On thinking probabilistically

Michael E. McIntyre

Centre for Atmospheric Science
University of Cambridge, UK

www.atm.damtp.cam.ac.uk/people/mem/index.html#thinking-probabilistically

The Master said:
To know what you know, and to know what you do not know: that is knowledge.
– Analects of Confucius

Disclaimer:

This talk **won't** provide operational answers to all our problems of statistical testing and statistical inference.

It will try, however, to indicate recent conceptual advances that help us to **know what we're doing** when using probability theory.

There's an over-arching conceptual framework covering all the traditional methods **including** Bayesian methods, and clarifying the old subjectivity-objectivity dilemmas.

Main take-home message: **probabilities are always conditional** (on the **information** you have and the **assumptions** you make)

– so the fundamental ideal of ultra-orthodox ('hardcore frequentist') statistical theory – that probabilities are absolutely objective, i.e. properties of things in the outside world – is, and always was, a delusion.

Indeed, it's a delusion highly injurious to good science.

Why?

- For a coherent account of what science is, we need to distinguish between data and models.
- To build good models, we need to be able to use probabilities as model properties.
- We also, of course, need to make assumptions.
- Same for a coherent account of what ordinary perception is – *despite* its subjective feeling of directness, e.g. of directly 'seeing what's there'.
- (Cf. the way music works, acausality illusions etc. I tried to explain all this in the "Lucidity and Science" papers of 1997 – websearch "lucidity principles".)
- What are models? They are partial and approximate representations of reality.
- What are data? Data consist of information coming directly from the outside world. One example is patterns of photons hitting the retinas of your eyes:

Walking-lights animation from
http://www.atm.damtp.cam.ac.uk/people/mem/

Kobe Lecture

So ‘seeing what’s there’ is **not** wholly objective!!
Unconscious **assumptions** are always involved!

‘‘No organism can afford to be conscious of matters with which it could deal at unconscious levels.’’
– Gregory Bateson (1972)

That of course applies to ourselves as well as to other species. So if we **think** we’re making no assumptions, it means only that **all** our assumptions are unconscious.

The hardcore frequentist statisticians of the early 20th century seem to have fallen into this trap. They thought they could attain absolute objectivity: “Let the data speak for themselves.”

- Brain is fitting a particular model (piecewise-3D skeletal motion) to data consisting, essentially, of 12 moving points in a 2D plane.

- The model-fitting process is wholly unconscious, and involves unconscious assumptions including something like Bayesian prior "probabilities". Details in ww.atm.damtp.cam.ac.uk/people/mem/index.html#thinking-probabilistically

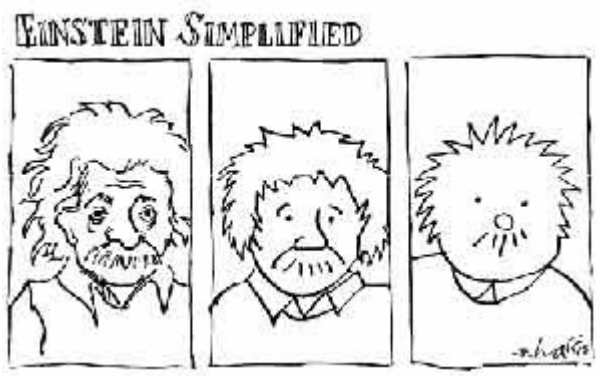
- Illustrates "observations are theory-laden"; there's always *some* subjectivity.

- **YET** there are such things as **goodness-of-fit**, **model simplicity**, **large domain of applicability** etc… Some models *are* much better than others. But none can be "absolutely true".

- **All this applies to science, because science fits models to data and is thus an extension of ordinary perception.**

- (Science wars: "science as mere opinion" vs "science as absolute truth" is a false dichotomy!)

- Actually use/need **hierarchy of models**
- (consciously or unconsciously)
  (Natural selection has equipped us therewith)



What are these things called "models"?

They are partial and approximate representations of reality.

Note:

- Children's model boats, cars, houses, aircraft... or Crick & Watson's DNA model are models in this sense (as well as real objects) (can be regarded as)
- Same objects in virtual reality are likewise models in same sense.
- So models can be made of computer code.

  NB: Inspecting the computer code would not tell you that it represented a model boat, car,... The same model can have different representations.
- Models can be made of neural patterns and protein circuitry.
- Models can be made of mathematical equations.

  (Cf. Eugene Wigner's "unreasonable effectiveness" of mathematics.)

(further discussion in my ***Kobe Lecture*** (websearch will find it)

The models we use to make sense of sounds – to do 'auditory scene analysis' – use the harmonic series as a basic building block.

From "Lucidity and Science" Part 1:

Reminder of what 'combinatorially large' means numerically; note that the number of ways to make a structure as simple as a linear, one-dimensional chain with 10 different links is 3,628,800, and with 100 different links, of the order of 10^{158}, i.e. of the order of

100,000,000,000,000,000,000,000,000,000,000,000,
000,000,000,000,000,000,000,000,000,000,000,000,
000,000,000,000,000,000,000,000,000,000,000,000,
000,000,000,000,000,000,000,000,000,000,000,000,
000,000,000,000,000,000,000,000.

Perceptual processing must deal with still larger numbers of possibilities, corresponding to internal models with far more complicated structures.....

How does the brain cope with the problem of combinatorial largeness – of selecting from the combinatorially large number of possible models to fit to the data?

One way is by perceptual grouping: – another wholly unconscious process first studied by the Gestalt psychologists:

SOME GREAT NAMES IN GAMBLING

Jakob Bernoulli
(1654-1705)

Pierre Simone de Laplace
(1749-1827)

Bruno de Finetti
(1906-1985)

“ARBITRAGE”

“DUTCH BOOKS”

BUT how many of you have ever heard of

Richard Threlkeld Cox?

Cox's theorem(s) of 1946 illuminate the foundations of probability theory, emphasizing the status of probability as a **model property** (and, incidentally, *why* natural selection equipped us with unconscious probability theory, along with other kinds of unconscious mathematics).

This brings with it a new clarity, power, and **flexibility** in the uses of probability theory – via an overarching framework that **includes** all the traditional frequentist thought-experiments (dice-throwing, sampling from large populations, etc.), and **includes** so-called Bayesian methods too.

Frequentist thought-experiments are, of course, useful in their place. It's the hardcore frequentist ideology, conflating data with models, that's injurious to good science.

The conceptual framework built on Cox's theorems is available in two landmark books:

Jaynes, E. T., *Probability Theory: The Logic of Science*, 727 pp. *…beware* **maxent**

MacKay, D. J. C. *Information Theory, Inference, and Learning Algorithms*, 628 pp. ..*beware* **Cox's Theorems!**

Both published by Cambridge University Press in 2003.

Online versions are available, partial or complete – details at www.atm.damtp.cam.ac.uk/people/mem/index.html#thinking-probabilistically

Cox's theorems tell us that:

- All of probability theory follows from a single qualitative **primordial idea**,namely that
- for given **background knowledge or information** Z, our brains can assess the plausibility of **any proposition** A; call this plausibility $P(A|Z)$.
- Then, under very weak qualitative assumptions, we can **prove** that the quantities $P(A|Z)$ are **mathematically indistinguishable from probabilities**.
- That is, aside from trivial transformations, they obey the product rule $P(AB|Z) = P(A|BZ)\,P(B|Z) = P(B|AZ)\,P(A|Z)$ and the sum rule $P(A|Z) + P(\text{not } A|Z) = 1$.

Cox's theorems tell us that:

- All of probability theory follows from a single qualitative **primordial idea**,namely that
- for given **background knowledge or information** Z, our brains can assess the plausibility of **any proposition** A; call this plausibility $P(A|Z)$.
- Then, under very weak qualitative assumptions, we can **prove** that the quantities $P(A|Z)$ are **mathematically indistinguishable from probabilities**.
- That is, aside from trivial transformations, they obey the product rule $P(AB|Z) = P(A|BZ)\,P(B|Z) = P(B|AZ)\,P(A|Z)$ and the sum rule $P(A|Z) + P(\text{not } A|Z) = 1$.
- Note that **Bayes' Theorem** or **Bayes' Rule** is just the second equality in the product rule.

Cox's theorems tell us that:

- All of probability theory follows from a single qualitative **primordial idea**,namely that
- for given **background knowledge or information** Z, our brains can assess the plausibility of **any proposition** A; call this plausibility $P(A|Z)$.
- Then, under very weak qualitative assumptions, we can **prove** that the quantities $P(A|Z)$ are **mathematically indistinguishable from probabilities**.
- That is, aside from trivial transformations, they obey the product rule $P(AB|Z) = P(A|BZ)\,P(B|Z) = P(B|AZ)\,P(A|Z)$ and the sum rule $P(A|Z) + P(\text{not } A|Z) = 1$.
- Note that **Bayes' Theorem** or **Bayes' Rule** is just the second equality in the product rule.
- **Conditioning statements are primordial**.
- So: **priors are always involved**, whether or not we are using Bayes' Theorem.

Making the $|Z$ explicit blows away all kinds of difficulties -- take for instance the notorious Monty Hall "3 doors" or "3 cards" problem:

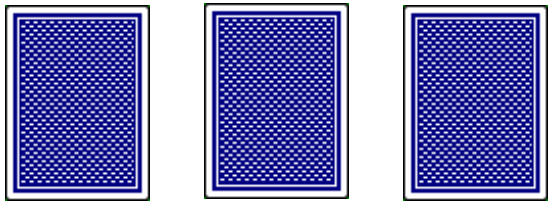
Monty Hall, host of *Let's Make a Deal*

Marilyn vos Savant,
who
received nearly
10,000 wrong answers.

3-cards version: Player 1 ('Monty') puts three cards face down. One is an ace, the others, ordinary cards.

Z includes the rules of the game (MacKay Ex 3.8ff), most crucially that: Player 1 ('Monty') knows where ace is, and will always flip ANOTHER card. Label positions as

1 2 3

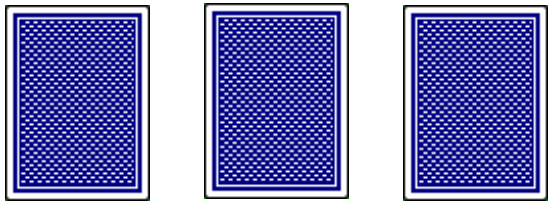


As Player 2, I assume Z *also* includes **priors**:

$$P(A_1|Z) = P(A_2|Z) = P(A_3|Z) = 1/3$$ (**subjective!!**)

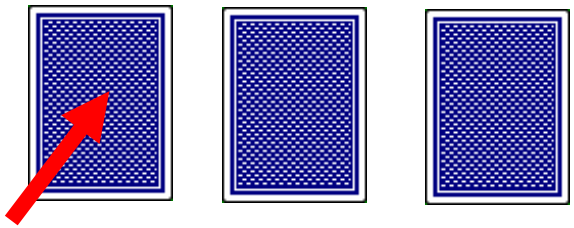
Z includes the rules of the game (MacKay Ex 3.8ff), most crucially that: Player 1 ('Monty') knows where ace is, and will always flip ANOTHER card. Label positions as

1 2 3



As Player 2, I assume Z *also* includes **priors**:

$$P(A_1|Z) = P(A_2|Z) = P(A_3|Z) = 1/3$$ (**subjective!!**)

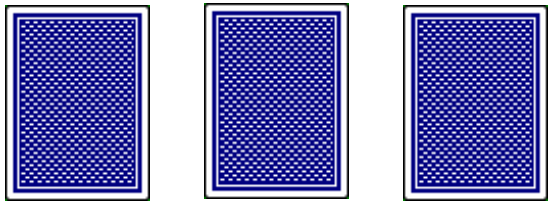


I finger card 1: Z updated to Z' but I still have

$$P(A_1|Z') = P(A_2|Z') = P(A_3|Z') = 1/3.$$

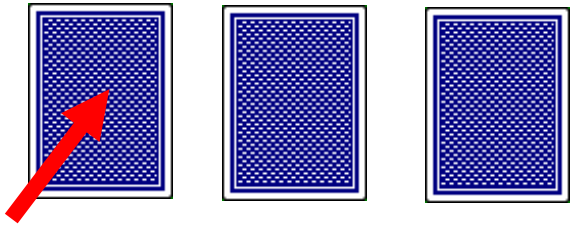
Z includes the rules of the game (MacKay Ex 3.8ff), most crucially that: Player 1 ('Monty') knows where ace is, and will always flip ANOTHER card. Label positions as

1 2 3



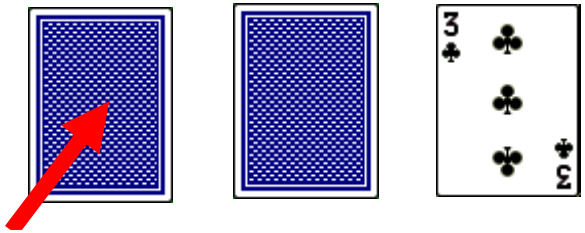
As Player 2, I assume Z *also* includes **priors**:

$$P(A_1|Z) = P(A_2|Z) = P(A_3|Z) = 1/3 \quad \textbf{(subjective!!)}$$



I finger card 1: Z updated to Z' but I still have

$$P(A_1|Z') = P(A_2|Z') = P(A_3|Z') = 1/3.$$

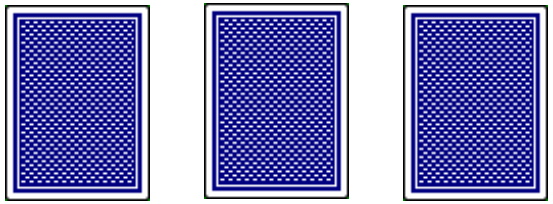


Z' updated to Z'', but (if Player 1 unbiased)

$$P(A_1|Z'') = 1/3, \; P(A_2|Z'') = \mathbf{2/3}, \; P(A_3|Z'') = \mathbf{0}$$

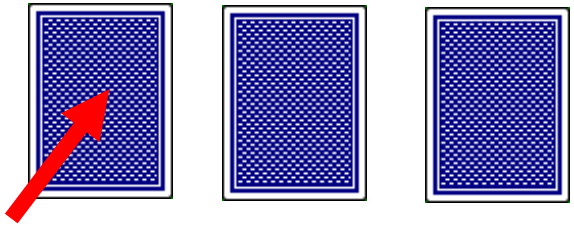
Z includes the rules of the game (MacKay Ex 3.8ff), most crucially that: Player 1 ('Monty') knows where ace is, and will always flip ANOTHER card.
Label positions as

1 2 3



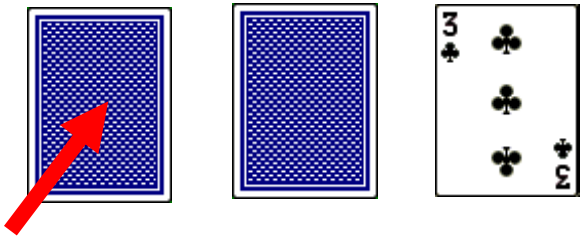
As Player 2, I assume Z *also* includes **priors**:

$$P(A_1|Z) = P(A_2|Z) = P(A_3|Z) = 1/3 \quad \textbf{(subjective!!)}$$



I finger card 1: Z updated to Z' but I still have

$$P(A_1|Z') = P(A_2|Z') = P(A_3|Z') = 1/3.$$



Z' updated to Z'', but (if Player 1 unbiased)

$$P(A_1|Z'') = 1/3, \; P(A_2|Z'') = \mathbf{2/3}, \; P(A_3|Z'') = \mathbf{0}$$

So Marilyn was right. **On these assumptions,** it's better to switch.
And NB again: **priors are involved**, whether or not we use Bayes' Theorem (Bayes' Rule).

Bayes' Rule and its **chain-consistency property**

- Bayes' Rule is just part of the product rule, usually written

$$P(M\,|\,DZ) = \frac{P(D\,|\,MZ)\;P(M\,|\,Z)}{P(D\,|\,Z)}$$

- The standard use is to let some set of candidate models or hypotheses M to define the pdf $P(D|MZ)$, and use $P(M|Z)$ to express one's prior judgement of the plausibility of model M. Remember, there's no escape from subjectivity; so it's a good thing to be forced to make your prejudices explicit.
- If new data are acquired, we can set D to be the statement that those data take whatever values they do. (more loosely, but succinctly, D = 'the data just acquired')
- This keeps the data-model distinction clear, and forces us to make our assumptions explicit.
- Chain-consistency: can refine prior by repeating the above

Conclusions

- Frequentist thought-experiments can be useful (especially with computer-aided Monte Carlo). But they're only a tiny subset of what's relevant to science.
- The **hardcore frequentist** dogma (that P values are absolute properties **of real things in the outside world**) conflates **reality** with **models of reality**. That's injurious to science.
- Hardcore frequentism also makes taboo a consideration of background information, and the associated **prior probabilities**. That's **catastrophic to science**.
- Intuition can be very treacherous – e.g. the way my fingers got burned in the 3-cards (Monty Hall) problem!
- But, in that quicksand of conscious vs unconscious probabilistic thinking, **Cox's theorems give us a rock to stand on** and free us to think flexibly and creatively.
- I think **Cox's theorems should be taught to all science undergraduates** and have the same status for statistical inference as thermodynamics for physics.