > 0 \ \exists N_0$ s.t. $n > N_0$

$|\frac{1}{n} - 0| < \varepsilon$

$\Rightarrow \frac{1}{n} < \varepsilon$

$\Rightarrow n > \frac{1}{\varepsilon}$

i.g.) $\Rightarrow N_0 = \lfloor \frac{1}{\varepsilon} \rfloor + 1$

i.g.) $\varepsilon = 10^{-6}$

$N_0 = \lfloor \frac{1}{10^{-6}} \rfloor + 1 = 10^6 + 1$

$$a_{N_0} = \frac{1}{10^6+1} < 10^{-6}$$

$$a_{N+1} = \frac{1}{10^6+2} < 10^{-6}$$

Back To Continuity of Probability

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

Proof:

define $B_1 = A_1$

$B_2 = A_2 \setminus A_1$ $\qquad B_i \cap B_j = \phi \quad i \neq j$

$B_3 = A_3 \setminus (A_1 \cup A_2)$

$$\bigcup_{i=1}^{n} B_i = \bigcup_{i=1}^{n} A_i \qquad \bigcup_{i \in \mathbb{N}} B_i = \bigcup_{i \in \mathbb{N}} A_i$$

$$P\left(\bigcup_{i \in \mathbb{N}} A_i\right) = P\left(\bigcup_{i \in \mathbb{N}} B_i\right) = \sum_{i=1}^{\infty} P(B_i) \quad \because \text{ by countable additivity}$$

$$= \lim_{n \to \infty} \sum_{i=1}^{n} P(B_i)$$

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

Q.E.D.

Corollary :

If $A_i^{\ i\in\mathbb{N}}$'s is a sequence of nested events:

(a). If A_i is increasing, sequence of events i.e.

$A_i \subseteq A_{i+1} \quad \forall i \in \mathbb{N}$, then

$$\boxed{\mathbb{P}\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \lim_{n\to\infty} \mathbb{P}(A_n)}$$

(b) If A_i is decreasing, i.e.

$A_i \supseteq A_{i+1} \quad \forall i \in \mathbb{N}$, then

$$\boxed{\mathbb{P}\left(\bigcap_{i\in\mathbb{N}} A_i\right) = \lim_{n\to\infty} \mathbb{P}(A_n)}$$

Proof

a) from continuity of probability

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

if $A_i \subseteq A_{i+1}$

$$\bigcup_{i=1}^{n} A_i = A_n \Rightarrow P\left(\bigcup_{i\in N} A_i\right) = \lim_{n\to\infty} P(A_n)$$

$$= P\left(\lim_{i\to\infty} A_i\right) = \lim_{i\to\infty} P(A_i)$$

$\therefore \lim_{x\to a} f(x) = f\left(\lim_{x\to a} x\right)$; continuous function

b) if $A_i \supseteq A_{i+1} \Rightarrow A_i^c \subseteq A_{i+1}^c$

by (a)

$$P\left(\bigcup_{i\in N} A_i^c\right) = \lim_{n\to\infty} P(A_n^c)$$

$$= P\left(\left[\bigcap_{i\in N} A_i\right]^c\right) = \lim_{n\to\infty} 1 - P(A_n)$$

$$1 - P\left(\bigcap_{i\in N} A_i\right) = 1 - \lim_{n\to\infty} P(A_n)$$

Why is it called continuity?

$$\lim_{i \to \infty} \mathbb{P}(A_i) = \mathbb{P}(\lim_{i \to \infty} A_i)$$

The lim operator passes probability, like continuous functions:

$$\lim_{x \to x_0} f(x) = f(\lim_{x \to x_0} x) = f(x_0)$$

Remark(Zero Probability and Almost Sure Events)

$(\Omega, \mathcal{F}, \mathbb{P})$: Probability Space

If $A \in \mathcal{F}$, $\mathbb{P}(A) = 0$,

Can we conclude that $A = \emptyset$?

No! ; select $\{1/2\}$ from $[0,1]$

A is called a zero probability event

If $A \in \mathcal{F}$, $\mathbb{P}(A) = 1$,

Can we conclude that $A = \Omega$? No ; $\setminus\{1/2\}$ from $[0,1]$

A is called an event that occurs almost surely (a.s.)

Conditional Probability

Def: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and the event $B \in \mathcal{F}$ be such that $\mathbb{P}(B) > 0$. For any event $A \in \mathcal{F}$, the conditional probability that A occurs given B is denoted as $\mathbb{P}(A|B)$ and is defined by:

$$\boxed{P(A|B) = \frac{P(A \cap B)}{P(B)}}$$

$$\Rightarrow \quad P(A \cap B) = P(A|B)\,P(B)$$

Example: Two fair dice are thrown. Given that the first die shows 3, what is the probability that the sum exceeds 6?

Intuition: The second die must be $\boxed{4,5,6}$. Answer: $\boxed{1/2}$

Principled answer:

$\Omega = \{1,2,\cdots,6\} \times \{1,2,\cdots,6\}$

$\mathcal{F} = 2^{\Omega}$

$|\mathcal{F}| = |2^{\Omega}| = 2^{36}$

$\mathbb{P}(A) = \frac{|A|}{|\Omega|} \quad \forall A \in \mathcal{F}$

So we defined the probability space $(\Omega, \mathcal{F}, \mathbb{P})$

Assume

A = the event that the total number exceeds 6

$A = \{ (a,b) \mid a,b \in \{1,2,\cdots 6\},\ a+b > 6 \}$

B = the event that the first die shows 3

$B = \{ (3,1), (3,2), \ldots (3,6) \}$

Find $\mathbb{P}(A \mid B)$: $\frac{P(A \cap B)}{P(B)} = \frac{3/36}{6/36} = 1/2$

Remark: $\mathbb{P}(A|B)$ is left undefined when $\mathbb{P}(B)=0$, i.e. when B is a zero probability event. In other words, we cannot condition on zero probability events.

Theorem (Conditional Probability is a Probability Measure Itself)

Assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $B \in \mathcal{F}$ and $\mathbb{P}(B) > 0$.

$\forall A \in \mathcal{F}$: $\mathbb{P}_B(A) = \mathbb{P}(A|B)$

is a probability measure, i.e. it satisfies Kolmogorov's axioms.

Proof:

(1) $0 \leq \mathbb{P}_B(A) \leq 1$ $A \cap B \subseteq B$

$$0 \leq P(A \cap B) \leq P(B) \leq 1$$

$$0 \leq \frac{P(A \cap B)}{P(B)} \leq 1$$

(2) $\mathbb{P}_B(\Omega) = 1$

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

(3) Countable Additivity

For disjoint events $A_1, A_2, \ldots$

$$\mathbb{P}_B\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}_B(A_i)$$

$$P_B\left(\bigcup_{i \in N} A_i\right) = \frac{P([\cup A_i] \cap B)}{P(B)} = \frac{P(\cup A_i \cap B)}{P(B)}$$

claim $i \neq j \quad (A_i \cap B) \cap (A_j \cap B) = \emptyset$

$\therefore \quad C_i = A_i \cap B$ are disjoint

$$= \frac{\sum_i P(A_i \cap B)}{P(B)} = \sum_i \frac{P(A_i \cap B)}{P(B)}$$

Exercise : Assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $B \in \mathcal{F}$, $\mathbb{P}(B) > 0$.

(a) Show that the "reduced σ-field" $\mathcal{F} \cap B = \{B \cap A \mid A \in \mathcal{F}\}$ is a σ-field itself.

(b) Show that $(\overset{B}{B \cap \Omega}, \overset{\mathcal{F}_B}{B \cap \mathcal{F}}, \overset{P_B(A)}{\mathbb{P}_B(A)})$ is also a probability space.

Remark: This exercise presents the "reduced sample space" view of conditional probability. In this viewpoint, conditioning on B is equivalent to selecting B as the new sample space.

$F \cap B$ is a sigma field

$C \in F \cap B$ show $C^c \in F \cap B$

complement of C with respect

$$B \setminus C = B \cap \underbrace{(A \cap B)}_{\in F}{}^c$$

Theorem (The Law of Total Probability):

Assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $\exists\, A, B \in \mathcal{F}$ s.t. $0 < \mathbb{P}(B) < 1$.

Then: $\mathbb{P}(A) = \mathbb{P}(A|B)\,\mathbb{P}(B) + \mathbb{P}(A|B^c)\,\mathbb{P}(B^c)$

Proof:

$$A = A \cap \Omega = A \cap (B \cup B^c)$$

$$= (A \cap B) \cup (A \cap B^c)$$

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

$$= P(A|B)P(B) + P(A|B^c)P(B^c)$$

note. $0 < P(B) < 1$

$P(B) \neq 0 \wedge P(B^c) \neq 0$

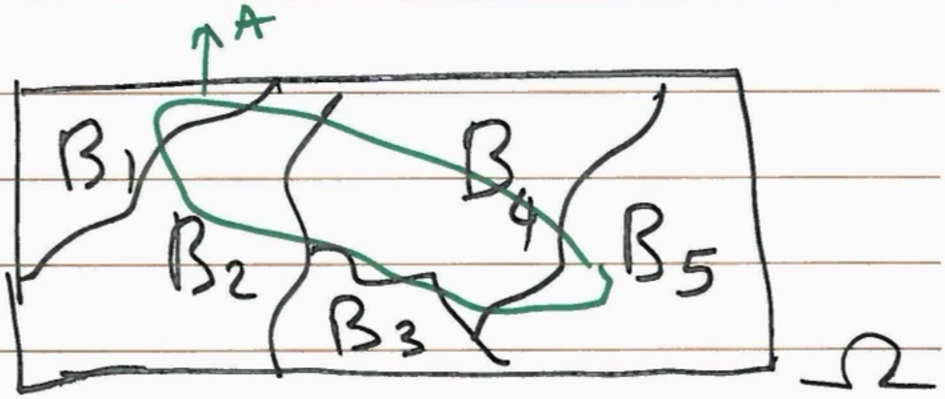
Theorem (General Total Probability)

Assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $B_1, B_2, \dots$ is a partition of Ω such that $\mathbb{P}(B_i) > 0$, then:

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A|B_i)\,\mathbb{P}(B_i)$$

Remark: For a finite partition $B_1, B_2, \dots, B_n$ the theorem is still true, i.e.

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|B_i)\,\mathbb{P}(B_i)$$



Proof:

$$A = A \cap \Omega = A \cap \left(\bigcup_{i=1}^{5} B_i\right)$$

$$= \bigcup_{i=1}^{n} (A \cap B_i) \qquad \because (A \cap B_i) \cap (A \cap B_j) = \phi \quad \forall i \neq j$$

$$P(A) = P\left(\bigcup_{i=1}^{n} A \cap B_i\right) = \sum_{i=1}^{n} P(A \cap B_i)$$

$$= \sum P(A \mid B_i) P(B_i)$$

Infinite case

$$A = A \cap \Omega = A \cap \left(\bigcup_{i \in N} B_i \right)$$

$$= \bigcup_{i \in N} \left(A_i \cap B_i \right)$$

$$P(A) = P\left(\bigcup_{i \in N} (A \cap B_i) \right)$$ by continuity of prob.

$$= \lim_{n \to \infty} \sum_{i=1}^{n} P(A \cap B_i) = \lim_{n \to \infty} \sum P(A|B_i) B_i$$

$$= \sum_{i=1}^{\infty} P(A|B_i) B_i$$

Example: A USC student is participating in a charity fundraiser event. CEO's of small sized, medium sized, and large companies are participating in the event as well. 50% of the CEO's are from large companies, 25% are from medium companies, and 25% are from small companies.

Probability of raising funds from ~~the~~ the CEO of a large company is 20%, while it is 70% for the CEO of a medium company and 30% for the CEO of

a small company.

What is the probability of raising funds from a randomly selected CEO?

Solution:

B_1 small
B_2 medium
B_3 large

A raising fund

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3)$$

$$= 0.3 \cdot 0.25 + 0.7 \cdot 0.25 + 0.2 \cdot 0.5$$

Theorem (The Baye's Rule)

Assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $A, B \in \mathcal{F}$ and $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$.

Then:

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B)\,\mathbb{P}(A|B)}{\mathbb{P}(A)}$$

$$= \frac{\mathbb{P}(B)\,\mathbb{P}(A|B)}{\mathbb{P}(A|B)\,\mathbb{P}(B) + \mathbb{P}(A|B^c)\,\mathbb{P}(B^c)}$$

Proof:

$$P(B|A) = \frac{P(A\cap B)}{P(A)} = \frac{P(A|B)\cdot P(B)}{P(A)}$$

$$= \frac{P(A|B)\cdot P(B)}{P(A|B)\cdot P(B) + P(A|B^c)\cdot P(B^c)}$$

$$\underbrace{P(H|E)}_{\text{Posterior}} = \frac{\underbrace{P(E|H)}_{\text{likelihood}}\cdot \underbrace{P(H)}_{\text{prior}}}{\underbrace{P(E)}_{P(E|H)\cdot P(H)+P(E|H^c)\cdot P(H^c)}}$$

$$P(H^c|E) = \frac{P(E|H^c)\cdot P(H^c)}{P(E|H^c)\cdot P(H^c) + P(E|H)\cdot P(H)}$$

$$\Rightarrow P(H|E) \propto P(E|H)\cdot P(H)$$

$$P(H^c|E) \propto P(E|H^c)\cdot P(H^c)$$

Theorem (General Baye's Rule)

Assume that $(\Omega, \mathcal{F}, \mathbb{P})$, and $A, B_i \in \mathcal{F}$ $\forall i \in \mathbb{N}$. Furthermore, B_i's form a partition and $\forall i \in \mathbb{N}$, $\mathbb{P}(B_i) > 0$. Then:

$$\mathbb{P}(B_i \mid A) = \frac{P(A \mid B_i) \cdot P(B_i)}{P(A)}$$

$$= \frac{P(A \mid B_i) \cdot P(B_i)}{\sum_{j=1}^{\infty} P(A \mid B_j) P(B_j)}$$

Example: (Charity Fundraiser)

Assume that the USC student raised some funds from a CEO in the event. What is the probability that it was from the CEO of a large company?

$P(B_3) = 0.5$

$$P(B_3|A) = \frac{P(A|B_3) \cdot P(B_3)}{P(A|B_1) \cdot P(B_1) + P(A|B_2) \cdot P(B_2) + P(A|B_3) \cdot P(B_3)}$$

Remark: Total probability and Baye's Rule deal with conditional probabilities of events given partitions of a sample Space:

When The (Unconditional) Probability of An Event Is Needed — USE TOTAL PROB

When The (Conditional) Probability of One of The Partitions Is Needed — USE BAYE'S

(you need evidence)

Remark: In Baye's Rule, A is an event that we observe, while B_i's are events that we do not observe, but would like to make some inference about.

Before any observation we know the <u>prior probabilities</u> $\mathbb{P}(B_i)$, and also know the <u>conditional probabilities</u> $\mathbb{P}(A|B_i)$.

$P(H_i)$

$P(E|H_i)$

After observing A, we compute

the posterior probabilities

$\mathbb{P}(B_i|A)$. We actually
$P(H_i|E)$

"update our belief" about

occurence of B_i, given that A

was observed.

Theorem: (The Multiplication Rule)

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $A_1, A_2, \ldots, A_n$ be events, i.e. $A_i \in \mathcal{F} \ \forall i \in \{1, 2, \ldots, n\}$.

Then:

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbb{P}(A_1)\mathbb{P}(A_2 | A_1) \ldots \mathbb{P}(A_n | A_1 \cap A_2 \cap A_3 \ldots \cap A_{n-1}) = \boxed{\mathbb{P}(A_1)\prod_{i=2}^{n} \mathbb{P}(A_i | A_1 \cap A_2 \cap \cdots \cap A_{i-1})}$$

Provided that the conditional probabilities are defined, i.e. $\forall l \in \{1, 2, \ldots, n\}$ $\mathbb{P}\left(\bigcap_{j=1}^{l} A_j\right) \neq 0$.

Proof (by induction)

$n=1$: $\mathbb{P}(A_1) = P(A_1)$

$n=2$: $\mathbb{P}(A_1 \cap A_2) = P(A_1) \cdot P(A_2 | A_1)$

Next, assume $\mathbb{P}\left(\bigcap_{i=1}^{k} A_i\right) = \mathbb{P}(A_1) \times$

$\times \mathbb{P}(A_2|A_1) \ldots \mathbb{P}(A_k | A_1 \cap A_2 \cap \ldots \cap A_{k-1})$.

One must show that

$$\mathbb{P}\left(\bigcap_{i=1}^{k+1} A_i\right) = \mathbb{P}(A_1) \prod_{i=2}^{k+1} \mathbb{P}(A_i | A_1 \cap A_2 \cap \ldots \cap A_{i-1})$$

$$P\left(\bigcap_{i=1}^{k+1} A_i\right) = P\left(\bigcap_{i=1}^{k} A_i \cap A_{k+1}\right)$$

$$= P\left(\bigcap_{i=1}^{k} A_i\right) P\left(A_{k+1} \,\Big|\, \bigcap_{i=1}^{k} A_i\right)$$

$$= P(A_1)\, P(A_2|A_1) \cdots P(A_k | A_1 \cap A_2 \cdots \cap A_{k-1}) \cdot P(P_{k+1} | A_1 \cap A_2 \cdots \cap A_k)$$

$$= P(A_1) \prod_{i=2}^{k+1} P(A_i | A_1 \cap A_2 \cap \cdots A_{i-1})$$

First order Markovian Property

$$P(A_i | A_1 \cap A_2 \cap \cdots \cap A_{i-1}) = P(A_i | A_{i-1})$$

Theorem: (General Multiplication Rule):

The multiplication rule holds for an infinite number of events $A_1, A_2, \ldots$, where $A_i \in \mathcal{F} \; \forall i \in \mathbb{N}$

i.e.

$$\mathbb{P}\left(\bigcap_{i=1}^{\infty} A_i\right) = \mathbb{P}(A_1) \prod_{i=2}^{\infty} \mathbb{P}(A_i \mid A_1 \cap \cdots \cap A_{i-1})$$

provided that all conditional probabilities are defined, i.e.

$$\mathbb{P}\left(\bigcap_{j=1}^{l} A_j\right) \neq 0 \quad \forall l \in \mathbb{N}$$

Proof: Take a $\lim_{n\to\infty}$ from the multiplication rule and use continuity of probability.

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1)\prod_{i=2}^{n} P(A_i \mid A_1 \cap A_2 \cdots \cap A_{i-1})$$

$$\lim_{n\to\infty} P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1)\prod_{i=2}^{\infty} P(A_i \mid A_1 \cap A_2 \cdots \cap A_{i-1})$$

$$= P\left(\lim_{n\to\infty}\bigcap_{i=1}^{n} A_i\right) = P\left(\bigcap_{i\in\mathbb{N}} A_i\right) \quad \text{Q.E.D.}$$

– Independence

Assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $A, B \in \mathcal{F}$. A, B are called independent iff

$$\boxed{P(A \cap B) = P(A) \cdot P(B)}$$

and if $\mathbb{P}(B) > 0$, it means:

$$\boxed{P(A|B) = \frac{P(A|B)}{P(B)} = \frac{P(A) \cdot P(B)}{P(B)} = P(A)}$$

Question: Can an event be independent from itself?

$P(A \cap A) = P(A)^2$

$P(A) = 0$ or $P(A) = 1$

almost sure and zero prob event

Example: $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, $A, B \in \mathcal{F}$, and $0 < \mathbb{P}(B) < 1$.

If A and B are independent, show that the probability of A given B is equal to the probability of A given B^c.

In other words, the probability of A doesn't change whether or not B occurs.

$$\mathbb{P}(A|B) = \mathbb{P}(A|B^c)$$

$$A = (A \cap B) \cup (A \cap B^c)$$

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

$$= P(A|B)P(B) + P(A|B^c)P(B^c)$$

$$= P(A|B)P(B) + P(A|B^c)(1-P(B))$$

$$= P(A)P(B) + P(A|B^c)(1-P(B))$$

$$P(A) - P(A)P(B) = P(A|B^c)(1-P(B))$$

$$= P(A)(1-P(B)) = P(A|B^c)(1-P(B))$$

$$\Rightarrow P(A) = P(A|B^c)$$

$$\therefore P(A|B) = P(A|B^c)$$

Exercise

show A^c is independent from B and B^c

$\Rightarrow$ σ-field generated by A is independent from σ-field generated by B

$F_A = \{\phi, A, A^c, \Omega\}$

$F_B = \{\phi, B, B^c, \Omega\}$

because every member of F_A is independent from every member of F_B

Remark: The concept of mutually exclusive events should not be confused with that of independent events.

Mutually exclusive sets (events) are sets whose overlap is empty. We do not need a probability measure to assess whether they are mutually exclusive or not. On the other hand, independent events are meaningful within the concept of probability. Without having a probability space equipped with a probability measure, one cannot define independence.

Is there event mutually exclusive with itself

$A \cap A = \phi \quad A = \phi$

When two mutually exclusive events independent

$P(A \cap B) = P(A)P(B) = 0$

at least one of event is 0 probability

Conditional Independence

Assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and C is $\mathcal{F}$-measurable, i.e. $C \in \mathcal{F}$. Also assume that $\mathbb{P}(C) > 0$.

$A, B \in \mathcal{F}$ are called conditionally independent given C iff:

$$\boxed{P(A \cap B \mid C) = P(A \mid C) P(B \mid C)}$$

Use in Bayesion Network / Graphical models / Naïve Bayes Model

Example: Assume that A, B are independent events. Are they conditionally independent given any $C \in \mathcal{F}$, for which $\mathbb{P}(C) > 0$? No

$P(A \cap B) = P(A)P(B)$

if $P(A \cap B | C) \stackrel{?}{=} P(A|C)P(B|C)$

$$\frac{P(A \cap B \cap C)}{P(C)} \stackrel{?}{=} \frac{P(A \cap C)}{P(C)} \frac{P(B \cap C)}{P(C)}$$

Does not hold for all C

Provide counter example

Counter-example (to disprove):

Roll of a die:

$\{A\} = \{1, 2\}$ $\quad B = \{2, 3, 4\}$, $C = \{2, 5\}$

$P(A \cap B \cap C) = 1/6 \quad P(A \cap B) = 1/6 \quad P(A) = 1/3 \quad P(B) = \frac{1}{2}$

$P(C) = 1/3 \quad\quad P(A) \cdot P(B) = P(A \cap B)$: independent

$P(A \cap C) = 1/6 \quad P(B \cap C) = 1/6$

$$\frac{1/6}{1/3} \neq \frac{1/6}{1/3} \frac{1/6}{1/3}$$

$\therefore \; P(A \cap B | C) \neq P(A|C) P(B|C)$

Independence for More Than Two Events

For an indexed family of events $\{A_i \in \mathcal{F} \mid i \in I\}$, A_i's are independent if for all J, finite subsets of I,

$$\mathbb{P}\left(\bigcap_{i\in J} A_i\right) = \prod_{i\in J} \mathbb{P}(A_i) \quad \forall J \subseteq I, \; |J| < \infty$$

Therefore, if A, B, C are independent then:

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

Note: I can be infinite

Example: We have a well-shuffled pack of 52 playing cards. Is the suit of a card independent from its rank?

suit = {D, C, S, H}

ranks = {2, 3, ... 9, J, Q, K, A}

i.e.) suit = H rank = Q

$P(\text{suit}=H) = 13/52 = 1/4$

$P(\text{rank}=Q) = 4/52 = 1/13$

$P(\text{suit}=H \wedge \text{rank}=Q) = 1/52$

$\therefore P(\text{suit}=H)P(\text{rank}=Q) = P(\text{suit}=H \wedge \text{rank}=Q)$

the above calculation would be true for all pairs

Pairwise Independence.

$\{A_i \in \mathcal{F} \mid i \in I\}$ is called pairwise independent if

$$\mathbb{P}(A_i \cap A_j) = P(A_i)P(A_j) \quad i \neq j$$

Question: which of the following implications is true?

pairwise independence $\not\Rightarrow$ Independence

Independence $\Rightarrow$ Pairwise Independence

This is because if $\{A_i \mid i \in I\}$ are independent, $\{A_i, A_j \mid i \neq j\}$, $\{A_i, A_j, A_k \mid i, j, k \text{ are distinct}\}$, etc must be independent but the reverse is not true.

Exercise: Consider mutually exclusive events B_1, B_2, B_3, and C on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with $\mathbb{P}(B_1) = \mathbb{P}(B_2) = \mathbb{P}(B_3) = p$ and $\mathbb{P}(C) = q$, where $3p+q \leq 1$. Assume that $p = -q + \sqrt{q}$, and show that the events $B_1 \cup C$, $B_2 \cup C$, and $B_3 \cup C$ are pairwise independent. Also, is there any $p>0$ and $q>0$

such that these three events are independent.

Pairwise independent

$$P((B_i \cup C) \cap (B_j \cup C))$$
$$= P(B_i \cup C)\, P(B_j \cup C)$$
$$= P((B_i \cap B_j) \cup C) = P(C) = q$$

$$P(B_i \cup C) = P(B_i) + P(C) = p + q = \sqrt{q}$$

$\therefore$ $B_i \cup C$ and $B_j \cup C$ are pairwise independent

Independent

$$P((B_1 \cup C) \cap (B_2 \cup C) \cap (B_3 \cup C)) = P(B_1 \cup C) P(B_2 \cup C) P(B_3 \cup C)$$

$$\Rightarrow P((B_1 \cap B_2 \cap B_3) \cup C) = q\sqrt{q}$$

$$P(C) = q\sqrt{q} = q$$

$q = 0$ or $q = 1$

$\Rightarrow$ both not acceptable

This serve as a general counter example for each that are pairwise indep but not indep

The Borel - Cantelli Lemmas

An event that occurs infinitely often:

Assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $A_i \in \mathcal{F}$

$\forall i \in \mathbb{N}$.

Define:

$$B_1 = \bigcup_{i=1}^{\infty} A_i$$

$$B_2 = \bigcup_{i=2}^{\infty} A_i$$

$$\vdots$$

$$B_n = \bigcup_{i=n}^{\infty} A_i$$

If $A_{i.o.}$ is an event that occurs **infinitely often**, then $\therefore$

$x \in A_{i.o.}$ must be in all B_n's,

i.e. $\forall n, x \in B_n \Rightarrow x \in \bigcap_{i=1}^{\infty} B_i$

$$\Rightarrow A_{i.o.} \subseteq \bigcap_{i=1}^{\infty} B_i = \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_n$$

On the other hand, if

$x \in B_K \; \forall K \; \forall K \in \mathbb{N}$, we can

show that x is a member of

an infinite number of A_i's, i.e.

$x \in A_{i.o.}$:

$x \in B_1 \Rightarrow x \in \bigcup_{i=1}^{\infty} A_i \Rightarrow \exists i > 0 \text{ s.t } x \in A_i$

$\vdots$

$x \in B_2 \Rightarrow x \in \bigcup_{i=2}^{\infty} A_i \Rightarrow \exists i > 1 \text{ s.t } x \in A_i$

$\vdots$

$x \in B_{l+1} \Rightarrow x \in \bigcup_{i=l+1}^{\infty} \Rightarrow \exists i > l \text{ s.t } x \in A_i$

$\vdots$

$\forall k \; x \in B_k \Rightarrow x \in \bigcap_{i=1}^{\infty} B_i$

Therefore, there is an infinite sequence of integers $\{j_l\}$ s. t.

$x \in A_{j_l} \quad \forall \; l \in \mathbb{N}. \Rightarrow$

$x \in A_{i.o.} \Rightarrow \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i \subseteq A_{i.o}$

$A_{i.o.} = \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i$

Lemma (Borel – Cantelli):

Assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, $A_1, A_2, \ldots,$ are events, i.e. $A_i \in \mathcal{F}\ \forall i \in \mathbb{N}$.

Define $A_{i.o.} = \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i$. Then:

(1) $\sum_{i=1}^{\infty} \mathbb{P}(A_i) < \infty \Rightarrow \mathbb{P}(A_{i.o.}) = 0$

(2) $\sum_{i=1}^{\infty} \mathbb{P}(A_i) = \infty$ and A_i's are independent $\Rightarrow \mathbb{P}(A_{i.o.}) = 1$.

Proof: $A_{i.o.} = \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i$

$A_{i.o.} \subseteq B_n$

$\Rightarrow P(A_{i.o.}) \leq P(B_n) = P\left(\bigcup_{i=n}^{\infty} A_i\right) \leq \sum_{i=n}^{\infty} P(A_i)$

$\Rightarrow P(A_{i.o.}) \leq \sum_{i=n}^{\infty} P(A_i)$

$\Rightarrow \lim_{n\to\infty} P(A_{i.o.}) \leq \lim_{n\to\infty} \sum_{i=n}^{\infty} P(A_i)$

Example (Erdös – Rényi):

Consider an experiment in which a coin is tossed independently an infinite number of times. Let

$A_i = \{H \text{ occurs in the } i^{th} \text{ coin toss}\}$

(1) Assume that $\mathbb{P}(A_i) = 1/2$.

Then $\sum_{i=1}^{\infty} \mathbb{P}(A_i) = \infty$

By the second Borel - Cantelli lemma, it follows that infinitely many heads will occur with probability 1 (almost surely)

(2) Next assume that $\mathbb{P}(A_i) = \left(\frac{1}{2}\right)^i$.

Then $\sum_{i=1}^{\infty} \mathbb{P}(A_i) = 1$

Hence, by the first Borel-Cantelli lemma, the probability that infinitely many heads occur is 0.

In other words, almost surely, only finitely many heads will occur!

It might appear surprising that in the first example, where probability of getting heads is $1/i$, for any i we select (no matter how large), there occurs a heads

beyond i almost surely.

The reason is that the decay rate $1/i$ of the probability of observing heads is not "adequately fast."

On the other hand, in the second example, the rate of decay $(\frac{1}{2})^i$ is adequately fast that after a finite i, there will almost surely be no heads.

Exercise: (Infinite Monkey):

A monkey sits in front of a computer and starts pressing keys randomly on the keyboard. Using the Borel - Cantelli Lemmas, show that the monkey will type any text of your choice (e.g. ABRACADABRA or the whole declaration of independence), with probability 1.

You can assume that the monkey selects the keys uniformly at random, and that it selects the successive keys independently.

assume $P(A_i) = t$ $\left(\frac{1}{26}\right)^P$

$\sum_{i=1}^{\infty} P(A_i) = \infty$