

LESSON 4

Probability : Sample Space,
Events, Classical And Axiomatic

Definitions of Probability,

Properties of Probability Measures

What is probability?

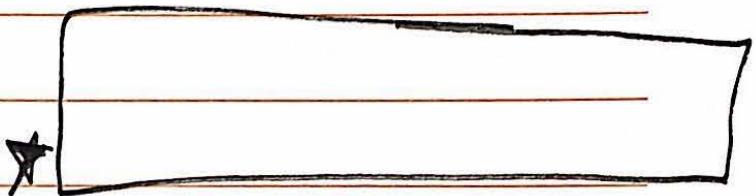
It is a mathematical theory of
'probabilistic' uncertainty.

Remark: There are other
types of uncertainty that

Cannot be modeled using
probability.

Example : Make a random
number generator select

$$x \in [0, 1]:$$



what is the probability
of selecting \star ? 0

Is it probable to
select $\star ?$

no

Is it possible to select
 $\star ?$

yes

probable \rightarrow possible

possible $\not\rightarrow$ probable

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

Possibility is modelled by See 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? $\frac{1}{2}$
 $\frac{1}{2}$ $> \frac{1}{2}$

Classical Definition of

probability :

Ratio of favorable outcomes

and the total number of

possible outcomes, if all

outcomes are equally likely

$$P(A) = \frac{N_A}{N}$$

Example: Five balls, one's red, the

rest are blue.

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

Relative Frequency Definition:

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

Example: We flip a coin

10,000,000 times.

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

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 $\Omega = \{0, 1, 2, \dots\}$

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

- Interested in the speed with which the coin hits the ground $\Omega = \{R\} \text{ or } \{D, C\}$

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

$$\Omega = [0, \infty]$$

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

Def (Informal) : An event is

Said to have occurred if the outcome of the experiment ω belongs to it.

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

$$E_1 = \emptyset, E_2 = \{H\}, E_3 = \{T\}, E_4 = \{H, T\} = \Omega$$

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

$$E = \{2, 4, 6\} \equiv \boxed{\text{Seeing an even number}}$$

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

$$E = [50, 120] \equiv \boxed{\text{person lives more than 50 years}}$$

Assignment of probability to

events: \rightarrow nothing occurred, impossible event

$$P(\emptyset) = 0$$

$$P(H) = p \quad P(T) = 1-p$$

$$P(\Omega) = 1$$

\downarrow Something occurred
= Sure event

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

Roll of a die

$$P(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{6} \quad A \subseteq \Omega$$

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 occurred

$$A \cup B$$

A and B occurred

$$A \cap B$$

A occurred, but B didn't occur

$$A \cap B^c = A \setminus B$$

To model events, we need

Some structure that
includes all possible events

σ -field of events

(non-empty)

Def: A subset of 2^{Ω} , $F \subseteq 2^{\Omega}$

is called a σ -field (or σ -algebra)
if

(1) $\emptyset \in F$ (redundant)

$\emptyset \in F, \Omega \in F$

(2) $A \in F \Rightarrow A^c \in F$

closure under complement

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

closure under countable union

This structure allows us to

model events like

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

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

Example : The smallest σ -field for any Ω

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

Example : $A \subseteq \Omega$ $\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$
 smallest sigma field generated by $A \subseteq \Omega$

Example : is 2^Ω a σ -field?

yes

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

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

Exercise : Using axioms (2), (3)
show that \emptyset, Ω also in \mathcal{F}

$$A_i = \begin{cases} A & i \in \emptyset \\ A^c & i \in \mathbb{N} \end{cases}$$

$$\begin{aligned} A \in \mathcal{F}, A^c \in \mathcal{F} &\Rightarrow A_i \in \mathcal{F}, \forall i \in \mathbb{N} \\ \Rightarrow \bigcup_{i \in \mathbb{N}} A_i &\in \mathcal{F} = A \cup A^c \cup A \cup A^c \dots = \Omega \\ \Rightarrow \Omega &\in \mathcal{F} \Rightarrow \Omega^c \in \mathcal{F} \Rightarrow \emptyset \in \mathcal{F} \end{aligned}$$

Lemma : if A_1, A_2, \dots, A_n is a finite sequence of events
 $\bigcup_{i=1}^n A_i \in \mathcal{F}$
define $B_i = \begin{cases} A_i & 1 \leq i \leq n \\ \emptyset & i > n \end{cases}$

$$\begin{aligned} \bigcup_{i \in \mathbb{N}} B_i &\in \mathcal{F} \quad \because B_i \in \mathcal{F} \\ \Rightarrow \bigcup_{i=1}^n A_i &\in \mathcal{F} \end{aligned}$$

Probability Measure (Kolmogorov)

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

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

$$P: \mathcal{F} \xrightarrow{\text{Set}} [0, 1]$$

s.t.

(1) Probability of Ω

$$P(\Omega) = 1$$

(2) Probability of countable unions

of disjoint events

(Countable additivity property)

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

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

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

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

(1) Probability of \emptyset

$$P(\emptyset) = 0$$

(2) Probability of Ω

$$P(\Omega) = 1$$

(3) Countable additivity

We call (Ω, \mathcal{F}, P)

a probability space or a
probability model or a
probability triplet

Aside: 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: $\Omega = \{H, T\}$

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

$$P(\emptyset) = 0$$

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

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

$$P(\Omega) = 1$$

Roll of a die

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

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

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

$$P(A) = \sum_{i \in A} p_i \quad \text{i.g.) } P(\{2, 3, 6\}) \\ \checkmark A \subseteq \Omega \\ = p_2 + p_3 + p_6$$

Lemma

(1) Additivity for a finite number of events

$$A_1, \dots, A_n \in \mathcal{F}, A_i \cap A_j = \emptyset \quad \checkmark i \neq j$$

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

(2) Probability of complement

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

(3) Preservation of order
Monotonicity

$$A \subseteq B \Rightarrow P(A) \leq P(B)$$

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

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

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

(5) Inclusion-Exclusion

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

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

(6) Inclusion-Exclusion

(General form)

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

not necessarily disjoint

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n P(A_i) - \sum_{i < j} P(A_i \cap A_j) \\ &\quad + \sum_{i < j < k} P(A_i \cap A_j \cap A_k) + \dots \\ &\quad + (-1)^{n+1} P\left(\bigcap_{i=1}^n A_i\right) \end{aligned}$$

for $n=3$

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) &= P(A_1) + P(A_2) + P(A_3) \\ &\quad - \{P(A_1 \cap A_2) + P(A_1 \cap A_3) + P(A_2 \cap A_3)\} \\ &\quad + P(A_1 \cap A_2 \cap A_3) \end{aligned}$$

for $n=4$

$$P(A_1 \cup A_2 \cup A_3 \cup A_4) =$$

(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

$$B_i = \begin{cases} A_i & 0 \leq i \leq n \\ \emptyset & i > n \end{cases}$$

, B_i 's are mutually exclusive

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

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

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

$$(2) \quad P(A) = 1 - P(A^c)$$

Note that

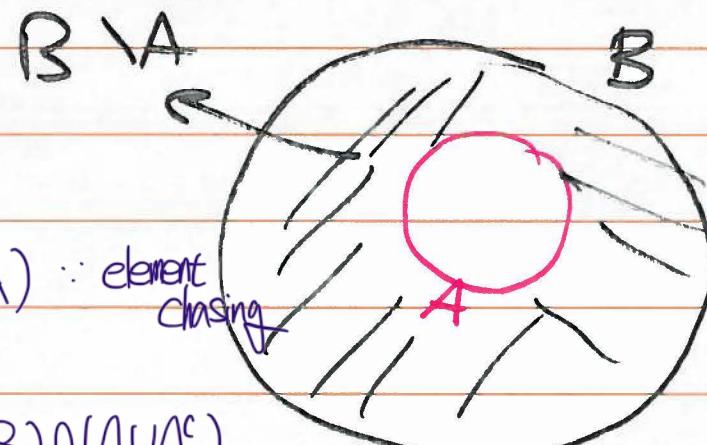
$$A \cup A^c = \Omega$$

$$A \cap A^c = \emptyset$$

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

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

$$(3) \quad A \subseteq B \quad P(A) \leq P(B)$$



$$B = A \cup (B \setminus A) \quad \because \text{element chasing}$$

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

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

$$A \cap (B \setminus A) = \emptyset \quad \therefore A \text{ and } B \setminus A \text{ are disjoint}$$

$$P(B) = P(A \cup (B \setminus A)) = P(A) + P(B \setminus A)$$
$$\Rightarrow P(A) \leq P(B) \quad \therefore P(B \setminus A) \geq 0$$

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

Important Corollary:

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

(A) Boole's Inequality

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

Union bound / Subadditivity of the probability measure

define a sequence of event

$$B_1 = A_1$$

$$B_2 = A_2 \setminus A_1$$

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

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

⋮

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

Claim B_i 's are mutually exclusive (disjoint)

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

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

$$\Rightarrow \bigcup_{i=1}^n B_i = \bigcup_{i=1}^n A_i \quad \text{← Why?}$$

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

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

Claim $B_i \subseteq A_i \Rightarrow P(B_i) \leq P(A_i)$

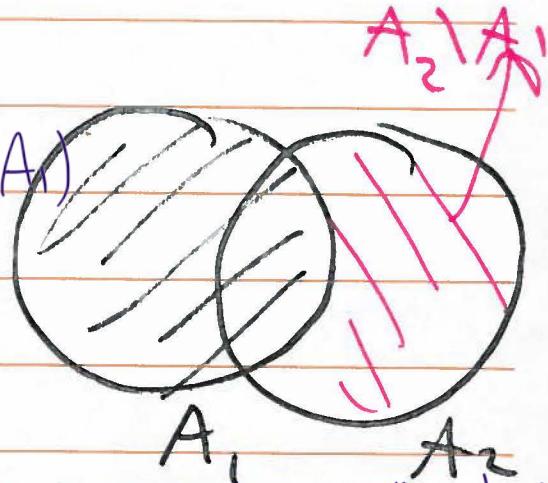
$$\Rightarrow \sum_{i=1}^{\infty} P(B_i) \leq \sum_{i=1}^{\infty} P(A_i)$$

$$\Rightarrow P\left(\bigcup_{i \in \mathbb{N}} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i) \quad Q.E.D.$$

(5) Inclusion Exclusion

$$\begin{aligned} P(A_1 \cup A_2) &= P(A_1) + P(A_2) \\ &\quad - P(A_1 \cap A_2) \end{aligned}$$

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



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

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

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

$$\text{Claim } A_2 \setminus A_1 = A_2 \setminus (A_1 \cap A_2)$$

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

because $A_1 \cap A_2 \subseteq A_2$

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

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

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

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

Induction on n

$$n=1 \Rightarrow 1=1 \text{ true } \checkmark$$

Assume for $n=k$

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

To use induction, we must

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

$$\begin{aligned} S(k+1) &: 1+3+5+\dots+(2k-1)+(2k+1) \\ &= k^2 + 2k+1 = (k+1)^2 \end{aligned}$$

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

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

Back to proving inclusion-exclusion

$$\begin{aligned} n=2 \quad P(A_1 \cup A_2) &= P(A_1) + P(A_2) \\ &\quad - P(A_1 \cap A_2) \end{aligned}$$

Assume that

$$n=k: \quad P(A_1 \cup A_2 \cup \dots \cup A_k)$$

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

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

$$+ (-1)^{k+1} P(A_1 \cap A_2 \cap \dots \cap A_k)$$

We must prove that for

$$n = k+1: \quad ?$$

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

(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

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

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

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

If you add * to all term $P(A_1 \cup A_2 \dots \cup A_{l+1})$

, you will get inclusive exclusion for $l+1$

A large rectangular frame with a red border, containing ten horizontal blue lines for handwriting practice.

A large rectangular frame with a red border, containing ten horizontal blue lines for handwriting practice.

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

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

This lemma defines a sequence

$$\text{of numbers } p_n = P\left(\bigcup_{i=1}^n A_i\right)$$

and states that the limit of

the sequence p_n is equivalent

to the probability $P\left(\bigcup_{i \in \mathbb{N}} A_i\right)$.

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

$$a_n : \mathbb{N} \rightarrow \mathbb{R}$$

Def: limit of a sequence

We say a_n converges to L as

$n \rightarrow \infty$ and write

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

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

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

Example: $\lim_{n \rightarrow \infty} \frac{1}{n} = 0$

Should prove that

$$\forall \varepsilon > 0 \quad \exists N. \quad \text{s.t.} \quad n > N_0$$

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

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

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

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

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

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

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

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

Back to Continuity of Probability.

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

Proof:

$$\text{define } B_1 = A_1$$

$$B_2 = A_2 \setminus A_1$$

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

$$B_i \cap B_j = \emptyset \quad i \neq j$$

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

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

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

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

Q.E.D.

Corollary :

If A_i is a sequence of nested events:

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

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

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

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

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

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

Proof

a) from continuity of probability

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

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

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

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

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

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

by (a)

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

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

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

Why is it called continuity?

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

The lim operator passes probability.

like continuous Functions:

$$\lim_{x \rightarrow x_0} f(x) = f\left(\lim_{x \rightarrow x_0} x\right) = f(x_0)$$

Remark(Zero Probability and

Almost Sure Events)

(Ω, \mathcal{F}, P) : Probability Space

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

Can we conclude that $A = \emptyset$?

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

A is called a zero probability event

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

Can we conclude that

$A = \Omega$? No : $\backslash \{\frac{1}{2}\}$ from $[0, 1]$

A is called an event that

occurs almost surely (a.s.)

Conditional Probability

Def : Let (Ω, \mathcal{F}, P) be a probability space and the event $B \in \mathcal{F}$ be such that $P(B) > 0$. For any event

$A \in \mathcal{F}$, the conditional probability that A occurs given B is denoted as $P(A|B)$ and is defined by :

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

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

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:

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

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

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

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

So we defined the probability space (Ω, \mathcal{F}, P)

Assume

~~A~~ = the event that the

total number exceeds 6

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

B = the event that the first die

shows 3

$$B = \left\{ (3,1), (3,2), \dots, (3,6) \right\}$$

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

Remark: $P(A|B)$ is left

undefined when $P(B)=0$, 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 (Ω, \mathcal{F}, P) is

a probability space and $B \in \mathcal{F}$

and $P(B) > 0$.

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

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

Proof:

$$(1) 0 \leq P_B(A) \leq 1 \quad A \cap B \subseteq B$$

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

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

$$(2) P_B(\Omega) = 1$$

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

(3) Countable Additivity

For disjoint events A_1, A_2, \dots

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

$$P_B \left(\bigcup_{i \in \mathbb{N}} A_i \right) = \frac{P(\bigcup_i A_i \cap B)}{P(B)} = \frac{P(\bigcup_i A_i \cap B)}{P(B)}$$

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

$\therefore C_i = A_i \cap B$ 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 (Ω, \mathcal{F}, P)

is a probability space and

$$B \in \mathcal{F}, P(B) > 0.$$

(a) Show that the "reduced

σ -field" $\mathcal{F} \cap \mathcal{B} = \{B \cap A \mid A \in \mathcal{F}\}$

is a σ -field itself.

(b) Show that $(B \cap \Omega, \mathcal{F}_B, P_B(A))$

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

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

Theorem (The Law of Total Probability):

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

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



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 (Ω, \mathcal{F}, P) is a probability space and

B_1, B_2, \dots is a partition of Ω such that $P(B_i) > 0$, then:

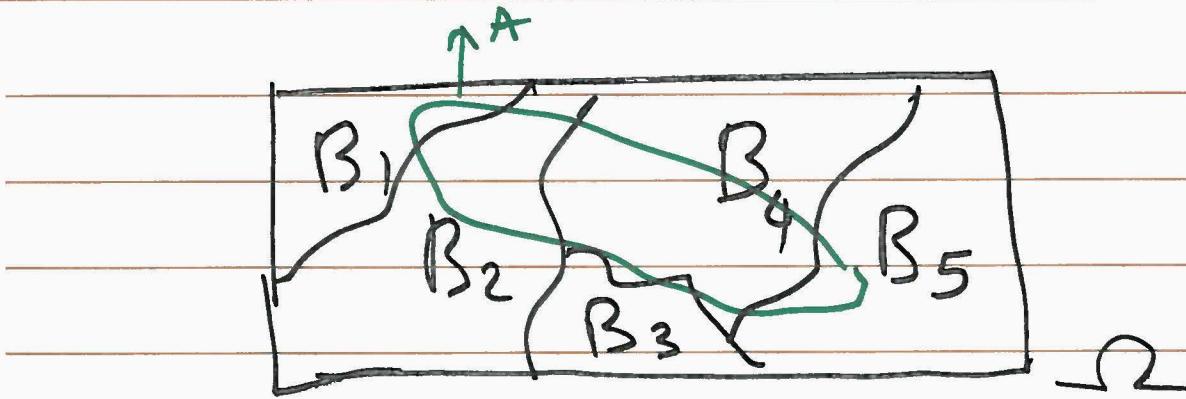
$$P(A) = \sum_{i=1}^{\infty} P(A|B_i) P(B_i)$$

Remark: For a finite partition

B_1, B_2, \dots, B_n the theorem is

still true, i.e.

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



Proof:

$$\begin{aligned}
 A &= A \cap \Omega = A \cap \left(\bigcup_{i=1}^5 B_i \right) \\
 &= \bigcup_{i=1}^5 (A \cap B_i) \quad \because \forall i \neq j \quad (A \cap B_i) \cap (A \cap B_j) = \emptyset
 \end{aligned}$$

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

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

Infinite case

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

$$= \bigcup_{i \in \mathbb{N}} (A \cap B_i)$$

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

$$\begin{aligned} &= \lim_{n \rightarrow \infty} \sum_{i=1}^n P(A \cap B_i) = \lim_{n \rightarrow \infty} \sum_{i=1}^n P(A|B_i) B_i \\ &= \sum_{i=1}^{\infty} P(A|B_i) B_i \end{aligned}$$

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

Then :

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

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

Proof:

$$\begin{aligned} P(B|A) &= \frac{P(A \cap B)}{P(A)} = \frac{P(A|B) \cdot P(B)}{P(A)} \\ &= \frac{P(A|B) \cdot P(B)}{P(A|B) \cdot P(B) + P(A|B^c) \cdot P(B^c)} \end{aligned}$$

$$\begin{aligned} P(H|E) &= \frac{P(E|H) \cdot P(H)}{P(E)} \\ \text{Posterior} &\quad \left. \begin{array}{l} \text{likelihood} \\ \text{prior} \end{array} \right\} \\ &= \frac{P(E|H) \cdot P(H) + P(E|H^c) \cdot P(H^c)}{} \end{aligned}$$

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

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

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

Theorem (General Baye's Rule)

Assume that (Ω, \mathcal{F}, P) , and

$A, B_i \in \mathcal{F} \forall i \in N$. Furthermore,

B_i 's form a partition and

$\forall i \in N, P(B_i) > 0$. Then:

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

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

Example : (Charity Fundraiser)

Assume that the USC student
raised some funds in the event
from a CEO

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:

Unconditional

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

When The 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 prior probabilities

$P(B_i)$, and also know the

$P(H_i)$

conditional probabilities $P(A|B_i)$.

$P(E|H_i)$

After observing A, we compute

the posterior probabilities

$P(B_i | A)$. We actually
 $P(H_i | E)$

"update our belief" about

occurrence of B_i , given that A
was observed.

Theorem: (The Multiplication Rule)

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

Then:

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

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

Provided that the conditional probabilities are defined, i.e. $\forall j \in \{1, 2, \dots, n\}$

$$P\left(\bigcap_{j=1}^k A_j\right) \neq 0.$$

Proof (by induction)

$$n=1 : \quad P(A_1) = P(A_1)$$

$$n=2 : \quad P(A_1 \cap A_2) = P(A_1) \cdot P(A_2 | A_1)$$

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

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

One must show that

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

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

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

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

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

First order Markovian Property

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

Theorem: (General Multiplication Rule):

The multiplication rule holds for an infinite number of events

A_1, A_2, \dots , where $A_i \in F \forall i \in \mathbb{N}$

i.e.

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

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

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

Proof: Take a lim from the
 $n \rightarrow \infty$

multiplication rule and use

Continuity of probability.

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

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

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

Q.E.D.

- Independence

Assume that (Ω, \mathcal{F}, P) is

a probability space and

$A, B \in \mathcal{F}$. A, B are

called independent iff

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

and if $P(B) > 0$, it means:

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

Question: Can an event be independent from itself?

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

$$P(A) = 0 \text{ or } P(A) = 1$$

almost sure and zero prob event

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

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.

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

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

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

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

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

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

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

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

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

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

Exercise

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

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

$$\mathcal{F}_A = \{\emptyset, A, A^c, \Omega\}$$

$$\mathcal{F}_B = \{\emptyset, B, B^c, \Omega\}$$

because every member of \mathcal{F}_A is independent from
every member of \mathcal{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

$$A \cap A = \emptyset \quad A = \emptyset$$

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 (Ω, \mathcal{F}, P) is a

probability space and C is

\mathcal{F} -measurable, i.e. $C \in \mathcal{F}$. Also

assume that $P(C) > 0$.

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

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

Use in Bayesian Network / Graphical models
/ Naive Bayes Model

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

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

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

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

Does not hold for all C

Provide counter example

Counter-example (\Rightarrow disprove):

Roll of a die:

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

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

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

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

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

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

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 ,

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

Therefore, if A, B, C are independent

$$\text{then: } P(A \cap B \cap C) = P(A)P(B)P(C)$$

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

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

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

Note: I can be infinite

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

$$\text{Suit} = \{D, C, S, H\}$$

$$\text{Ranks} = \{2, 3, \dots, 9, J, Q, K, A\}$$

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 .

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

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

Question: Which of the following implications is true?

pairwise independence $\not\Rightarrow$ Independence

Independence \Rightarrow Pairwise Independence

This is because if $\{A_i | i \in I\}$ are independent, $\{A_i, A_j | i \neq j\}$, 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 (Ω, \mathcal{F}, P) , with $P(B_1) = P(B_2) = p$, $P(B_3) = q$, and $P(C) = r$.

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

$$P((B_i \cup C) \cap (B_j \cup C))$$

$$= P(B_i \cup C) P(B_j \cup C)$$

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

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

$\therefore B_1 \cup C$ and $B_2 \cup C$ are pairwise independent

Independent

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

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

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

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

$$q = 0 \text{ or } q = 1$$

\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 (Ω, \mathcal{F}, 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 :

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

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

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

On the other hand, if

$x \in B_k \vee k \in \mathbb{N}$, we can

show that x is a member of

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

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

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

⋮

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

⋮

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

⋮

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

Therefore, there is an infinite

sequence of integers $\{j_l\}$ s.t.

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

$$x \in A_{i_{l,0}} \Rightarrow \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i \subseteq A_{i_{l,0}}$$

$$A_{i_{l,0}} = \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i$$

Lemma (Borel-Cantelli):

Assume that (Ω, \mathcal{F}, P) is a

probability space, A_1, A_2, \dots

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:

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

$$(2) \sum_{i=1}^{\infty} P(A_i) = \infty \text{ and } A_i's$$

are independent $\Rightarrow P(A_{i.o.}) = 1$.

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

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

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

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

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

A large rectangular frame with a brown border, designed for handwriting practice. It contains ten horizontal lines spaced evenly apart.

A large rectangular frame with a brown border, designed for handwriting practice. It contains ten horizontal lines spaced evenly apart.

Example (Erdős–Rényi):

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

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

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

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

By the second Borel-Cantelli

lemma, it follows that infinitely many heads will occur with probability 1 (almost surely)

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

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

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 γ_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 γ_i of the probability of

observing heads is not

"adequately fast."

On the other hand, in

the second example, the rate of

decay $(\gamma_2)^i$ 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.