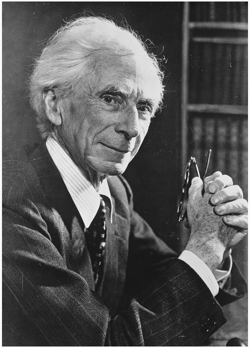
Bayesian Probability

Michael Habeck

Microscopic Image Analysis, Jena University Hospital
MPI for Multidisciplinary Sciences, Göttingen

Facets of Probability – April 1st, 2026



Probability is the most important concept in modern science, especially as nobody has the slightest notion what it means.

Bertrand Russell (1929)

$$\Pr(\,A \mid X\,)$$

How plausible is hypothesis A given knowledge X ?

False — True

0 ⟵ degree of implication/deducibility ⟶ 1

Bayesian probability is ...

... a quantitative measure of **uncertainty**, or degree of belief

... **epistemic** rather than ontic: it reflects our state of knowledge, not a physical property of the world

... always **conditional**:

$$\Pr(A \mid X)$$

where $A \equiv$ any proposition, $X \equiv$ given premises

Subjective, but not arbitrary

Bayesian probability is often criticized as being **”subjective”**

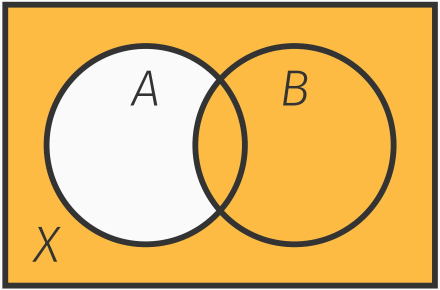
But it is more accurate to see probability as

> ”***subjectively objective***: *Probability exists in your mind: if you’re ignorant of a phenomenon, that’s an attribute of you, not an attribute of the phenomenon.”* (Eliezer Yudokowsky)

Probability obeys **objective rules**:

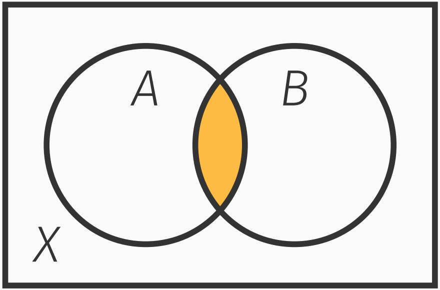
> *”you can’t change probabilities by wishing”*

Implication: *if A, then B*



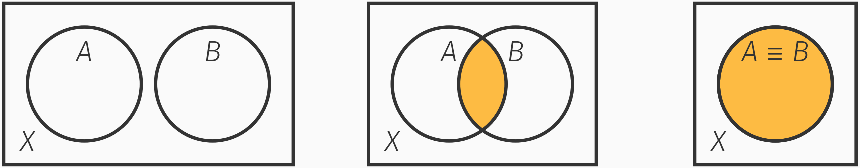
A	B	$A \to B$
1	1	1
1	0	0
0	1	1
0	0	1

Modus ponens: $A \to B, A \vdash B$



A	B	$A \to B$
1	1	1
1	0	0
0	1	1
0	0	1

Probability as degree of implication



$\Pr(B \mid A \wedge X) = 0$ $\qquad$ $0 < \Pr(B \mid A \wedge X) < 1$ $\qquad$ $\Pr(B \mid A \wedge X) = 1$

$$\Pr(B \mid A \wedge X) = \frac{|A \cap B|}{|A|}$$

Product rule

$$\underbrace{\frac{|A \cap B|}{|A|}}_{\Pr(B|A \wedge X)} \; \underbrace{\frac{|A|}{|X|}}_{\Pr(A|X)} = \frac{|A \cap B|}{|X|} = \underbrace{\frac{|A \cap B|}{|B|}}_{\Pr(A|B \wedge X)} \; \underbrace{\frac{|B|}{|X|}}_{\Pr(B|X)}$$

The algebra of probable inference

Product rule

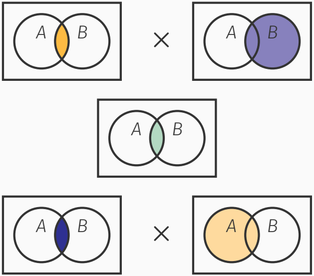
$$\Pr(A \mid B \wedge X) \times \Pr(B \mid X)$$

$$=$$

$$\Pr(A \wedge B \mid X)$$

$$=$$

$$\Pr(B \mid A \wedge X) \times \Pr(A \mid X)$$



Sum rule

$$\Pr(A \mid X) + \Pr(\neg A \mid X) = 1$$

The algebra of probable inference

Cox's theorem

Degrees of implication satisfy the sum rule and the product rule

Cox's axioms

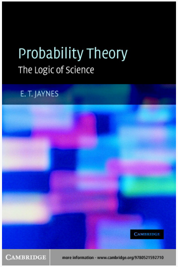
- $\Pr(\neg A \mid X) = F[\Pr(A \mid X)]$
- $\Pr(A \wedge B \mid X) = G[\Pr(A \mid X), \Pr(B \mid A \wedge X)]$

Richard T. Cox: The Algebra of Probable Inference (1961)

Probability theory as extended logic

"the mathematical rules of probability theory are [...] the unique consistent rules for conducting inference (i.e. plausible reasoning) of any kind"

Edwin T. Jaynes: Probability Theory – The Logic of Science (2003)



Deductive reasoning with probabilities

$$\frac{A \rightarrow B \qquad A}{B}$$

Deductive reasoning with probabilities

$$\frac{A \to B \quad A}{B}$$

Product rule: $\Pr(B \mid A \wedge X) \Pr(A \mid X) = \Pr(A \wedge B \mid X)$

$$\Pr(B \mid A \wedge \underbrace{A \to B}_{X}) \overset{\substack{\text{assuming} \\ \Pr(A \mid X) > 0}}{=} \frac{\Pr(A \wedge B \mid X)}{\Pr(A \mid X)}$$

$$\overset{\substack{\text{if } A \to B \text{ then} \\ A \wedge B = A}}{=} \frac{\Pr(A \mid X)}{\Pr(A \mid X)}$$

$$= 1$$

Plausible reasoning with probabilities

$$\frac{A \to B \qquad B}{A \text{ more plausible}}$$

Plausible reasoning with probabilities

$$\frac{A \to B \qquad B}{A \text{ more plausible}}$$

Product rule: $\Pr(B \mid A \wedge X)\Pr(A \mid X) = \Pr(A \mid B \wedge X)\Pr(B \mid X)$

$$\Pr(A \mid B \wedge \underbrace{A \to B}_{X}) \overset{\substack{\text{assuming} \\ \Pr(B|X)>0}}{=} \frac{\Pr(B \mid A \wedge X)\Pr(A \mid X)}{\Pr(B \mid X)} \overset{\Pr(B|A\wedge X)=1}{=} \frac{\Pr(A \mid X)}{\Pr(B \mid X)} \overset{\Pr(B|X)\leq 1}{\geq} \Pr(A \mid X)$$

$$\Pr(\theta \mid D, M)$$

D: data

θ: model parameters

M: model (all assumptions about θ and D)

$$\Pr(D \mid \theta, M)\ \Pr(\theta \mid M) = \Pr(\theta \mid D, M)\ \Pr(D \mid M)$$

$$\Pr(D \mid \theta, M)\, \Pr(\theta \mid M) \overset{D \leftrightarrow \theta}{\Longrightarrow} \Pr(\theta \mid D, M)\, \Pr(D \mid M)$$

Modeling   Inference

$$\underbrace{\Pr(D \mid \theta, M)}_{\textit{Likelihood}} \underbrace{\Pr(\theta \mid M)}_{\textit{Prior}} \overset{D \leftrightarrow \theta}{\Longrightarrow} \underbrace{\Pr(\theta \mid D, M)}_{\textit{Posterior}} \underbrace{\Pr(D \mid M)}_{\textit{Evidence}}$$

Bayes' rule

$$\underbrace{\Pr(\boldsymbol{\theta} \mid D, M)}_{\textit{Posterior}} = \frac{\overbrace{\Pr(D \mid \boldsymbol{\theta}, M)}^{\textit{Likelihood}}}{\underbrace{\Pr(D \mid M)}_{\textit{Evidence}}} \underbrace{\Pr(\boldsymbol{\theta} \mid M)}_{\textit{Prior}}$$

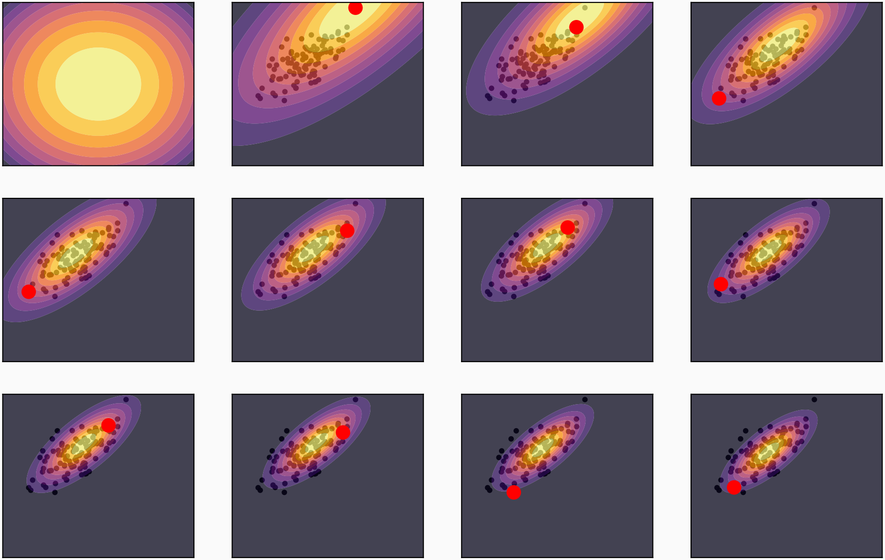
$$\Pr(D \mid M) = \int \Pr(D \mid \boldsymbol{\theta}, M)\, \Pr(\boldsymbol{\theta} \mid M)\, \mathrm{d}\boldsymbol{\theta}$$

Bayes' rule as evolutionary process

$$\underbrace{\Pr(\theta \mid D, M)}_{\textit{new population}} = \frac{\overbrace{\Pr(D \mid \theta, M)}^{\textit{fitness}}}{\underbrace{\Pr(D \mid M)}_{\textit{average fitness}}} \underbrace{\Pr(\theta \mid M)}_{\textit{current population}}$$

$$\Pr(D \mid M) = \int \Pr(D \mid \theta, M) \Pr(\theta \mid M) \, d\theta$$

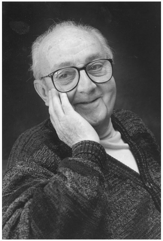
Bayes' rule as evolutionary process



Assigning probabilities

- **Pragmatism** (e.g. conjugate priors)
- **Symmetry and invariance** expressing ignorance
- The **maximum entropy principle**
- Solomonoff's **universal prior** favoring simpler explanations
- **Domain knowledge**, including constraints imposed by physical laws
- **Quantum mechanics**: probabilities via the Born rule

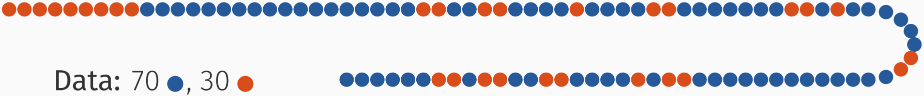
Probabilistic modeling



All models are wrong, but some models are useful.

George Box

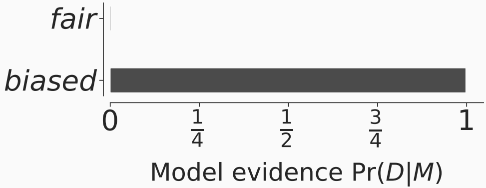
Tossing a coin



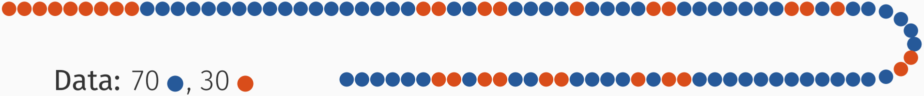
Data: 70 ●, 30 ●

Fair coin: $\Pr(\bullet \mid \mathit{fair}) = \frac{1}{2}$, $\Pr(\bullet \mid \mathit{fair}) = \frac{1}{2}$

Biased coin: $\Pr(\bullet \mid \theta, \mathit{biased}) = \theta$, $\Pr(\bullet \mid \theta, \mathit{biased}) = 1 - \theta$



Tossing a coin

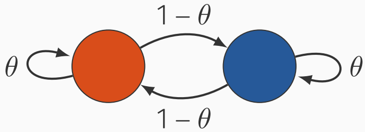


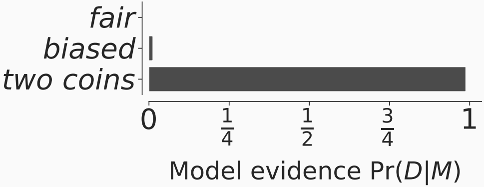
Data: 70 ●, 30 ●

Fair coin: Pr(● | *fair*) = $\frac{1}{2}$, Pr(● | *fair*) = $\frac{1}{2}$

Biased coin: Pr(● | θ, *biased*) = θ, Pr(● | θ, *biased*) = $1-\theta$

Two coins:





Bayesian epistemology

Data rarely rule out a model completely — they shift probability across competing models.

Models aren't true or false, just more or less plausible given the evidence.

Learning = updating and reweighting beliefs as new data arrive.

Scientific inference is a process of gradual refinement rather than abrupt rejection.

Conclusion

Consider probabilities as **degrees of implication**

Comply with the rules: the **sum rule** and the **product rule** (Bayes' rule is a corollary)

Come up with **useful models**

Compute the **posterior** and the **evidence**

Compare and challenge your models