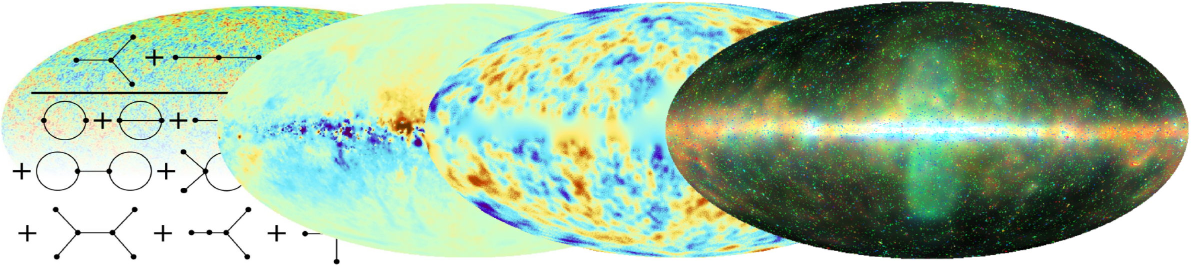

Information Theory

a short lecture

PD Dr. Torsten Enßlin
MPI for Astrophysics
LMU Munich

1. From Logic to Probability

Cox theorem (1946)

1.1 Aristotelian logic

A and B be statements or propositions	*e.g.* $A =$ "it rains", $B =$ "there is a cloud"
I background information	
$I =$ "if A is true, then B is also true" $=$ "$A \Rightarrow B$"	*e.g.* $I =$ "it rains only if there is a cloud"

deduction = syllogism

strong syllogism:

$I \Rightarrow$ "if B is false then A is false" $= (\overline{B} \Rightarrow \overline{A})$	*e.g.* $I \Rightarrow$ "no cloud $\rightarrow$ no rain"

weak syllogism:

$I \Rightarrow$ "if B (is true) then A is more plausible" $= J$	*e.g.* $I \Rightarrow$ "a cloud $\rightarrow$ maybe rain" $= J$

weaker syllogism:

$J \Rightarrow$ "if A, then B becomes more plausible"	*e.g.* $J \Rightarrow$ "rain $\rightarrow$ maybe a cloud"

1.2 Boolean Algebra

Boolean operations on statements A and B

- **“and”**: $AB =$ “both, A and B are true”
  conjunction or logical product
- **“or”**: $A + B =$ “at least one of the propositions A, B is true”
  disjunction or logical sum
- **“identity”**: “$A = B$” $=$ “A always has the same truth value as B”
  logical equivalence
- **“denial”**: $\overline{A} =$ “not A” $=$ “A is false”
  negation or logical complement
  $A =$ “$\overline{A}$ is false”, “$A{=}\overline{A}$” is always false

Notation:

- $AB + C = (AB) + C$
- $\overline{AB} = \overline{(AB)}$= “$AB$ is false”

Axioms of Boolean Algebra

idempotency:	$AA = A$
	$A + A = A$
commutativity:	$AB = BA$
	$A + B = B + A$
associativity:	$A(BC) = (AB)C = ABC$
	$A + (B + C) = (A + B) + C = A + B + C$
distributivity:	$A(B + C) = AB + AC$
	$A + (BC) = (A + B)(A + C)$ $(*)$
duality:	$\overline{AB} = \overline{A} + \overline{B}$
	$\overline{A + B} = \overline{A}\,\overline{B}$
implication:	"$A \Rightarrow B$" $\equiv$ "$A = AB$" $=$"A and AB have the same truth value"
	$(A \Rightarrow B) =$ ("it rains" is as true as "it rains and there is a cloud")

axiom set is over-complete, *e.g.* 2nd distrib. $(*)$ follows from other axioms:

$$\overline{A} + \overline{B}\,\overline{C} \overset{\text{duality}}{=} \overline{A} + \overline{B + C} \overset{\text{duality}}{=} \overline{A\,(B + C)} \underset{\text{distr.}}{\overset{1^{\text{st}}}{=}} \overline{AB + AC} \overset{\text{duality}}{=} \overline{AB}\,\overline{AC} \overset{\text{duality}}{=} (\overline{A} + \overline{B})\,(\overline{A} + \overline{C})$$

$$\overline{A} \to A', \overline{B} \to B', \overline{C} \to C' \Rightarrow (*) \quad \square$$

1.3 Plausible Reasoning

Aim: extend binary logic to continous plausible reasoning

Notation: $\pi(A|B)$= plausibility π of "A given B"

= "conditional plausibility that A is true, given that B is true"

1.3.1 Desiderata

I	Degrees of plausibility are represented by real numbers.
II	Qualitative correspondence with common sense.
1.	Aristotelian logic should be included.
III	Self consistency of the plausibility value system:
1.	If a conclusion can be reasoned in several ways, their results must agree.
2.	Equivalent knowledge states are represented by equivalent plausibilities.
3.	All available information must be included in any reasoning.

Convention: $C =$ "A is more plausible than B" $\Rightarrow \pi(A|C) > \pi(B|C)$, $\pi(\overline{A}|C) < \pi(\overline{B}|C)$

$D \xrightarrow{\text{update}} D'$ with $\pi(A|D') > \pi(A|D)$ and $\pi(B|AD') = \pi(B|AD)$

$\Rightarrow \pi(AB|D') \geq \pi(AB|D)$, $\pi(\overline{A}|D') < \pi(\overline{A}|D)$

1.3.2 The Product Rule

Decomposition of $AB|C$ = "A and B given C"

1. a) Decide whether B is true under C by specifying $\pi(B|C)$
   b) If this is the case, decide if A is also true by specifying $\pi(A|BC)$.
2. a) Decide whether A is true under C by specifying $\pi(A|C)$
   b) Given A, decide if B is also true by specifying $\pi(B|AC)$

Desideratum III.1$\Rightarrow$ $\exists$ plausibility function $f(x,y) = z$:

$$\pi(AB|C) = f(\pi(B|C), \pi(A|BC)) = f(\pi(A|C), \pi(B|AC)).$$

Convention/desideratum II: $f(x,y)$ continuous and monotonic in both x,y.
Decomposition of $ABC|D$

$$f(f(x,y),z) = f(x,f(y,z)).$$

Cox (1946): $\exists$ transformed plausibility system ω:

$$\omega(f(x,y)) = \omega(x)\,\omega(y) \text{ or } f(x,y) = \omega^{-1}(\omega(x)\,\omega(y)).$$

Product rule: $\omega(AB|C) = \omega(A|BC)\,\omega(B|C) = \omega(B|AC)\,\omega(A|C).$

1.3.3 True and False

True: assume "A certain given C" = "$C \Rightarrow A$"
$\Rightarrow$(i) $AB|C = B|C$, (ii) $A|BC = A|C$
$\omega(B|C) \overset{\text{(i)}}{=} \omega(AB|C) = \omega(A|BC)\,\omega(B|C) \overset{\text{(ii)}}{=} \omega(A|C)\,\omega(B|C)$ true for any B
$\Rightarrow \omega(A|C) = 1$

False: assume "A is impossible, given C" = "$C \Rightarrow \overline{A}$"
$\Rightarrow$(iii) $AB|C = A|C$, (iv) $A|BC = A|C$
$\omega(A|C) \overset{\text{(iii)}}{=} \omega(AB|C) = \omega(A|BC)\,\omega(B|C) \overset{\text{(iv)}}{=} \omega(A|C)\,\omega(B|C)$ true for any B

$$\Rightarrow \omega(A|C) = \begin{cases} 0 \\ \infty \end{cases}$$

[solution $\omega(A|C) = -\infty$ ruled out by the special case $A = B$]

plausibilities $\omega \in [0,1]$, implausibilities $\omega' \in [1,\infty]$, related by $\omega = \frac{1}{\omega'}$

Convention: plausibilities $\omega \in [0,1]$
$\omega(A|B) = 0$ expressing A is false given B
$\omega(A|B) = 1$ expressing A if true given B

1.3.4 Negation

Aristotelian logic:

- A is either true or false
- $A\overline{A}$ is always false
- $A+\overline{A}$ is always true

Negation function $S : [0,1] \to [0,1]$

$$\omega(\overline{A}|B) = S(\omega(A|B)),$$

S monotonically decreasing, $S(0) = 1$, and $S(1) = 0$, Cox (1946): consistency requires

$$\begin{aligned} S(x) &= (1-x^m)^{1/m} \qquad x \in [0,1],\ 0 < m < \infty. && (1) \\ \Rightarrow \omega(\overline{A}|B) &= S(\omega(A|B)) = (1-\omega^m(A|B))^{1/m} && (2) \\ \omega^m(\overline{A}|B) &= 1-\omega^m(A|B) && (3) \end{aligned}$$

- **sum rule:** $\omega^m(\overline{A}|B) + \omega^m(A|B) = 1$
- **product rule:** $\omega^m(AB|C) = \omega^m(A|BC)\,\omega^m(B|C) = \omega^m(B|AC)\,\omega^m(A|C),\ \ \omega^m \to P$

1.4 Probability

Convention: plausibility system with exponent $m = 1$ defines **probabilities**

$$P(x) = \omega^m(x)$$

1.4.1 Probability system

$P(A|B) =$"probability of A given B"

product rule:	$P(AB|C) = P(A|BC)\,P(B|C) = P(B|AC)\,P(A|C)$
sum rule:	$P(A|B) + P(\overline{A}|B) = 1$

Probabilities can be based on

- logic (extended to uncertainty)
- relative frequencies of events (frequentist definition)

$$P(\text{specific event} \mid \text{generic event}) = \lim_{n\to\infty} \frac{n(\text{specific event})}{n(\text{generic event})}$$

- set theory (Kolmogorov system)
- consistent bet ratios (de Finetti approach)

1.4.2 Marginalization

$P(A,B|C) \xrightarrow{\text{marginalization}} P(A|C)$, note new notation $A,B := AB$ for "and"

(i) options B and $\overline{B}$, exclusive ($B\overline{B}$ always false) and exhaustive ($B+\overline{B}$ allways true):

$$\begin{aligned} P(A,B|C) &= P(B|AC)\,P(A|C) \\ P(A,\overline{B}|C) &= P(\overline{B}|AC)\,P(A|C) \\ \Rightarrow\; P(A,B|C)+P(A,\overline{B}|C) &= \underbrace{\left[P(B|AC)+P(\overline{B}|AC)\right]}_{1} P(A|C) = P(A|C) \end{aligned}$$

$P(A|C) = P(A,B|C)+P(A,\overline{B}|C)$ is "B-marginalized probability of A"

(ii) options $\{B_i\}_{i=1}^{n}$ in I, mutually exclusive ($B_iB_j|I$ false for $i \neq j$), exhaustive ($\sum_i B_i|I$ true):

$$P(A|I) = \sum_{i=1}^{n} P(A,B_i|I) \text{ is } B\text{-marginalized probability of } A \text{ given } I \tag{4}$$

Notation: $P(A) = P(A|I)$, $P(A|B) = P(A|BI)$ if context I is obvious

Warning: if context is not obvious confusion is guaranteed.

End