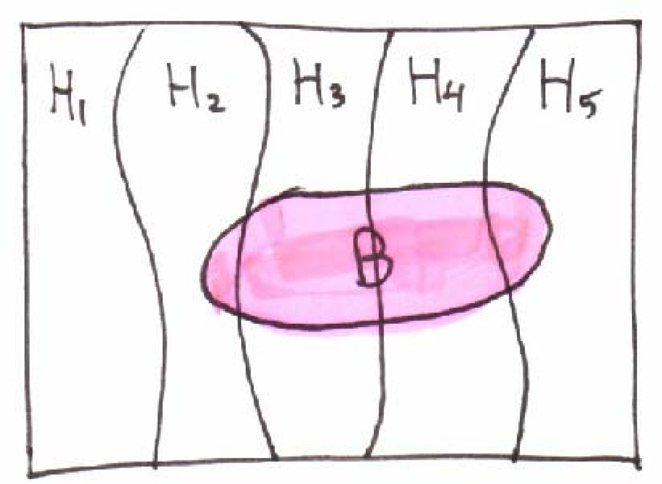
Opinionated Lessons in Statistics

by Bill Press

#2 Bayes

Bayes Theorem



Thomas Bayes
1702 - 1761

(same picture as before)

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

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

Law of And-ing

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"

- As a theorem relating probabilities, Bayes is unassailable
- But we will also use it in inference, where the H's are hypotheses, while B is the data
  - "what is the probability of an hypothesis, given the data?"
  - some (defined as frequentists) consider this dodgy
  - others (Bayesians like us) consider this fantastically powerful and useful
  - in real life, the "war" between Bayesians and frequentists is long since over, and most statisticians adopt a mixture of techniques appropriate to the problem
    - for a view of the "war", see Efron paper on the course web site
- Note that you generally have to know a complete set of EME hypotheses to use Bayes for inference
  - perhaps its principal weakness

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

Example: Trolls Under the Bridge

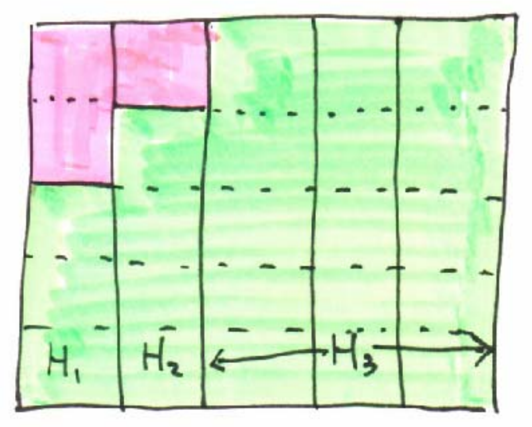
Trolls are bad. Gnomes are benign.
Every bridge has 5 creatures under it:

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.”



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

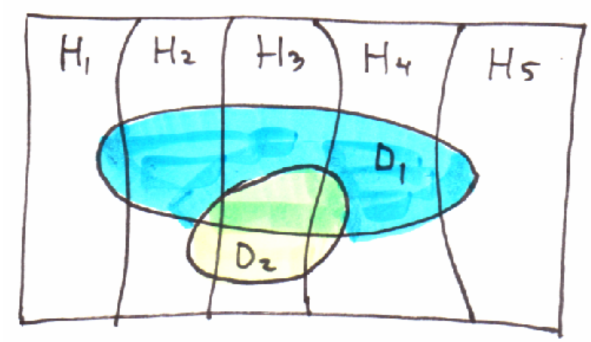
so, $$P(H_1|T) = \frac{\frac{2}{5}\cdot\frac{1}{5}}{\frac{2}{5}\cdot\frac{1}{5}+\frac{1}{5}\cdot\frac{1}{5}+0\cdot\frac{3}{5}} = \frac{2}{3}$$

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 posterior probabilities.

Commutivity/Associativity of Evidence

$P(H_i|D_1D_2)$ desired



We see D_1:

$P(H_i|D_1) \propto P(D_1|H_i)P(H_i)$

Then, we see D_2:

$P(H_i|D_1D_2) \propto P(D_2|H_iD_1)P(H_i|D_1)$ ← this is now a prior!

But,

$= P(D_2|H_iD_1)P(D_1|H_i)P(H_i)$

$= P(D_1D_2|H_i)P(H_i)$

this being symmetrical shows that we would get the same answer regardless of the order of seeing the data

All priors $P(H_i)$ are actually $P(H_i|D)$, conditioned on previously seen data! Often write this as $P(H_i|I)$. ← background information

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

Example: Hempel’s Paradox

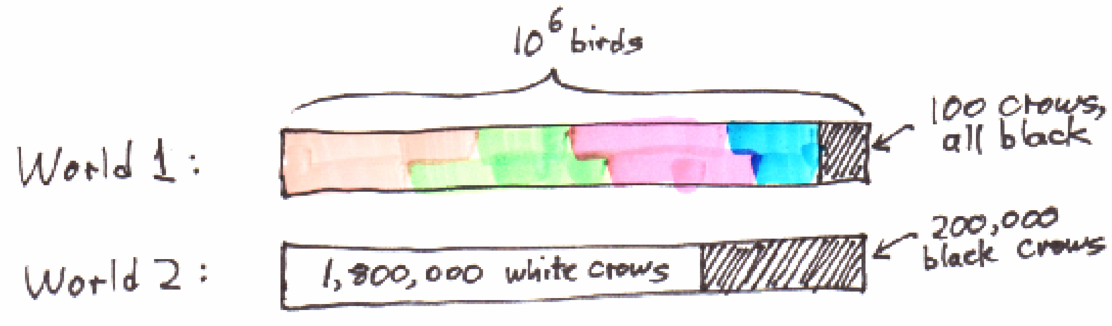
Folk wisdom: A case of a hypothesis adds support to that hypothesis.

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

All crows are black $\Leftrightarrow$ 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!



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 <u>some</u> kind of background information about the universe of hypotheses, otherwise data has no meaning at all.

Congratulations! You are now a Bayesian.

Bayesian viewpoint:
Probabilities are modified by data. This makes them intrinsically subjective, because different observers have access to different amounts of data (including their "background information" or "background knowledge").

Notice in particular that the connection of probability to "frequency of occurrence of repeated events" is now complicated! (Would have to "repeat" the exact state of knowledge of the observer.)