2.1 Conditional Probability.

Notation "Suppress" probability measure space $(\Omega, \mathcal{A})$.

ex) $P \equiv$ A probability measure P on Ω with σ-algebra $\mathcal{A}$.

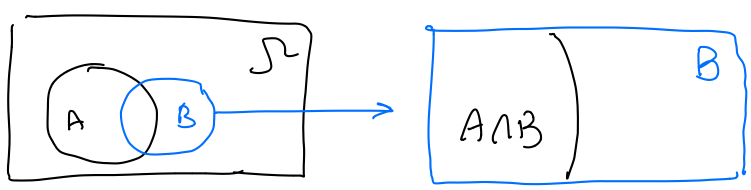
Note Very often, it doesn't matter what underlying measure space is, it only matters what P looks like in the events we're interested in.

Terminology event = measurable set = set in σ-algebra $\mathcal{A}$.

Sample space = Ω.

Def. Conditional Probability of A given B.

If $P(B) > 0$, then $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$.



2.2-3 Independence

"A fantastically useful property to have".

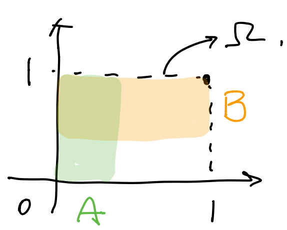
Def. (These are definitions used in different context, but not equivalent).

i) Independence for 2 events.

Events A and B are independent if $P(A \cap B) = P(A)P(B)$

Intuition A and B are not related → ex) coinflip & dice roll.

Example



Probability of the overlap is the product of probabilities of individual events

ii) Independence of multiple events

Finite set of events $A_1, A_2, \cdots, A_n$ are (mutually) independent

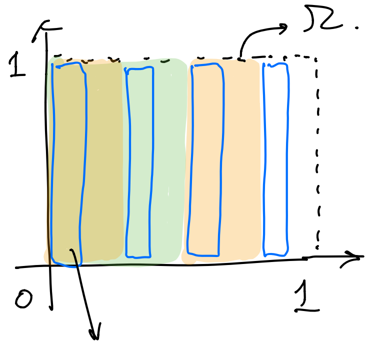
if for any subset $S \subset \{1, 2, \cdots, n\}$, $P(\bigcap_{i \in S} A_i) = \prod_{i \in S} P(A_i)$.

Remark. Mutual independence implies pairwise independence.
(any pair is independent)

However, the converse is not true.

(Mutual Independence $\underset{\not\Leftarrow}{\Rightarrow}$ Pairwise Independence)

Example.



A_1, A_2, $\boxed{A_3}$

are mutually independent, and pairwise independent.

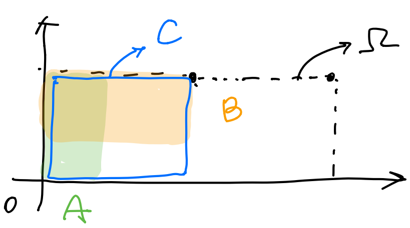
$P(A_1 \cap A_2 \cap A_3) = \frac{1}{8} = P(A_1)P(A_2)P(A_3) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}$

iii) Conditional Independence.

A and B are conditionally independent given C

if $P(A \cap B | C) = P(A|C)P(B|C)$ and $P(C) > 0$.

Example



A and B are not independent, but conditionally independent given $\boxed{C}$.

Remark. Independence $\underset{\not\Leftarrow}{\not\Rightarrow}$ Conditional Independence.

iv) (Mutual) Independence of an infinite sequence of events

Infinite set of events $A_1, A_2, \cdots,$ are (mutually) independent if for any finite subset $S \subset \{1, 2, \cdots\}$, $P(\bigcap_{i \in S} A_i) = \prod_{i \in S} P(A_i)$.

↳ Avoiding circular definition.

Note Nearly the same compared to ii).

v) Conditional Independence of multiple events.

Finite set of events $A_1, A_2, \cdots, A_n$ are conditionally independent given C if for any finite subset $S \subset \{1, 2, \cdots, n\}$,

$P(\bigcap_{i \in S} A_i | C) = \prod_{i \in S} P(A_i | C)$ and $P(C) > 0$.

Prop. Suppose $P(B) > 0$.

Then A, B are independent if and only if $P(A|B) = P(A)$.

pf) ($\Rightarrow$) $P(A \cap B) = P(A) P(B)$, $P(B) > 0 \Rightarrow P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A)$.

($\Leftarrow$) $P(A) = \underline{P(A|B)} = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A) P(B) = P(A \cap B)$. ∎

↳ Implies $P(B) > 0$.

Example $\Omega = \{1, 2, 3, 4\}$. $\mathcal{A} = 2^{\Omega}$. $P(k) = \frac{1}{4}$ $\forall k \in \Omega$.

$A = \{1, 2\}$. $B = \{1, 3\}$, $C = \{2, 3\}$.

i) Are A, B independent?

pf) $P(A) \cdot P(B) = \frac{1}{2} \cdot \frac{1}{2} = P(A \cap B) = \frac{1}{4}$.

$\therefore$ A and B are independent. ∎

ii) Are A, B, C pairwise independent?

pf) $P(A) \cdot P(B) = \frac{1}{2} \cdot \frac{1}{2} = P(A \cap B) = \frac{1}{4}$. A and B are independent,

$P(A) \cdot P(C) = \frac{1}{2} \cdot \frac{1}{2} = P(A \cap C) = \frac{1}{4}$. A and C are independent,

$P(B) \cdot P(C) = \frac{1}{2} \cdot \frac{1}{2} = P(B \cap C) = \frac{1}{4}$. B and C are independent,

$\therefore$ A, B, C is pairwise independent. ▨

iii) Are A, B, C mutually independent?

pf) $P(A)P(B)P(C) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \neq P(A \cap B \cap C) = 0$.

$\therefore$ A, B, C is not mutually independent. ▨

iv) Are A, B conditionally independent given C?

pf) $P(A \cap B \mid C) = \dfrac{P(A \cap B \cap C)}{P(C)} = 0$

$$P(A \mid C) P(B \mid C) = \frac{P(A \cap C)}{P(C)} \cdot \frac{P(B \cap C)}{P(C)} = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \neq P(A \cap B \mid C)$$

$\therefore$ A, B are not conditionally independent given C. ▨

2.4-5 Three Essential Rules for Probability

Remark. $\boxed{P(A \cap B) = P(A \mid B) P(B)}$ if $P(B) > 0$.

Thm. **Bayes' Rule**

$$P(B \mid A) = \frac{P(A \mid B) P(B)}{P(A)} \quad \text{if } P(A) > 0 \text{ and } P(B) > 0.$$

Remark. Bayesian Statistics.

For example, let's say that data we observe are determined by some event A; and probabilistic model with parameters determined by some event B, When we define a probabilistic model, we often find in terms of the probability of seeing a particular data, given some parameter values.

Bayes' rule allows us to invert the conditional probabilities in subsets, that is, if we observe some data, we can infer the probability of parameter values given the data.

Thm. Chain Rule.

If for events $A_1, A_2, \cdots A_n$ and $P(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) > 0$,

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1) P(A_2 | A_1) P(A_3 | A_1 \cap A_2) \cdots P(A_n | A_1 \cap A_2 \cap \cdots \cap A_{n-1}).$$

pf). By mathematical induction.

If $n=2$, $P(A_1 \cap A_2) = P(A_1) P(A_2 | A_1)$, so the chain rule holds.

Suppose the chain rule holds at $n-1$.

Let $B = A_1 \cap A_2 \cap \cdots \cap A_{n-1}$.

$$P(B) = P(A_1) P(A_2 | A_1) \cdots P(A_{n-1} | A_1 \cap A_2 \cap \cdots \cap A_{n-2}).$$

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(B \cap A_n) = P(A_n | B) P(B)$$
$$= P(A_1) P(A_2 | A_1) \cdots P(A_n | A_1 \cap A_2 \cap \cdots \cap A_{n-1}).$$

So the chain rule holds at n. ▨

<u>Remark</u>. Well-definedness of the chain rule.

$P(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) > 0$

$\Rightarrow P(A_1) > 0,\ P(A_1 \cap A_2) > 0, \cdots P(A_1 \cap \cdots \cap A_{n-1}) > 0$ by monotonicity.

$\Rightarrow P(A_2 | A_1),\ P(A_3 | A_1 \cap A_2), \cdots, P(A_n | A_1 \cap A_2 \cap \cdots \cap A_{n-1})$ is well-defined.

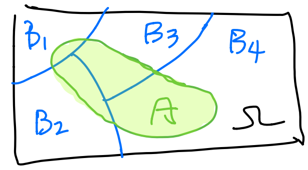
Finite or Countably infinite.
e.g. $\{B_1, B_2, \cdots, B_n\}$ or $\{B_1, B_2, \cdots\}$.

<u>Def</u>. A partition of Ω

is a collection $\{B_i\}_i \subset 2^{\Omega}$

They are pairwise disjoint.

Such that i) $\bigcup_i B_i = \Omega$ and ii) $B_i \cap B_j = \phi \quad (i \neq j)$



<u>Thm</u>. Partition Rule.

$P(A) = \sum_i P(A \cap B_i)$ for any partition $\{B_i\}$ of Ω.

pf) $A = A \cap \Omega = A \cap \left(\bigcup_i B_i\right) = \bigcup_i (A \cap B_i)$.

countable additivity. $(A \cap B_i) \cap (A \cap B_j) = \phi \quad (i \neq j)$.

$$P(A) = P\left(\bigcup_i (A \cap B_i)\right) = \sum_i P(A \cap B_i)$$ ▨

<u>Prop</u>. Conditional probability measure given B.

If $P(B) > 0$, then $Q(A) = P(A|B)$ defines a probability measure Q.

$\rightsquigarrow$ Every "tools" developed for probability measure can be applied to conditional probability measure.

pf) i) $Q(\phi) = P(\phi | B) = P(\phi \cap B)/P(B) = P(\phi)/P(B) = 0$ ▨

ii) For pairwise disjoint sets $E_1, E_2, \cdots, E_n$,

$$Q\left(\bigcup_{i=1}^{\infty} E_i\right) = P\left(\bigcup_{i=1}^{\infty} E_i \,\middle|\, B\right) = P\left(\left(\bigcup_{i=1}^{\infty} A_i\right) \cap B\right) / P(B) = P\left(\bigcup_{i=1}^{\infty} A_i \cap B\right) / P(B)$$
$$= \bigcup_{i=1}^{\infty} P(A_i \cap B) / P(B) = \bigcup_{i=1}^{\infty} Q(A_i)$$ ▨

iii) $Q(\Omega) = P(\Omega | B) = P(\Omega \cap B)/P(B) = P(B)/P(B) = 1$ ▨

$\therefore$ By i), ii), Q is a measure, and by iii), Q is a probability measure. ▨