Statistical Machine Translation
Probability Theory I

Jakub Waszczuk

Heinrich Heine Universität Düsseldorf

Winter Semester 2018/19

Today

Our goals for today:

- Motivations and intuitions behind probability theory
- Probability space: definition, some examples
- Basic calculation rules:
  - Sum rule
  - Product rule

Example (following E. T. Janyes)

Let's consider the following situation:

> *Some dark night a policeman walks down a street, apparently deserted. Suddenly he hears a burglar alarm, looks across the street, and sees a jewelry store with a broken window. Then a gentleman wearing a mask comes crawling out through the broken window, carrying a bag which turns out to be full of expensive jewelry.*

Example (following E. T. Janyes)

Let's consider the following situation:

> *Some dark night a policeman walks down a street, apparently deserted. Suddenly he hears a burglar alarm, looks across the street, and sees a jewelry store with a broken window. Then a gentleman wearing a mask comes crawling out through the broken window, carrying a bag which turns out to be full of expensive jewelry.*

Question: is the gentleman wearing a mask dishonest?

Example (following E. T. Janyes)

Let's consider the following situation:

> *Some dark night a policeman walks down a street, apparently deserted. Suddenly he hears a burglar alarm, looks across the street, and sees a jewelry store with a broken window. Then a gentleman wearing a mask comes crawling out through the broken window, carrying a bag which turns out to be full of expensive jewelry.*

Question: is the gentleman wearing a mask dishonest? Hints:

- Street deserted
- In the dark of night
- Burglar alarm goes off
- Broken window
- The gentleman crawls out from the jewelry store
- With a bag full of jewelry
- He wears a mask

Example (following E. T. Janyes)

Is this a logical conclusion?

$$\text{street deserted} \wedge \text{in the dark of night} \wedge \text{burglar alarm goes off} \wedge$$
$$\text{broken window} \wedge \text{gentleman crawls out from the jewelry store} \wedge$$
$$\text{with a bag full of jewelry} \wedge \text{he wears a mask}$$
$$\vdots$$
$$\text{the gentelement with the mask is dishonest}$$

Example (following E. T. Janyes)

Is this a logical conclusion?

$$\text{street deserted} \wedge \text{in the dark of night} \wedge \text{burglar alarm goes off} \wedge$$
$$\text{broken window} \wedge \text{gentleman crawls out from the jewelry store} \wedge$$
$$\text{with a bag full of jewelry} \wedge \text{he wears a mask}$$
$$\vdots$$
$$\text{the gentelement with the mask is dishonest}$$

There is a trace of doubt...

It might be that this gentleman was the owner of the jewelry store and he was coming home from a masquerade party, and didn't have the key with him. However, just as he walked by his store, a passing truck threw a stone through the window, and he was only protecting his own property.

Deductive vs plausible reasoning

Plausible reasoning

We would rather want to determine the *plausibility* of the conclusion:

$$P\Big(\text{the gentelement with the mask is dishonest} \,\Big|\, \text{street deserted} \wedge \text{in the dark of night} \wedge \text{burglar alarm goes off} \wedge \text{broken window} \wedge \text{gentleman crawls out from the jewelry store} \wedge \text{with a bag full of jewelry} \wedge \text{he wears a mask}\Big)$$

Deductive vs plausible reasoning

Plausible reasoning

We would rather want to determine the *plausibility* of the conclusion:

$$P\Big(\text{the gentelement with the mask is dishonest} \;\Big|\; \text{street deserted} \land \text{in the dark of night} \land \text{burglar alarm goes off} \land \text{broken window} \land \text{gentleman crawls out from the jewelry store} \land \text{with a bag full of jewelry} \land \text{he wears a mask}\Big)$$

Desiderata

A theory of plausible/rational reasoning:

- Events: $A, B, C, D, E, \ldots$
- Function P which describes the plausibility of events
- Language for describing complex events: $\land, \lor, \neg, \implies, \ldots$
- Numerical interpretation of the operators: $P(A \land B) = \ldots$, $P(\neg A) = \ldots$, etc.
- This theory should follow some basic intuitions on rational reasoning.

Home-baked theory of rational reasoning

Proposition

Probability $P(A)$ of an event A is within the range of $[0, 1]$, with the following interpretation:

1. *0 $\cong$ impossibility*
2. *1 $\cong$ certainty*
3. *graded scale between impossible and certain*

Home-baked theory of rational reasoning

Some desirable properties

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

Home-baked theory of rational reasoning

Some desirable properties

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

$$P(A \text{ and } B) = P(B \text{ and } A) \tag{2}$$

Home-baked theory of rational reasoning

Some desirable properties

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

$$P(A \text{ and } B) = P(B \text{ and } A) \tag{2}$$

$$P(A) \leq P(A \text{ or } B) \tag{3}$$

Home-baked theory of rational reasoning

Some desirable properties

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

$$P(A \text{ and } B) = P(B \text{ and } A) \tag{2}$$

$$P(A) \leq P(A \text{ or } B) \tag{3}$$

$$P(B) \leq P(A \text{ or } B) \tag{4}$$

Home-baked theory of rational reasoning

Some desirable properties

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

$$P(A \text{ and } B) = P(B \text{ and } A) \tag{2}$$

$$P(A) \leq P(A \text{ or } B) \tag{3}$$

$$P(B) \leq P(A \text{ or } B) \tag{4}$$

$$P(A \text{ and } B) \leq P(A) \tag{5}$$

Home-baked theory of rational reasoning

Some desirable properties

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

$$P(A \text{ and } B) = P(B \text{ and } A) \tag{2}$$

$$P(A) \leq P(A \text{ or } B) \tag{3}$$

$$P(B) \leq P(A \text{ or } B) \tag{4}$$

$$P(A \text{ and } B) \leq P(A) \tag{5}$$

$$\ldots \tag{6}$$

Home-baked theory of rational reasoning

Some desirable properties

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

$$P(A \text{ and } B) = P(B \text{ and } A) \tag{2}$$

$$P(A) \leq P(A \text{ or } B) \tag{3}$$

$$P(B) \leq P(A \text{ or } B) \tag{4}$$

$$P(A \text{ and } B) \leq P(A) \tag{5}$$

$$\ldots \tag{6}$$

Notation

Instead of using $\wedge$ and $\vee$ for conjunction and disjunction of events, we will now use:

- $\cap$ for conjunction ($A \cap B \;\equiv\;$ A occurs *and* B occurs)
- $\cup$ for disjunction ($A \cup B \;\equiv\;$ A occurs *or* B occurs)

Home-baked theory of rational reasoning

Definition

Let $\perp$ be an event that is completely impossible, i.e., $P(\perp) = 0$.

Home-baked theory of rational reasoning

Definition

Let $\perp$ be an event that is completely impossible, i.e., $P(\perp) = 0$.

Example

If you roll a 6-numbered dice, it lands on neither 1, 2, . . . , nor 6.

Home-baked theory of rational reasoning

Definition

Let $\bot$ be an event that is completely impossible, i.e., $P(\bot) = 0$.

Example

If you roll a 6-numbered dice, it lands on neither 1, 2, . . . , nor 6.

Proposition

Then:

$$P(A \cap \bot) = P(\bot) = 0 \tag{7}$$

$$P(A \cup \bot) = P(A) \tag{8}$$

In words: $\bot$ is an absorbing element of conjunction of and neutral element of disjunction.

Home-baked theory of rational reasoning

Definition

Conversely, let $\top$ *be an event that is 100% probable, i.e.,* $P(\top) = 1$.

Home-baked theory of rational reasoning

Definition

Conversely, let $\top$ *be an event that is 100% probable, i.e.,* $P(\top) = 1$*.*

Example

Example: if you through a 6-numbered dice, it gives a number between 1 and 6.

Home-baked theory of rational reasoning

Definition

Conversely, let $\top$ be an event that is 100% probable, i.e., $P(\top) = 1$.

Example

Example: if you through a 6-numbered dice, it gives a number between 1 and 6.

Proposition

Then:

$$P(A \cap \top) = P(A) \tag{9}$$

$$P(A \cup \top) = P(\top) = 1 \tag{10}$$

In words: $\bot$ is a neutral element of conjunction and absorbing element of disjunction.

Home-baked theory of rational reasoning

The operations which follow the laws we devised are $+$ for $\cup$ and $\times$ for $\cap$:

$$P(A \cup B) = P(A) + P(B) \geq P(A)$$
$$P(A \cap B) = P(A) \times P(B) \leq P(A)$$

Home-baked theory of rational reasoning

The operations which follow the laws we devised are $+$ for $\cup$ and $\times$ for $\cap$:

$$P(A \cup B) = P(A) + P(B) \geq P(A)$$
$$P(A \cap B) = P(A) \times P(B) \leq P(A)$$

$$P(A \cap \bot) = P(A) \times P(\bot) = P(A) \times 0 = 0$$
$$P(A \cup \bot) = P(A) + P(\bot) = P(A) + 0 = P(A)$$

Home-baked theory of rational reasoning

The operations which follow the laws we devised are $+$ for $\cup$ and $\times$ for $\cap$:

$$P(A \cup B) = P(A) + P(B) \geq P(A)$$
$$P(A \cap B) = P(A) \times P(B) \leq P(A)$$

$$P(A \cap \bot) = P(A) \times P(\bot) = P(A) \times 0 = 0$$
$$P(A \cup \bot) = P(A) + P(\bot) = P(A) + 0 = P(A)$$

$$P(A \cap \top) = P(A) \times P(\top) = P(A) \times 1 = P(A)$$
$$P(A \cup \top) = P(A) + P(\top) = P(A) + 1 = 1$$

Home-baked theory of rational reasoning

The operations which follow the laws we devised are $+$ for $\cup$ and $\times$ for $\cap$:

$$P(A \cup B) = P(A) + P(B) \geq P(A)$$
$$P(A \cap B) = P(A) \times P(B) \leq P(A)$$

$$P(A \cap \bot) = P(A) \times P(\bot) = P(A) \times 0 = 0$$
$$P(A \cup \bot) = P(A) + P(\bot) = P(A) + 0 = P(A)$$

$$P(A \cap \top) = P(A) \times P(\top) = P(A) \times 1 = P(A)$$
$$P(A \cup \top) = P(A) + P(\top) = P(A) + 1 = 1$$

Not really.

Home-baked theory of rational reasoning

Other problems

$$P(A \cap A) = P(A) \neq P(A) \times P(A)$$
$$P(A \cup A) = P(A) \neq P(A) + P(A)$$

Home-baked theory of rational reasoning

Other problems

$$P(A \cap A) = P(A) \neq P(A) \times P(A)$$
$$P(A \cup A) = P(A) \neq P(A) + P(A)$$

Notes

- Conjunction and disjunction are idempotent – $A \cup A = A$ and $A \cap A = A$
- Multiplication and addition are not idempotent
- The matter is more complex than it seemed...
- Fortunatelly, there is a wonderfully elegant solution satisfies all the necessary laws

Boolean algebra

Definition

Let M be a set. We say that $\mathcal{A} \subseteq \wp(M)$ is a Boolean algebra over M if:

1. $M \in \mathcal{A}, \emptyset \in \mathcal{A}$
2. *if $A \in \mathcal{A}$, then also $\overline{A} \in \mathcal{A}$, where $\overline{A} := M \setminus A$*
3. *if $A, B \in \mathcal{A}$, then also $A \cup B \in \mathcal{A}$*

Boolean algebra

Definition

Let M be a set. We say that $\mathcal{A} \subseteq \wp(M)$ is a Boolean algebra over M if:

1. $M \in \mathcal{A}, \emptyset \in \mathcal{A}$
2. *if* $A \in \mathcal{A}$*, then also* $\overline{A} \in \mathcal{A}$*, where* $\overline{A} := M \setminus A$
3. *if* $A, B \in \mathcal{A}$*, then also* $A \cup B \in \mathcal{A}$

What about $A \cap B$?

Boolean algebra

Definition

Let M be a set. We say that $\mathcal{A} \subseteq \wp(M)$ is a Boolean algebra over M if:

1. $M \in \mathcal{A}, \emptyset \in \mathcal{A}$
2. *if* $A \in \mathcal{A}$, *then also* $\overline{A} \in \mathcal{A}$, *where* $\overline{A} := M \setminus A$
3. *if* $A, B \in \mathcal{A}$, *then also* $A \cup B \in \mathcal{A}$

What about $A \cap B$?

1. From definition, if $A, B \in \mathcal{A}$, then $\overline{\overline{A} \cup \overline{B}} \in \mathcal{A}$

Boolean algebra

Definition

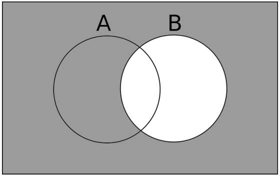
Let M be a set. We say that $\mathcal{A} \subseteq \wp(M)$ is a Boolean algebra over M if:

1. $M \in \mathcal{A}, \emptyset \in \mathcal{A}$
2. *if* $A \in \mathcal{A}$, *then also* $\overline{A} \in \mathcal{A}$, *where* $\overline{A} := M \setminus A$
3. *if* $A, B \in \mathcal{A}$, *then also* $A \cup B \in \mathcal{A}$

What about $A \cap B$?

1. From definition, if $A, B \in \mathcal{A}$, then $\overline{\overline{A} \cup \overline{B}} \in \mathcal{A}$
2. $A \cap B = \overline{\overline{A} \cup \overline{B}}$

$\overline{\text{B}}$



Boolean algebra

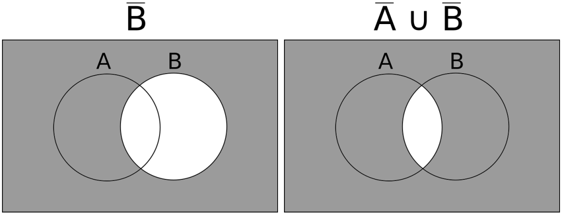
Definition

Let M be a set. We say that $\mathcal{A} \subseteq \wp(M)$ is a Boolean algebra over M if:

1. $M \in \mathcal{A}, \emptyset \in \mathcal{A}$
2. *if* $A \in \mathcal{A}$*, then also* $\overline{A} \in \mathcal{A}$*, where* $\overline{A} := M \setminus A$
3. *if* $A, B \in \mathcal{A}$*, then also* $A \cup B \in \mathcal{A}$

What about $A \cap B$?

1. From definition, if $A, B \in \mathcal{A}$, then $\overline{\overline{A} \cup \overline{B}} \in \mathcal{A}$
2. $A \cap B = \overline{\overline{A} \cup \overline{B}}$



Boolean algebra

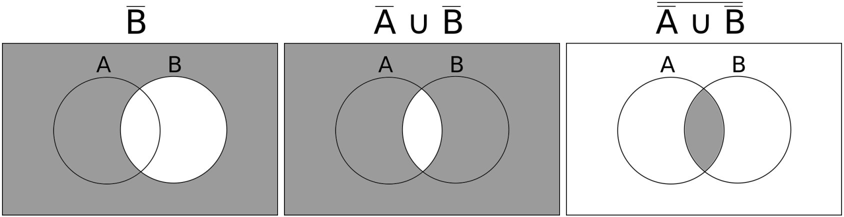
Definition

Let M be a set. We say that $\mathcal{A} \subseteq \wp(M)$ is a Boolean algebra over M if:

1. $M \in \mathcal{A}, \emptyset \in \mathcal{A}$
2. *if* $A \in \mathcal{A}$, *then also* $\overline{A} \in \mathcal{A}$, *where* $\overline{A} := M \setminus A$
3. *if* $A, B \in \mathcal{A}$, *then also* $A \cup B \in \mathcal{A}$

What about $A \cap B$?

1. From definition, if $A, B \in \mathcal{A}$, then $\overline{\overline{A} \cup \overline{B}} \in \mathcal{A}$
2. $A \cap B = \overline{\overline{A} \cup \overline{B}}$



Boolean algebra

Why should we care?

- Boolean algebras generalize both set and logic operations:
  1. $\cup \equiv \vee$
  2. $\cap \equiv \wedge$
  3. $\neg \equiv$ complement ($\overline{A}$)

- They capture the essential properties of set and logical operations:
  1. $A \cup B = B \cup A, \quad A \cap B = B \cap A$
  2. $\overline{\overline{A}} = A$
  3. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
  4. $\overline{A \cup B} = \overline{A} \cap \overline{B}, \quad \overline{A \cap B} = \overline{A} \cup \overline{B}$
  5. ...

- This gives us a language with desirable properties, which allows us to describe complex probabilistic events.

Boolean algebra: exercise

Dice

Let $M = \{1, 2, 3, 4, 5, 6\}$ be the set of possible outcomes of rolling a dice.

Tasks

- Assuming that every outcome is possible, determine the corresponding Boolean algebra of events $\mathscr{A}$.
- Is it possible to construct another Boolean alebra for M?

Probability space

Definition

*A probability space is a triple $(\Omega, \mathfrak{A}, P)$, where $\mathfrak{A} \subseteq \wp(\Omega)$ is a Boolean algebra (which represents the possible **events**), and $P : \mathfrak{A} \rightarrow [0, 1]$ is a probability function such that:*

1. $P(\Omega) = 1$;
2. $P(\emptyset) = 0$, *and*
3. *if $A_1, A_2, ..., A_n$ is a sequence of pairwise disjoint sets (mutually exclusive events), then:*

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) \qquad (11)$$

Probability space

Definition

A probability space is a triple $(\Omega, \mathfrak{A}, P)$, *where* $\mathfrak{A} \subseteq \wp(\Omega)$ *is a Boolean algebra (which represents the possible* ***events****), and* $P : \mathfrak{A} \to [0, 1]$ *is a probability function such that:*

1. $P(\Omega) = 1$*;*
2. $P(\emptyset) = 0$*, and*
3. *if* $A_1, A_2, ..., A_n$ *is a sequence of pairwise disjoint sets (mutually exclusive events), then:*

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) \tag{11}$$

Note that (3) *implies that for two given* $A, B \in \mathfrak{A}$ *such that* $A \cap B = \emptyset$*:*

$$P(A \cup B) = P(A) + P(B) \tag{12}$$

Probability space

Definition

A probability space is a triple $(\Omega, \mathfrak{A}, P)$*, where* $\mathfrak{A} \subseteq \wp(\Omega)$ *is a Boolean algebra (which represents the possible* ***events****), and* $P : \mathfrak{A} \to [0, 1]$ *is a probability function such that:*

1. $P(\Omega) = 1$*;*
2. $P(\emptyset) = 0$*, and*
3. *if* $A_1, A_2, ..., A_n$ *is a sequence of pairwise disjoint sets (mutually exclusive events), then:*

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) \tag{11}$$

Note that (3) *implies that for two given* $A, B \in \mathfrak{A}$ *such that* $A \cap B = \emptyset$*:*

$$P(A \cup B) = P(A) + P(B) \tag{12}$$

Note

Note one of the roles of the Boolean algebra: if $A, B \subseteq \Omega$ are events (to which probabilities are assigned), than $A \cup B$ (i.e., A or B occur), $A \cap B$ (i.e., both A and B occur) and $\overline{A}$ (i.e. A does not occur) are also events with assigned probabilities.

Probability spaces

Laplace

- The set of outcomes Ω is finite
- Every outcome $\omega \in \Omega$ is possible and equally likely

$$p(\omega) = \frac{1}{|\Omega|} \tag{13}$$

Probability spaces

Laplace

- The set of outcomes Ω is finite
- Every outcome $\omega \in \Omega$ is possible and equally likely

$$p(\omega) = \frac{1}{|\Omega|} \tag{13}$$

Example

An example is a fair cube (dice) with n sides:

- n is arbitrary, but must be finite (and > 0)
- Question: is each event equally likely?

Probability spaces

Laplace

- The set of outcomes Ω is finite
- Every outcome $\omega \in \Omega$ is possible and equally likely

$$p(\omega) = \frac{1}{|\Omega|} \tag{13}$$

Example

An example is a fair cube (dice) with n sides:

- n is arbitrary, but must be finite (and > 0)
- Question: is each event equally likely?

Exercise

Suppose you roll two 6-faced dices.

- How does the event of getting 2 with the first dice look like?

Probability spaces

Bernoulli

The probability space of the Bernoulli distribution has only two outcomes, 0 or 1:

- $\Omega = \{0, 1\}$
- $\mathfrak{A} = \wp(\Omega) = \{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$
- $p(1) = 1 - p(0)$

Example

The typical example is the event of coin tossing, with a possibly unfair coin ($p(1) \neq p(0)$).

Probability spaces

Discrete

Generalization of Laplace and Bernoulli:

- Ω is finite
- $\mathfrak{A} = \wp(\Omega)$, i.e., every conceivable event has a probability

Example

Suppose you throw a coin *n* times and that the probability of getting *heads* is *h*. What is the probability of throwing heads *k* times?

- $\Omega = \{0, 1, \ldots, N\}$
- The probability of throwing heads *k* times:

$$p(k) = h^k \times (1 - h)^{(n-k)} \times \binom{n}{k} \tag{14}$$

Exercise

Setup

- Hans has three children
- The probability of having a boy is $\frac{1}{2}$

Task

Answer the question:

- What's the probability that Hans has exactly one boy?

To this end:

- Determine the underlying probability space.
- Determine the event of Hans having exactly one boy.

Calculation rules

We now proceed to determine the calculation rules for the operators which allow us to build complex event exressions:

- complement ($\overline{A}$)
- $\cup$
- $\cap$

Calculation rules

Complement

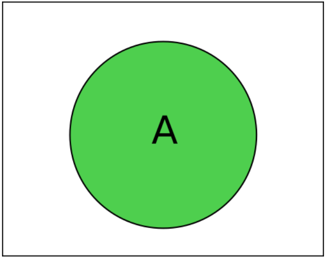
$$P(\overline{A}) = 1 - P(A) \tag{15}$$

Calculation rules

Complement

$$P(\overline{A}) = 1 - P(A) \quad (15)$$

Venn diagram: $P(A)$

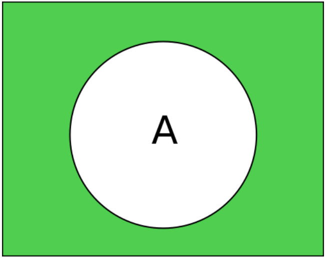


Calculation rules

Complement

$$P(\overline{A}) = 1 - P(A) \tag{15}$$

Venn diagram: $P(\overline{A})$



Calculation rules

Complement

$$P(\overline{A}) = 1 - P(A) \tag{15}$$

Derivation

$$P(A \cup \overline{A}) = P(\Omega) = 1, \text{ and}$$
$$A \cap \overline{A} = \emptyset$$

Therefore:

$$\begin{aligned} 1 &= P(A \cup \overline{A}) \\ &= P(A) + P(\overline{A}) \\ \iff P(A) &= 1 - P(\overline{A}) \end{aligned}$$

Calculation rules

Sum rule

The sum rule allows us to interpret the logical disjunction arithmetically:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{16}$$

Calculation rules

Sum rule

The sum rule allows us to interpret the logical disjunction arithmetically:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{16}$$

This is harder to derive... but still follows directly from the set theory and probability axioms.

Calculation rules

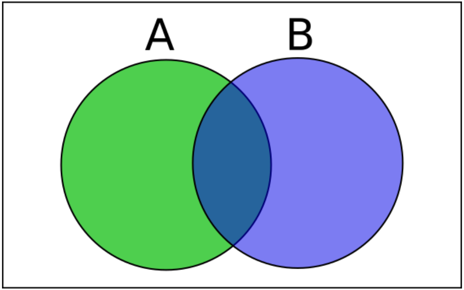
Sum rule

The sum rule allows us to interpret the logical disjunction arithmetically:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{16}$$

This is harder to derive... but still follows directly from the set theory and probability axioms.

Venn diagram



Calculation rules

Deriving the sum rule (for the record)

Using Venn diagrams, it's easy to verify that:

$$A \cup B = A \,\dot{\cup}\, (B \setminus A) \tag{17}$$

$$B \setminus A = B \setminus (A \cap B) \tag{18}$$

$$B = B \setminus (A \cap B) \,\dot{\cup}\, (A \cap B) \tag{19}$$

Calculation rules

Deriving the sum rule (for the record)

Using Venn diagrams, it's easy to verify that:

$$A \cup B = A \mathbin{\dot\cup} (B \setminus A) \tag{17}$$
$$B \setminus A = B \setminus (A \cap B) \tag{18}$$
$$B = B \setminus (A \cap B) \mathbin{\dot\cup} (A \cap B) \tag{19}$$

From which follow:

$$P(A \cup B) = P(A) + P(B \setminus (A \cap B)) \qquad \text{(via Eq. 17, 18)}$$
$$P(B) = P(B \setminus (A \cap B)) + P(A \cap B) \qquad \text{(via Eq. 19)}$$

Calculation rules

Deriving the sum rule (for the record)

Using Venn diagrams, it's easy to verify that:

$$A \cup B = A \,\dot{\cup}\, (B \setminus A) \tag{17}$$

$$B \setminus A = B \setminus (A \cap B) \tag{18}$$

$$B = B \setminus (A \cap B) \,\dot{\cup}\, (A \cap B) \tag{19}$$

From which follow:

$$P(A \cup B) = P(A) + P(B \setminus (A \cap B)) \qquad \text{(via Eq. 17, 18)}$$

$$P(B) = P(B \setminus (A \cap B)) + P(A \cap B) \qquad \text{(via Eq. 19)}$$

Eliminating $P(B \setminus (A \cap B))$ from both equations above gives the sum rule.

Conditional probability: intuition

Setup

- Hans has three children
- The probability of having a boy is $\frac{1}{2}$

Conditional probability: intuition

Setup

- Hans has three children
- The probability of having a boy is $\frac{1}{2}$

Update

- Suppose we know, that Hans has a daugher.

Conditional probability: intuition

Setup

- Hans has three children
- The probability of having a boy is $\frac{1}{2}$

Update

- Suppose we know, that Hans has a daugher.

Question

- What is the probability that he has exactly one son in this case?

Conditional probability: intuition

Proposition

Let:

- $(\Omega, \mathfrak{A}, P)$ *be a probability space*
- $C \in \mathfrak{A}$ *be the event that we know has occurred*

We design a modified probability space $(\Omega_C, \mathfrak{A}_C, P_C)$ *as follows:*

- $\Omega_C = C$ *(*Ω *restricted to outcomes consistent with* C*)*
- $\mathfrak{A}_C \subseteq \wp(C)$
- $P_C = P|_C$ *(*P *with domain restricted to* C*)*

Conditional probability: intuition

Proposition

Let:

- $(\Omega, \mathfrak{A}, P)$ *be a probability space*
- $C \in \mathfrak{A}$ *be the event that we know has occurred*

We design a modified probability space $(\Omega_C, \mathfrak{A}_C, P_C)$ *as follows:*

- $\Omega_C = C$ *(Ω restricted to outcomes consistent with C)*
- $\mathfrak{A}_C \subseteq \wp(C)$
- $P_C = P|_C$ *(P with domain restricted to C)*

Issues

- $P_C(\Omega_C) \neq 1$
- The underlying sample space and algebra of events are changed
- The new probability does not tell us anything about events partially out of C

Conditional probability: derivation

Assumptions

- $P_C(C) =$
- if $A \subseteq \overline{C}$, then $P_C(A) =$
- if $A \subseteq C$, then $P_C(A) =$

Conditional probability: derivation

Assumptions

- $P_C(C) = 1$
- if $A \subseteq \overline{C}$, then $P_C(A) =$
- if $A \subseteq C$, then $P_C(A) =$

Conditional probability: derivation

Assumptions

- $P_C(C) = 1$
- if $A \subseteq \overline{C}$, then $P_C(A) = 0$
- if $A \subseteq C$, then $P_C(A) =$

Conditional probability: derivation

Assumptions

- $P_C(C) = 1$
- if $A \subseteq \overline{C}$, then $P_C(A) = 0$
- if $A \subseteq C$, then $P_C(A) = \alpha P(A)$, where α is a normalization constant

Conditional probability: derivation

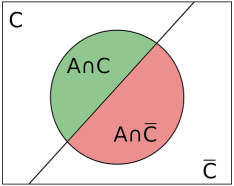
Assumptions

- $P_C(C) = 1$
- if $A \subseteq \overline{C}$, then $P_C(A) = 0$
- if $A \subseteq C$, then $P_C(A) = \alpha P(A)$, where α is a normalization constant

Derivation

If $A \subseteq C \cup \overline{C}$:

$$\begin{aligned} P_C(A) &= P_C((A \cap C) \cup (A \cap \overline{C})) \\ &= P_C(A \cap C) + P_C(A \cap \overline{C}) \\ &= P_C(A \cap C) = \alpha P(A \cap C) \end{aligned}$$



Conditional probability: derivation

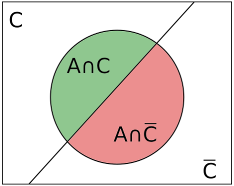
Assumptions

- $P_C(C) = 1$
- if $A \subseteq \overline{C}$, then $P_C(A) = 0$
- if $A \subseteq C$, then $P_C(A) = \alpha P(A)$, where α is a normalization constant

Derivation

If $A \subseteq C \cup \overline{C}$:

$$\begin{aligned} P_C(A) &= P_C((A \cap C) \cup (A \cap \overline{C})) \\ &= P_C(A \cap C) + P_C(A \cap \overline{C}) \\ &= P_C(A \cap C) = \alpha P(A \cap C) \end{aligned}$$



Therefore:

$$P_C(C) = \alpha P(C \cap C) = \alpha P(C) = 1 \implies \alpha = \frac{1}{P(C)}$$

Conditional probability: derivation

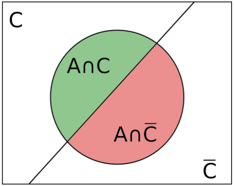
Assumptions

- $P_C(C) = 1$
- if $A \subseteq \overline{C}$, then $P_C(A) = 0$
- if $A \subseteq C$, then $P_C(A) = \alpha P(A)$, where α is a normalization constant

Derivation

If $A \subseteq C \cup \overline{C}$:

$$\begin{aligned} P_C(A) &= P_C((A \cap C) \cup (A \cap \overline{C})) \\ &= P_C(A \cap C) + P_C(A \cap \overline{C}) \\ &= P_C(A \cap C) = \alpha P(A \cap C) \end{aligned}$$



Therefore:

$$P_C(C) = \alpha P(C \cap C) = \alpha P(C) = 1 \implies \alpha = \frac{1}{P(C)}$$

Finally:

$$P_C(A) = \frac{P(A \cap C)}{P(C)}$$

Conditional probability: definition

Notation

We denote with $P(A|C)$ the probability of event $A \in \mathfrak{A}$ in a context where we now that $C \in \mathfrak{A}$ has occurred.

Definition

The conditional probability of $A|C$ is defined as:

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

Product rule

Definition

The product rule stems directly from the conditional probability $P(A|B) = \frac{P(A \cap B)}{P(B)}$*:*

$$P(A \cap B) = P(A|B)P(B) \tag{20}$$

Product rule

Definition

The product rule stems directly from the conditional probability $P(A|B) = \frac{P(A \cap B)}{P(B)}$*:*

$$P(A \cap B) = P(A|B)P(B) \tag{20}$$

Since $\cap$ *is commutative* ($A \cap B = B \cap A$)*, it also holds that:*

$$P(A \cap B) = P(B \cap A) = P(B|A)P(A) \tag{21}$$

Recap

- Probability theory can be seen as a theory of rational reasoning
- In this theory, events are modeled with a Boolean algebra, which:
  - Comprises operators ($\cup$, $\cap$, etc.) for defining complex events
  - Satisfies the intuitive laws of rational reasoning (e.g., $P(A) \leq (A \cup B)$)
- The probability itself is modeled as a function from events to $[0, 1]$
- In case of complex events, it can be calculated based on:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$P(A \cap B) = P(A|B)P(B)$$