# Machine Learning

## Section 3: From Logic to Probabilities

## 第3节：从逻辑到概率

Stefan Harmeling

13. October 2021

**Today:**

What exactly do we mean by $p(A)$ or $p(A, B)$, etc?

今天。
我们所说的p(A)或p(A, B)
等到底是什么意思？

# Let’s define...

## Syntax

- What are allowed strings? 什么是允许的字符串？
- e.g. $A \wedge B$, $A$, $A \rightarrow B$, $A \vee \neg B \vee C$

# Let's define...

## Syntax

- What are allowed strings?
- e.g. $A \wedge B$, $A$, $A \rightarrow B$, $A \vee \neg B \vee C$

## Semantics

- What do these strings mean?
- i.e. when is $A \wedge B$ true, when is it false?

这些字符串是什么意思?
即：A , B何时为真，何时为假?

# Propositional logic (1) — syntax

**Definition 3.1 (alphabet)**

*The alphabet* $\mathcal{A} = \mathcal{V} \cup \{\neg, \wedge, \vee, \rightarrow, ), (\}$ *consists of various symbols:*

- *a finite set* $\mathcal{V}$ *of symbols* $X, Y, \ldots$*; aka non-logical symbols, propositional variables*
- *junctors* $\neg, \wedge, \vee, \rightarrow$*; aka logical symbols, connectives, names of truth functions*
- *parenthesis* ), (

字母表A = V 8 {¬, , -, ), (}由各种符号组成。
符号的有限集合V X , Y , . . . ; 又称非逻辑符号，命题变量
junctors ¬, , , -, ; 又称逻辑符号，连接词，真值函数的名称

括号内），（

# Propositional logic (1) — syntax

## Definition 3.1 (alphabet)

*The alphabet $\mathcal{A} = \mathcal{V} \cup \{\neg, \wedge, \vee, \rightarrow, ), (\}$ consists of various symbols:*

- *a finite set $\mathcal{V}$ of symbols $X$, $Y$, ...; aka non-logical symbols, propositional variables*
- *junctors $\neg$, $\wedge$, $\vee$, $\rightarrow$; aka logical symbols, connectives, names of truth functions*
- *parenthesis ), (*

一个公式A是一个有限长度的符号串，沿着以下的归纳定义建立。

V中的所有符号都是公式

L如果A和B是公式，那么¬A, A , B, A - B, A B和（A）也是公式。

## Definition 3.2 (formula)

*A formula $A$ is a finite-length string of symbols build along the following inductive definition:*

- *all symbols in $\mathcal{V}$ are formulas*
- *if $A$ and $B$ are formulas then $\neg A$, $A \wedge B$, $A \vee B$, $A \rightarrow B$ and $(A)$ are also formulas*

F是所有公式的集合。F是所有字符串的一个子集，

*$\mathcal{F}$ is the set of all formulas. $\mathcal{F}$ is a subset of all strings, i.e. $\mathcal{F} \subset \mathcal{A}^*$.*

In the theory of formal grammars, $\mathcal{F}$ is a context-free language, i.e. definition 2 is a context-free grammar (type II of Chomsky's hierarchy).

# Propositional logic (2) — semantics

**Definition 3.3 (boolean assignment)**

*A boolean assignment $\omega$ assigns every propositional variable in $\mathcal{V}$ a truth value, i.e.*

$$\omega : \mathcal{V} \to \{0, 1\}$$

*where $0$ represents false and $1$ represents true.*

定义3.3（布尔赋值

一个布尔赋值ω给V中的每个命题变量分配一个真值，即

其中0代表假，1代表真。

# Propositional logic (2) — semantics

## Definition 3.3 (boolean assignment)

*A boolean assignment $\omega$ assigns every propositional variable in $\mathcal{V}$ a truth value, i.e.*

$$\omega : \mathcal{V} \to \{0, 1\}$$

*where 0 represents false and 1 represents true.*

## Definition 3.4 (entailment)

*A boolean assignment $\omega$ induces a truth function $q : \mathcal{F} \to \{0, 1\}$ that maps all formulas onto truth values as follows:*

- $q(X) := \omega(X)$ *for propositional variables* $X \in \mathcal{V}$
- $q(\neg A) := 1 - q(A)$ *and* $q(A \wedge B) := q(A)q(B)$ *and* $q((A)) = q(A)$
- $q(A \vee B) := q(\neg(\neg A \wedge \neg B))$ *and* $q(A \to B) := q(\neg A \vee B)$

*If $q(A) = 1$ we say that $\omega$ entails $A$ and write $\omega \models A$.*

如果q(A)=1，我们就说ω包含了A，并写成ω à A。

# From propositional logic to probabilities (1)

## Definition 3.5 (sample space)

*The set $\Omega$ of all boolean assignments is called sample space. Note, that for $n$ propositional variables it has $2^n$ elements.*

从命题逻辑到概率 (1)
定义3.5（样本空间
所有布尔赋值的集合Ω被称为样本空间。注意，对于n个命题变量，它有2n个元素。

# From propositional logic to probabilities (1)

### Definition 3.5 (sample space)

*The set $\Omega$ of all boolean assignments is called sample space. Note, that for n propositional variables it has $2^n$ elements.*

### Definition 3.6 (probability mass function)

*The probability mass function $f : \Omega \to [0, 1]$ assigns each boolean assignment a probability, such that*

- $0 \leq f(\omega) \leq 1$ *for all* $\omega \in \Omega$
- $\sum_{\omega \in \Omega} f(\omega) = 1$

定义3.6（概率质量函数
概率质量函数f Ω [0, 1]为每个布尔赋值
分配了一个概率，从而

# From propositional logic to probabilities (1)

## Definition 3.5 (sample space)

*The set $\Omega$ of all boolean assignments is called sample space. Note, that for n propositional variables it has $2^n$ elements.*

## Definition 3.6 (probability mass function)

*The probability mass function $f : \Omega \to [0, 1]$ assigns each boolean assignment a probability, such that*

- $0 \leq f(\omega) \leq 1$ *for all* $\omega \in \Omega$
- $\sum_{\omega \in \Omega} f(\omega) = 1$

一个事件E`Ω是样本空间Ω的一个子集。每个公式A自然会诱导出一个事件EA。

## Definition 3.7 (event)

*An event $E \subset \Omega$ is a subset of the sample space $\Omega$. Each formula $A$ naturally induces an event $E_A$:*

$$E_A := \{\omega \in \Omega \text{ such that } \omega \vDash A\} \subset \Omega$$

*Different formulas can induce the same event. Note that $\Omega = E_{A \vee \neg A}$.*

# From propositional logic to probabilities (2)

**Definition 3.8 (probability)**

*The probability $p(E)$ of some event $E$ is the probability mass of $E$, i.e.*

$$p(E) := \sum_{\omega \in E} f(\omega)$$

*The probability $p(A)$ of some formula $A$ is defined as the probability $p(E_A)$ of the induced event $E_A$, i.e.*

$$p(A) := p(E_A) = \sum_{\omega \in E_A} f(\omega) = \sum_{\omega \models A} f(\omega)$$

某个事件E的概率p（E）是E
的概率质量，即

某个公式A的概率p(A)被定义为诱发
事件EA的概率p(EA)，也就是说。

# From propositional logic to probabilities (2)

**Definition 3.8 (probability)**

*The probability $p(E)$ of some event $E$ is the probability mass of $E$, i.e.*

$$p(E) := \sum_{\omega \in E} f(\omega)$$

*The probability $p(A)$ of some formula $A$ is defined as the probability $p(E_A)$ of the induced event $E_A$, i.e.*

$$p(A) := p(E_A) = \sum_{\omega \in E_A} f(\omega) = \sum_{\omega \models A} f(\omega)$$

**Definition 3.9 (joint probability)**

*The joint probability $p(A, B)$ of several formulas $A$ and $B$ is the probability of their conjuction, i.e. $p(A, B) := p(A \wedge B)$, similarly for more than two. The joint probability of several events is the probability of their intersection (remember events are subsets of $\Omega$).*

定义3.9（联合概率

几个公式A和B的联合概率p(A, B)是它们结合的概率，即p(A, B)=p(A , B)，对于两个以上的公式也是如此。几个事件的联合概率是它们相交的概率（记住事件是Ω的子

# From propositional logic to probabilities (3)

**Theorem 3.10 (Kolmogorov's axioms and more)**

1. $0 \leq p(A) \leq 1$
2. $p(\Omega) = 1$
3. $p(A \vee B) = p(A) + p(B)$ *if* $E_A \cap E_B = \varnothing$
4. $p(A \vee B) = p(A) + p(B) - p(A, B)$
5. $p(A) = p(A, B) + p(A, \neg B)$
6. $p(A) + p(\neg A) = 1$
7. $p(A \vee B) = p(\neg(\neg A \wedge \neg B))$, $p(A \rightarrow B) = p(\neg A \vee B)$

*The first three are Kolmogorov's axioms which imply 4., 5., 6.*

Proof of 3:

$$p(A \vee B) = \sum_{\omega \in E_{A \vee B}} f(\omega) = \sum_{\omega \in E_A} f(\omega) + \sum_{\omega \in E_B} f(\omega) = p(A) + p(B)$$

The second equality holds because of $E_A \cap E_B = \varnothing$.

# From propositional logic to probabilities (4)

**Definition 3.11 (conditional probability)**

*For some formula or event $B$ with non-zero probability, i.e. $p(B) > 0$, the conditional probability $p(A|B)$ is the ratio of the joint probability $p(A, B)$ and $p(B)$, i.e.*

$$p(A|B) := \frac{p(A, B)}{p(B)}$$

# From propositional logic to probabilities (4)

**Definition 3.11 (conditional probability)**

*For some formula or event $B$ with non-zero probability, i.e. $p(B) > 0$, the conditional probability $p(A|B)$ is the ratio of the joint probability $p(A, B)$ and $p(B)$, i.e.*

$$p(A|B) := \frac{p(A, B)}{p(B)}$$

What about defining $p(A|B) = 1$ for $p(B) = 0$ (which looks like “ex falso quod libet”)?

对于某个概率不为零的公式或事件B，即p(B)>0，条件概率p(ASB)是联合概率p(A, B)和p(B)的比率，即
对于p(B)=0，定义p(ASB)=1（这看起来像 "ex falso quod libet"），怎么样？

# From propositional logic to probabilities (4)

**Definition 3.11 (conditional probability)**

*For some formula or event $B$ with non-zero probability, i.e. $p(B) > 0$, the conditional probability $p(A|B)$ is the ratio of the joint probability $p(A, B)$ and $p(B)$, i.e.*

$$p(A|B) := \frac{p(A, B)}{p(B)}$$

What about defining $p(A|B) = 1$ for $p(B) = 0$ (which looks like “ex falso quod libet”)? This is a bad idea, since it would imply $p(A|B) + p(\neg A|B) = 2 \neq 1$.

# From propositional logic to probabilities (4)

**Definition 3.11 (conditional probability)**

*For some formula or event $B$ with non-zero probability, i.e. $p(B) > 0$, the conditional probability $p(A|B)$ is the ratio of the joint probability $p(A, B)$ and $p(B)$, i.e.*

$$p(A|B) := \frac{p(A, B)}{p(B)}$$

What about defining $p(A|B) = 1$ for $p(B) = 0$ (which looks like "ex falso quod libet")? This is a bad idea, since it would imply $p(A|B) + p(\neg A|B) = 2 \neq 1$.

**Theorem 3.12 (Kolmogorov's axioms, Bayes' rule and more)**

- *For fixed $B$, the conditional probability $p(A|B)$ fulfills Kolmogorov's axioms.*
- *Bayes' theorem: $p(B|A) = p(A|B)p(B)/p(A)$*

*Note that Bayes' theorem is often falsely called "Bayes' rule". I also call it Bayes' rule ;).*

# From propositional logic to probabilities (5)

**Lemma 3.13 (implication vs. conditional probability)**

*Are $p(A \to B)$ and $p(B|A)$ the same thing? Assume $p(A) > 0$, otherwise $p(B|A)$ is not defined.*

1. $p(B|A) = \frac{p(A)-(1-p(A\to B))}{p(A)}$
2. $p(A) = 1$ *implies* $p(B|A) = p(A \to B)$
3. $p(A \to B) \geq p(B|A) \geq p(B, A)$
4. $p(A \to B) \geq 1 - p(A)$
5. $p(A \to B) = 1$ *if and only if* $p(B|A) = 1$

Proof:

$$p(B|A) = \frac{p(A,B)}{p(A)} = \frac{p(A) - p(A, \neg B)}{p(A)} = \frac{p(A) - (1 - p(\neg A \vee B))}{p(A)}$$
$$= \frac{p(A) - (1 - p(A \to B))}{p(A)}$$

# What exactly do we mean by $p(A)$ or $p(A, B)$, etc?

## Events as inputs

- events $E$ are subsets of $\Omega$ (the set of all events)
- $p(E)$ is the probability that event $E$ happens

事件E是Ω（所有事件的集合）的
子集
p（E）是事件E发生的概率

# What exactly do we mean by $p(A)$ or $p(A, B)$, etc?

Events as inputs事件作为输入

- events $E$ are subsets of $\Omega$ (the set of all events)
- $p(E)$ is the probability that event $E$ happens

Formulas as inputs公式作为输入

- a formula $A \vee B$ induces an event $E_{A \vee B}$
- $p(A \vee B)$ is the probability that formula $A \vee B$ is true (defined via the event sets)

公式A-B诱发了一个事件EA-B
p(A-B)是公式A-B为真的概率（通过事件集定义）。

# What exactly do we mean by $p(A)$ or $p(A, B)$, etc?

Events as inputs

- events $E$ are subsets of $\Omega$ (the set of all events)
- $p(E)$ is the probability that event $E$ happens

Formulas as inputs

- a formula $A \vee B$ induces an event $E_{A \vee B}$
- $p(A \vee B)$ is the probability that formula $A \vee B$ is true (defined via the event sets)

We can use anything that is either true or false as input for $p$.

我们可以用任何真或假的东西作为p的输入。

# Summary

## Probability notation

- $p$ is a function of events. Also a function of formulas, since they induce events.
- $p(\cdot)$, plug in anything that is either true or false.

p是事件的一个函数。也是公式的函数，因为它们会诱发事件。

p( )，插入任何为真或假的东西。

# Summary

## Probability notation

- $p$ is a function of events. Also a function of formulas, since they induce events.
- $p(\cdot)$, plug in anything that is either true or false.

## There are only two important rules:

$$
\begin{aligned}
p(A, B|C) &= p(A|B, C)\, p(B|C) && \text{product rule} \\
p(B|C) &= p(A, B|C) + p(\neg A, B|C) && \text{sum rule}
\end{aligned}
$$

# Summary

## Probability notation

- $p$ is a function of events. Also a function of formulas, since they induce events.
- $p(\cdot)$, plug in anything that is either true or false.

## There are only two important rules:

$$\begin{aligned} p(A, B|C) &= p(A|B, C)\ p(B|C) && \text{product rule} \\ p(B|C) &= p(A, B|C) + p(\neg A, B|C) && \text{sum rule} \end{aligned}$$

## ...with some variations:

$$\begin{aligned} p(A, B) &= p(A|B)\ p(B) && \text{product rule} \\ p(B) &= p(A, B) + p(\neg A, B) && \text{sum rule} \\ 1 &= p(A) + p(\neg A) \\ p(A, B) &= p(B|A)\ p(A) \\ p(B|A) &= p(A|B)\ p(B)/p(A) && \text{Bayes rule} \\ &\vdots \end{aligned}$$