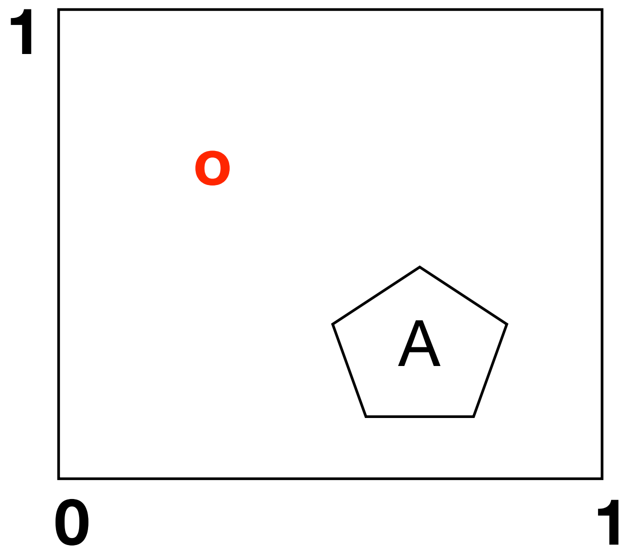
Lecture 2: probability concepts II.

Kolmogorov probability concept

example for sample space S of probabilistic outcomes of experiments:
x and y coordinates probed in (0,1) intervals of the two coordinates:



event A: outcome which does occur within polygon A

measurable probability space (S, σ) where σ is all the subsets

Kolmogorov probability concept

(Ω, F, P) probability space:

- sample space Ω (set of all possible outcomes)
- set of events F
- each event is a subset of Ω containing zero or more outcomes
- probability measure P: probability of some event A is P(A)

probability measure is a function on the collection of events that satisfies certain axioms

Axioms: (satisfied by frequentist definition of probabilities)

I. $P(A) \geq 0$ for an event A

II. $P(\Omega) = 1$ where Ω is the set of all possible outcomes

III. if $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$

disjoint

union of mutually exclusive

Example of a theorem:

Theorem: $P(\emptyset) = 0$

Proof: $A \cap \emptyset = \emptyset$, so

$P(A) = P(A \cup \emptyset) = P(A) + P(\emptyset)$, q.e.d.

Kolmogorov probability concept

Simple example: coin toss

Consider a single coin-toss, and assume that the coin will either land heads (H) or tails (T) (but not both). No assumption is made as to whether the coin is fair.

We may define:

$$\Omega = \{H, T\}$$
$$F = \{\varnothing, \{H\}, \{T\}, \{H, T\}\}$$

Kolmogorov's axioms imply that:

$$P(\varnothing) = 0$$

The probability of *neither* heads *nor* tails, is 0.

$$P(\{H, T\}) = 1$$

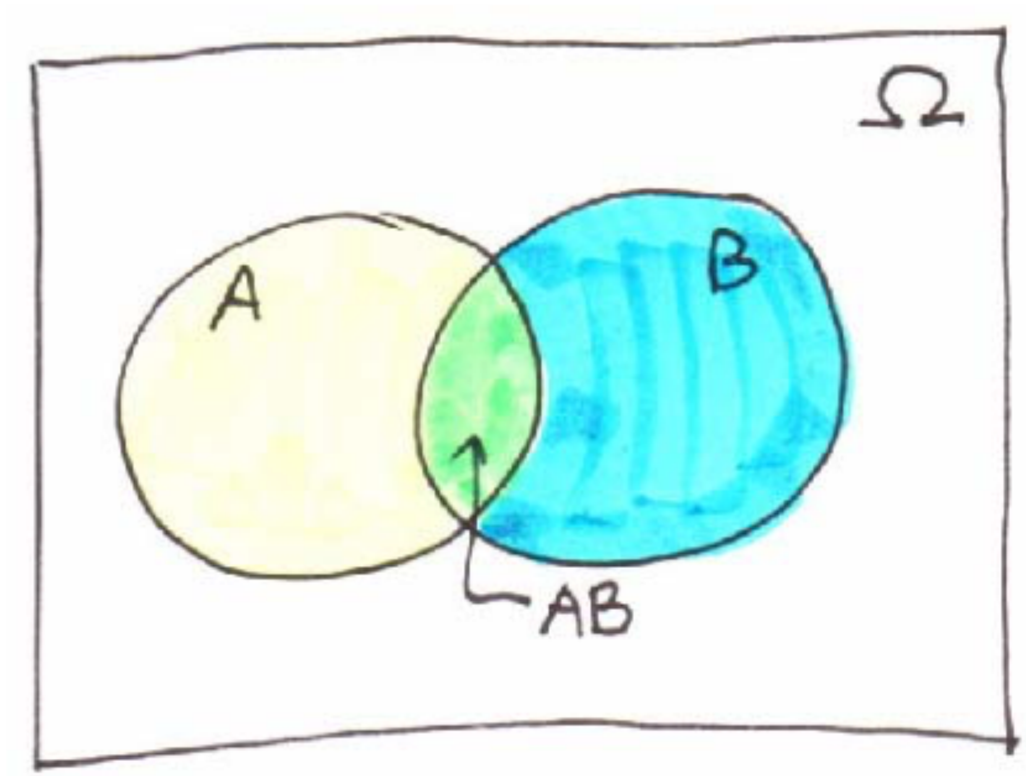
The probability of *either* heads *or* tails, is 1.

$$P(\{H\}) + P(\{T\}) = 1$$

The sum of the probability of heads and the probability of tails, is 1

Kolmogorov probability concept

Additivity or “Law of Or-ing”



Venn diagrams at web site of Probability, Mathematical Statistics, Stochastic Processes:

http://www.math.uah.edu/stat/

$$P(A \cup B) = P(A) + P(B) - P(AB)$$

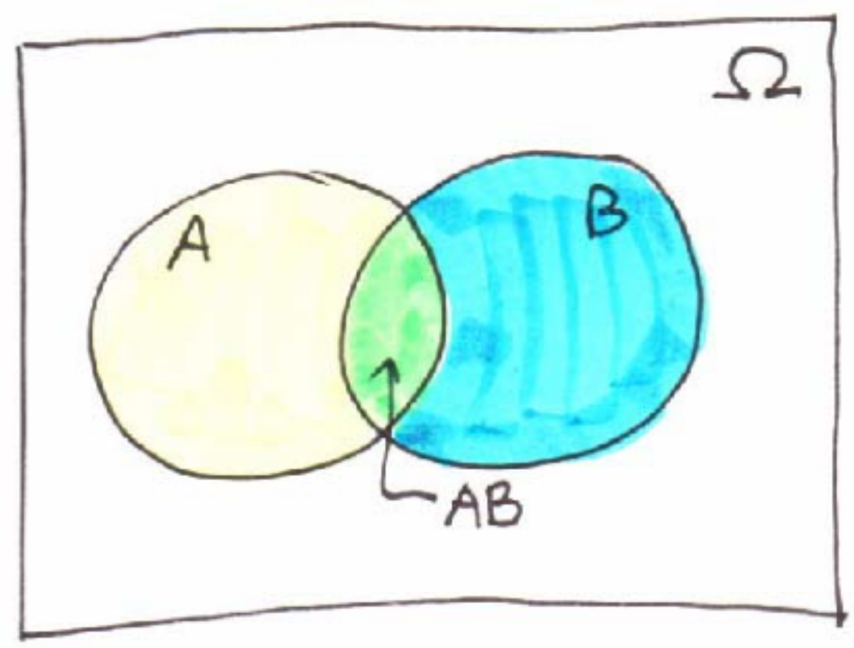
A or B

A and B

$$P(A \cap B)$$

Kolmogorov probability concept

Additivity or "Law of Or-ing"



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

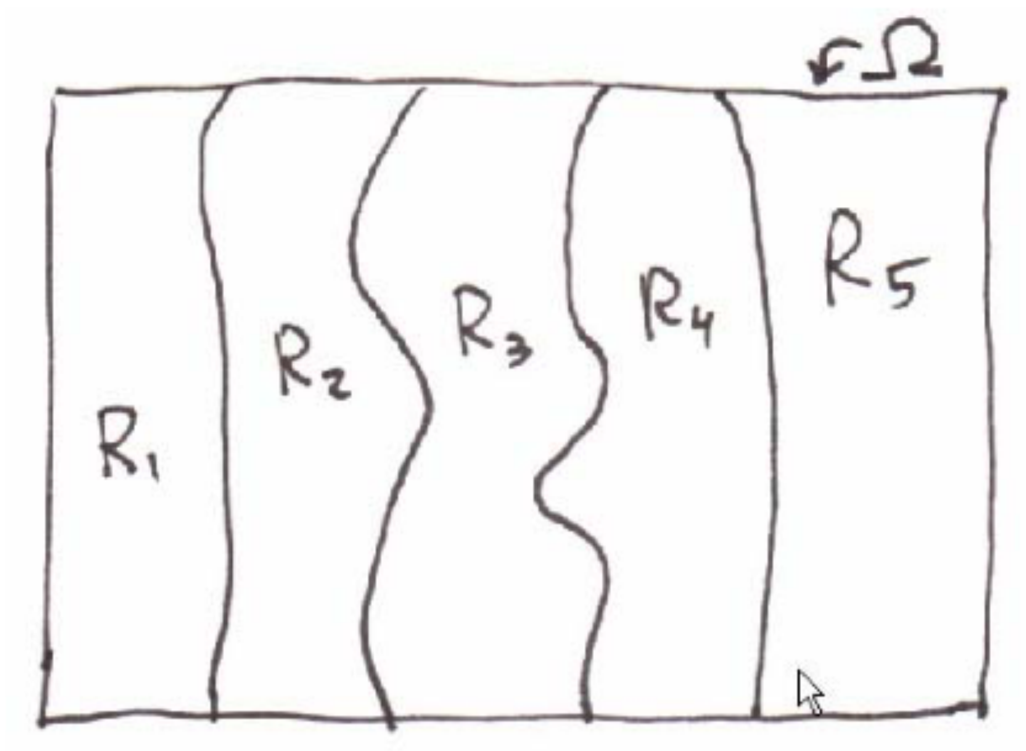
$$P(A \cup B) = P(A) + P(B \setminus (A \cap B)) \quad \text{(by Axiom 3)}$$

$$P(B) = P(B \setminus (A \cap B)) + P(A \cap B)\,.$$

Eliminating $P(B \setminus (A \cap B))$ from both equations gives us the desired result.

Kolmogorov probability concept

"Law of Exhaustion"



If R_i are exhaustive and mutually exclusive (EME)

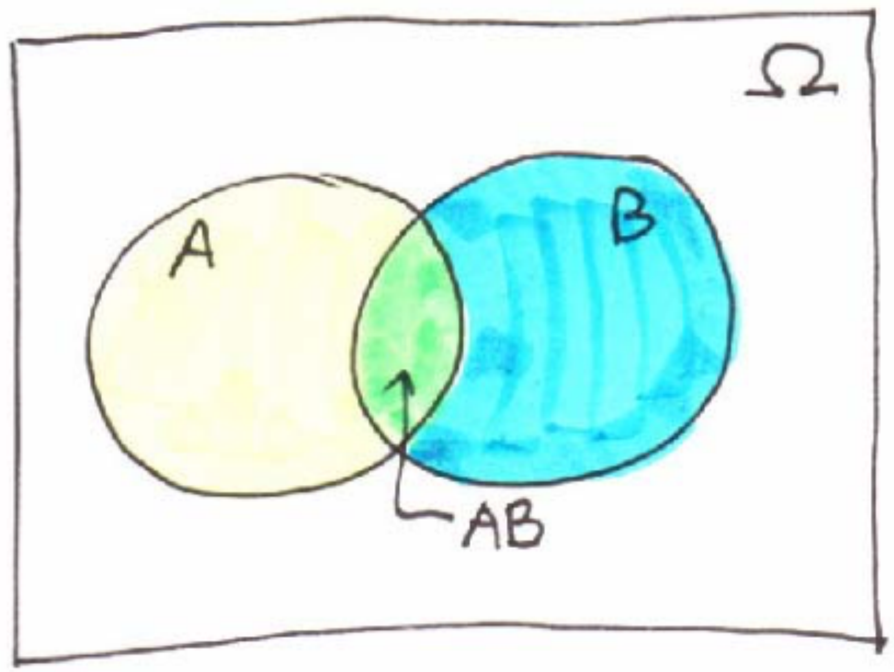
$$\sum_i P(R_i) = 1$$

This can be extended to the inclusion-exclusion principle

$$P\left(E^c\right) = P(\Omega \setminus E) = 1 - P(E)$$

Kolmogorov probability concept

Multiplicative Rule or "Law of And-ing"



(same picture as before)

"given"

$$P(AB) = P(A)P(B|A) = P(B)P(A|B)$$

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

"conditional probability"

"renormalize the outcome space"

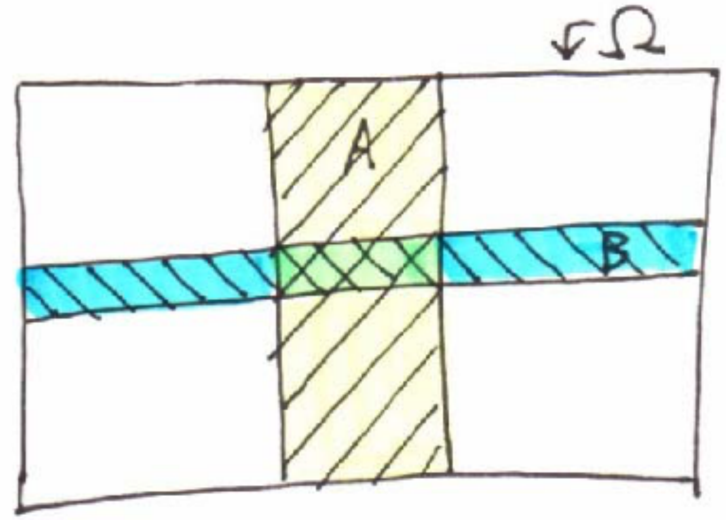
Kolmogorov probability concept

Similarly, for multiple And-ing:

$$P(ABC) = P(A)P(B|A)P(C|AB)$$

Independence:

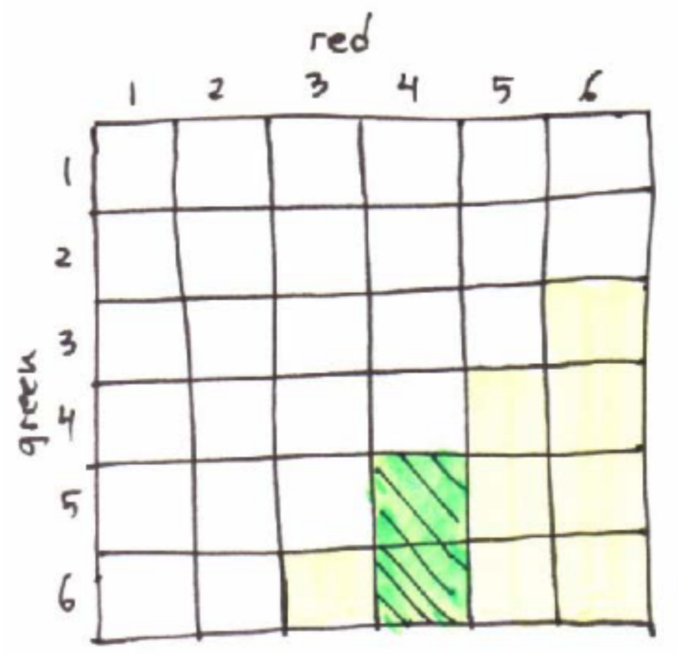
Events A and B are independent if
$P(A|B) = P(A)$
so $P(AB) = P(B)P(A|B) = P(A)P(B)$



Kolmogorov probability concept

A symmetric die has
$P(1) = P(2) = \ldots = P(6) = \frac{1}{6}$
Why? Because $\sum_i P(i) = 1$ and $P(i) = P(j)$.
Not because of "frequency of occurence in N trials".
That comes later!

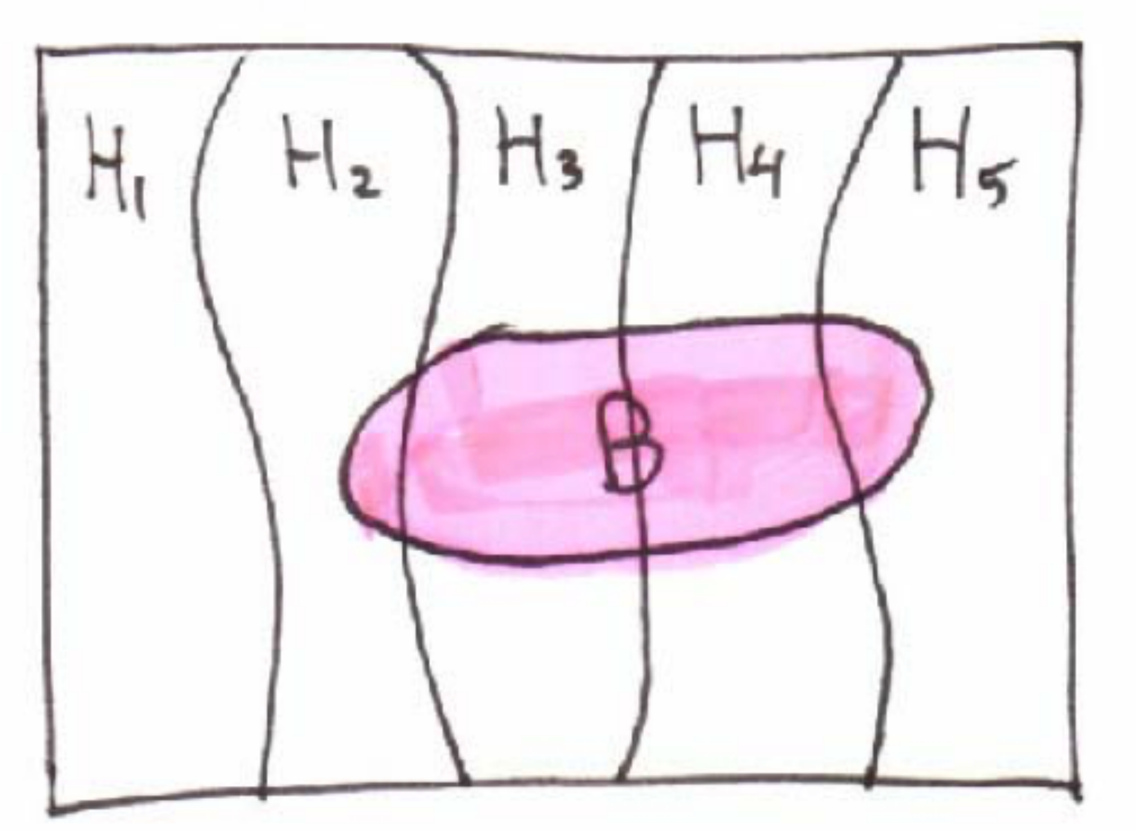
The sum of faces of two dice (red and green) is > 8.
What is the probability that the red face is 4?



$$P(R4\,|>8) = \frac{P(R4 \cap >8)}{P(>8)} = \frac{2/36}{10/36} = 0.2$$

Kolmogorov probability concept

Law of Total Probability or “Law of de-Anding”



H’s are exhaustive and mutually exclusive (EME)

$$P(B) = P(BH_1) + P(BH_2) + \ldots = \sum_i P(BH_i)$$

$$P(B) = \sum_i P(B|H_i)P(H_i)$$

Kolmogorov probability concept

1. $A \subseteq B$ if and only if the occurrence of A *implies* the occurrence of B.

2. $A \cup B$ is the event that occurs if and only if A occurs *or* B occurs.

3. $A \cap B$ is the event that occurs if and only if A occurs *and* B occurs.

4. A and B are disjoint if and only if they are *mutually exclusive*; they cannot both occur on the same run of the experiment.

5. $A \setminus B$ is the event that occurs if and only if A occurs *and* B does *not* occur.

6. A^c is the event that occurs if and only if A does *not* occur.

7. $(A \cap B^c) \cup (B \cap A^c)$ is the event that occurs if and only if *one but not both* of the given events occurs. Recall that this event is the *symmetric difference* of A and B, and is sometimes denoted $A \Delta B$.

8. $(A \cap B) \cup (A^c \cap B^c)$ is the event that occurs if and only if *both or neither* of the given events occurs.

Suppose now that $\mathscr{A} = \{A_i : i \in I\}$ is a collection of events for the random experiment, where I is a countable index set.

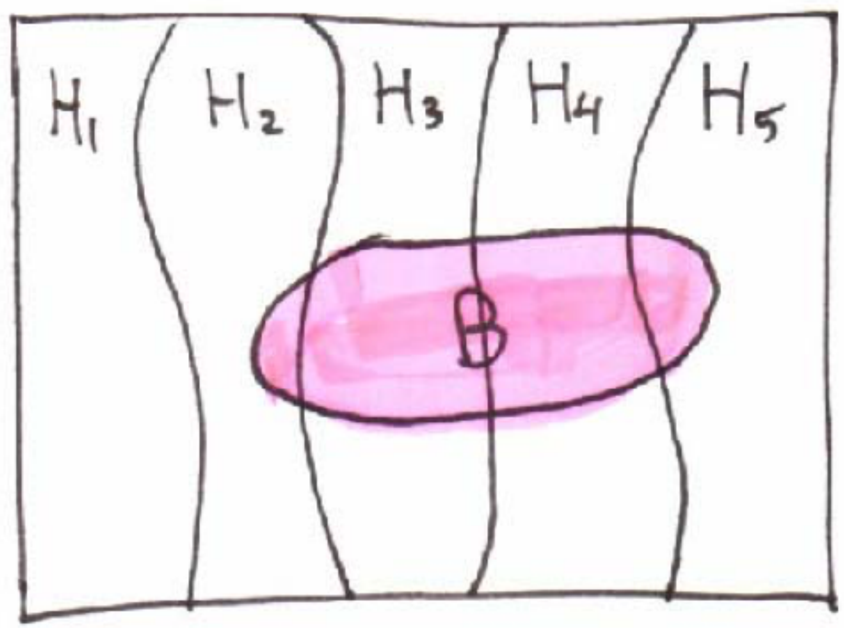
10. $\bigcup \mathscr{A} = \bigcup_{i \in I} A_i$ is the event that occurs if and only if *at least one* event in the collection occurs.

11. $\bigcap \mathscr{A} = \bigcap_{i \in I} A_i$ is the event that occurs if and only if *every* event in the collection occurs:

12. $\mathscr{A}$ is a pairwise disjoint collection if and only if the events are *mutually exclusive*; at most one of the events could occur on a given run of the experiment.

Bayes' theorem

Bayes Theorem



Thomas Bayes
1702 - 1761

(same picture as before)

Law of And-ing

$$P(H_i|B) = \frac{P(H_iB)}{P(B)}$$
$$= \frac{P(B|H_i)P(H_i)}{\sum_j P(B|H_j)P(H_j)}$$

Law of de-Anding

We usually write this as

$$P(H_i|B) \propto P(B|H_i)P(H_i)$$

this means, "compute the normalization by using the completeness of the H_i's"

Bayes' theorem

Let's work a couple of examples using Bayes Law:

Example: Trolls Under the Bridge

Trolls are bad. Gnomes are benign.
Every bridge has 5 creatures under it:

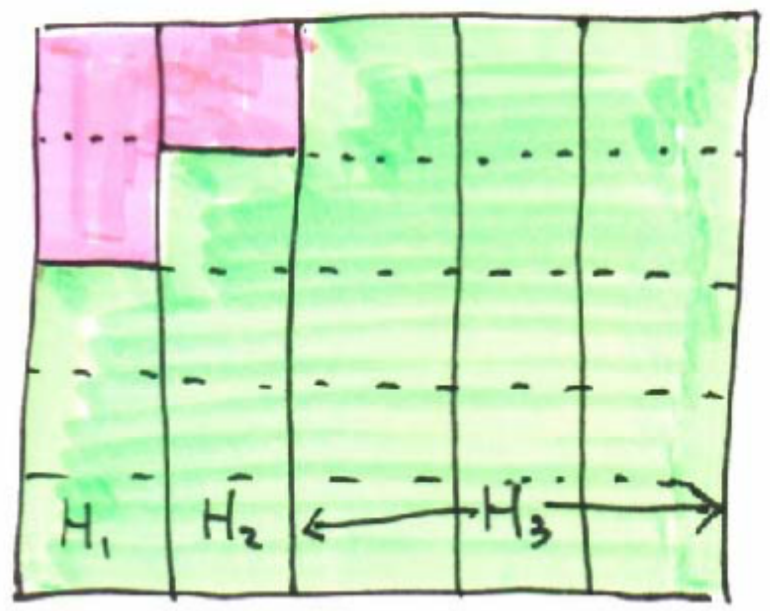
20% have TTGGG (H_1)
20% have TGGGG (H_2)
60% have GGGGG (benign) (H_3)

Before crossing a bridge, a knight captures one of the 5 creatures at random. It is a troll. "I now have an 80% chance of crossing safely," he reasons, "since only the case

20% had TTGGG (H1) → now have TGGG

is still a threat."

Bayes' theorem



$$P(H_i|T) \propto P(T|H_i)P(H_i)$$

so, $$P(H_1|T) = \frac{\frac{2}{5}\cdot\frac{1}{5}}{\frac{2}{5}\cdot\frac{1}{5}+\frac{1}{5}\cdot\frac{1}{5}+0\cdot\frac{3}{5}} = \frac{2}{3}$$

The knight's chance of crossing safely is actually only 33.3%
Before he captured a troll ("saw the data") it was 60%.
Capturing a troll actually made things worse!
(80% was never the right answer!)

Data changes probabilities!
Probabilities after assimilating data are called <u>posterior probabilities</u>.

Bayes' theorem

Bayes Law is a "calculus of inference", better (and certainly more self-consistent) than folk wisdom.

Example: Hempel's Paradox

Folk wisdom: A case of a hypothesis adds support to that hypothesis.

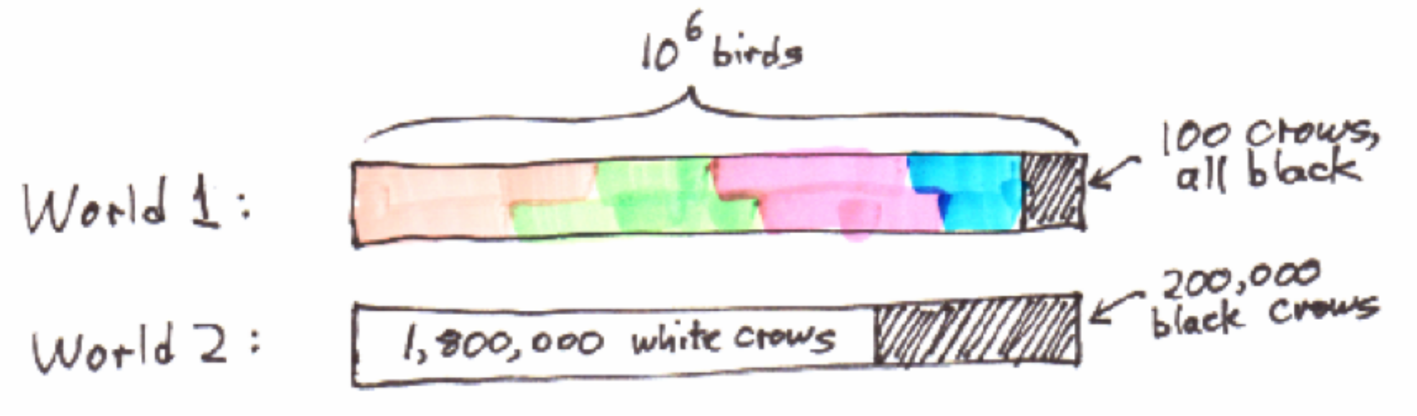
Example: "All crows are black" is supported by each new observation of a black crow.

All crows are black ⇔ All non-black things are non-crows

But this is supported by the observation of a white shoe.

So, the observation of a white shoe is thus evidence that all crows are black!

Bayes' theorem



I.J. Good: "The White Shoe is a Red Herring" (1966)

We observe one bird, and it is a black crow.
a) Which world are we in?
b) Are all crows black?

Important concept, Bayes odds ratio:

$$\frac{P(H_1|D)}{P(H_2|D)} = \frac{P(D|H_1)P(H_1)}{P(D|H_2)P(H_2)}$$

$$= \frac{0.0001\,P(H_1)}{0.1\,P(H_2)} = 0.001\,\frac{P(H_1)}{P(H_2)}$$

So the observation strongly supports H2 and the existence of white crows.

Hempel's folk wisdom premise is not true.

Data supports the hypotheses in which it is more likely compared with other hypotheses. (This is Bayes!)

We must have some kind of background information about the universe of hypotheses, otherwise data has no meaning at all.