Induction and Deduction in Statistics

{Kevin T. Kelly, Konstantin Genin}
Carnegie Mellon Universisty

Belgrade 2016

The First Cut in the Philosophical Pie

- All the **objects of human...enquiry** may naturally be divided into **two kinds**, to wit,
- ***Relations of Ideas***, and
- ***Matters of Fact***.

David Hume, *Enquiry*, Section IV, Part 1.

The First Cut in the Philosophical Pie

- Any ... **inference** in science belongs to one of **two kinds**:
  1. either it yields **certainty** in the sense that the conclusion is **necessarily true**, provided that the premises are true,
  2. or it does not.
- The first kind is that of **deductive inference**...
- The second kind will here be called **'inductive inference'**.
- R. Carnap, *The Continuum of Inductive Methods*, 1952, p. 3 .

The First Cut in the Philosophical Pie

- **Deductive inference:**
  - Truth preserving.
  - Stable (monotonic).
  - Non-ampliative.

- **Inductive inference:**
  - Everything else.

Inference in Science

Deduction

- Calculation
- Refuting universal H
- Verifying existential H
- Deciding between universal H, H'
- Predicting E from H
- Hypotheses compatible with E

Induction

- Inferring universal H
- Choosing between universal H_0 , H_1 , H_2 , ...

Inference in Science

Deduction

- Calculation
- Refuting universal H
- Verifying existential H
- Deciding between universal H, H'
- Predicting E from H
- Hypotheses compatible with E

Induction

- Inferring universal H
- Choosing between universal H_0, H_1, H_2, ...

Assuming **deterministic** data!

Inference in Science

Deduction

- Calculation
- Refuting universal H
- Verifying existential H
- Deciding between universal H, H'
- Predicting E from H
- Hypotheses compatible with E

Induction

- Inferring universal H
- Choosing between universal H_0, H_1, H_2, ...

Assuming **stochastic** data...

Inference in Science

Deduction

- Calculation
- Refuting universal H
- Verifying existential H
- Deciding between universal H, H'
- Predicting E from H
- Hypotheses compatible with E

Induction

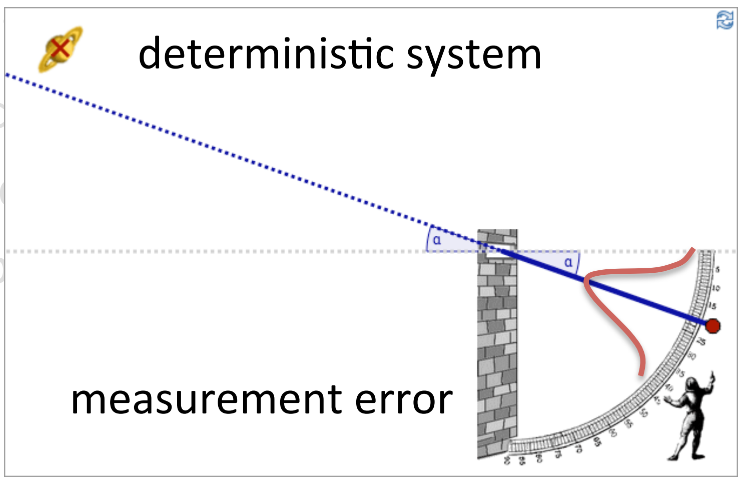
- Inferring universal H
- Choosing between universal H_0, H_1, H_2, ...
- Refuting universal H
- Verifying existential H
- Deciding between universal H, H'
- Predicting E from H
- Hypotheses compatible with E

Assuming **stochastic** data...

Inference in Science

Deduction

- Calculation
- Refuting universal *H*
- Ver
- Ded
- Pre
- Hyp



Induction

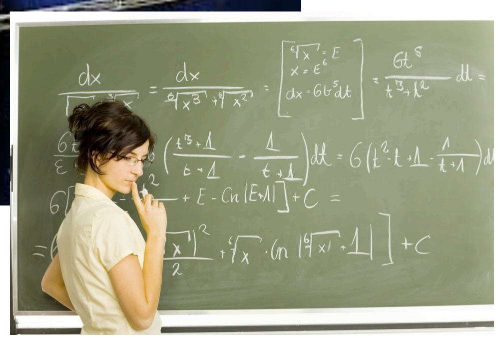
- Inferring universal *H*
- Choosing between universal H_0, H_1, H_2, ...
- Refuting universal *H*
- Verifying existential *H*
- Deciding between universal *H, H'*
- Predicting *E* from *H*
- Hypotheses compatible with *E*

and **all** continuous measurement is stochastic!

Inference in Science

Deduction

- Calculation
- Refuting universal *H*
- Verifying existential *H*



H’

Induction

- Inferring universal *H*
- Choosing between universal H_0, H_1, H_2, ...
- Refuting universal *H*
- Verifying existential *H*
- Deciding between universal *H, H’*
- Predicting *E* from *H*
- Hypotheses compatible with *E*

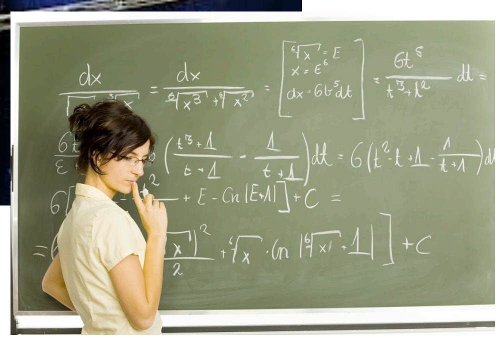
Also, **real** calculation occasions probable error from numerical approximation, human error, and cosmic rays.

Inference in Science

Deduction

- Calculation
- Refuting universal *H*
- Verifying existential *H*

H’



Induction

- Inferring universal *H*
- Choosing between universal H_0, H_1, H_2, ...
- Refuting universal *H*
- Verifying existential *H*
- Deciding between universal *H*, *H’*
- Predicting *E* from *H*
- Hypotheses compatible with *E*
- Calculation

Also, **real** calculation occasions probable error from numerical approximation, human error, and cosmic rays.

Inference in Science

Deduction

- Calculation
- Refuting universal *H*
- Verifying existential *H*
- Deciding between universal *H*, *H'*
- Predicting *E* from
- Hypotheses comp

Booooring!

Induction

- Inferring universal *H*
- Choosing between universal H_0, H_1, H_2, …
- Refuting universal *H*
- Verifying existential *H*
- Deciding between universal *H*, *H'*
- Predicting *E* from *H*
- Hypotheses compatible with *E*
- Calculation

A More Revealing First Cut

	Ideal	Statistical
Deduction	Refuting universal H_0 Verifying existential H_1 Deciding between universal H_0, H_1 Predicting E from H Hypotheses compatible with E Ideal calculation	Rejecting simple H_0 Accepting composite H_1 Deciding between simple H_0, H_1 Direct inference from simple H Confidence interval Real calculation
Induction	Inferring universal H_0 Choosing between universal H_0, H_1, H_1, ...	Inferring simple H_0 Model selection

Ideal Methods

Deductive

- Stable
- Guaranteed to avoid error

Inductive

- Unstable
- Not guaranteed to avoid error

Statistical Methods

Deductive

- Stable in chance
- Guaranteed low chance of error

Inductive

- Unstable in chance
- No guarantee of low chance of error.

Deeper Question

Can one represent deductive statistical methods as literally **deducing** their conclusions from **statistical information**?

Deeper Question

Can one represent deductive statistical methods as literally **deducing** their conclusions from **statistical information**?

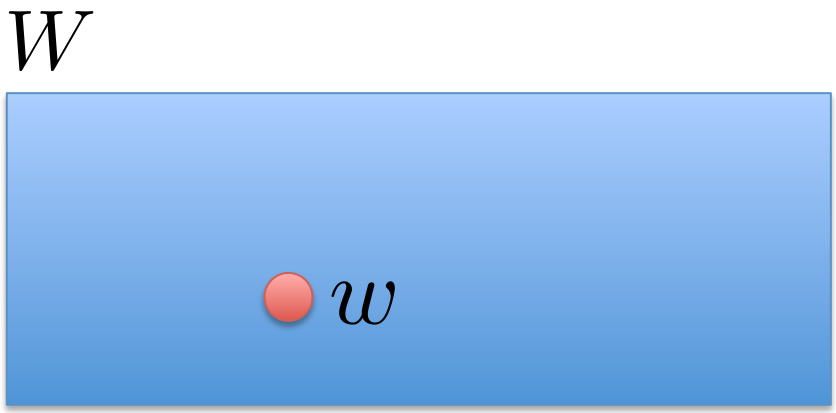
Yes.

The Structure of Ideal Information

- ✗ Logic
- ✗ Probability
- ✓ Topology

Worlds

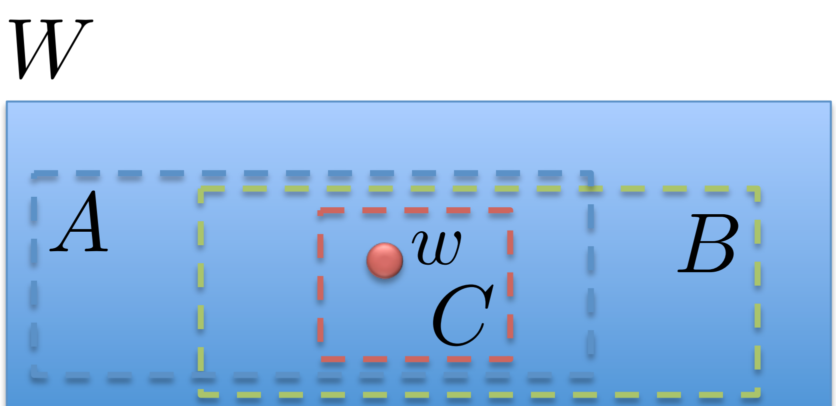
- The points in W are **possible worlds**.



The Structure of Information

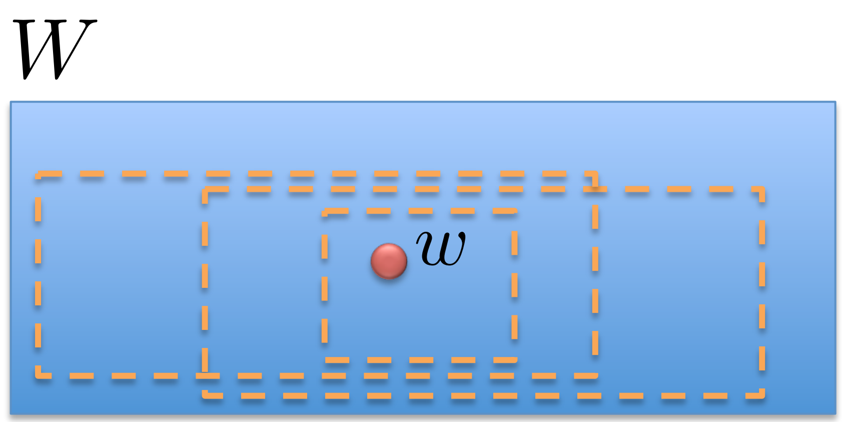
An **information basis** I is a countable set of propositions called **information states** such that :

1. each world makes some information state true;
2. each pair of true information states is entailed by a true information state.



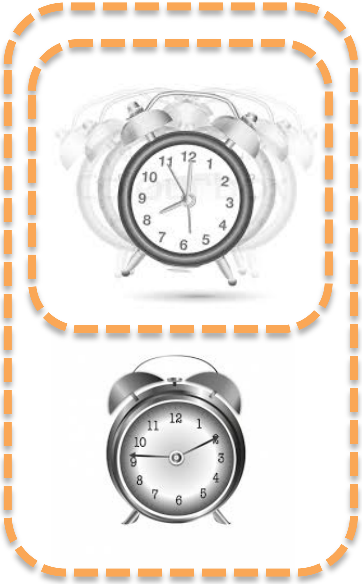
The Structure of Information

$$\mathcal{I}(w) \ := \ \{E \in \mathcal{I} : w \in E\}.$$



Simplest Example: Alarm Clock

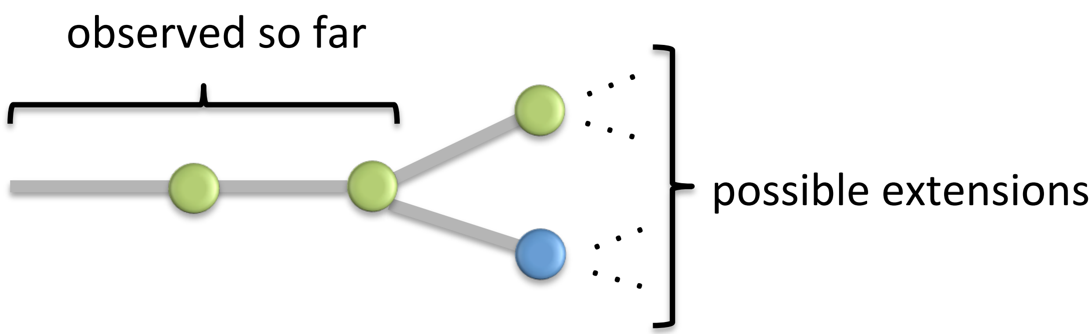
- The theorist is awakened by her graduate students only when her theory is refuted.



Example: Sequential Binary Experiment

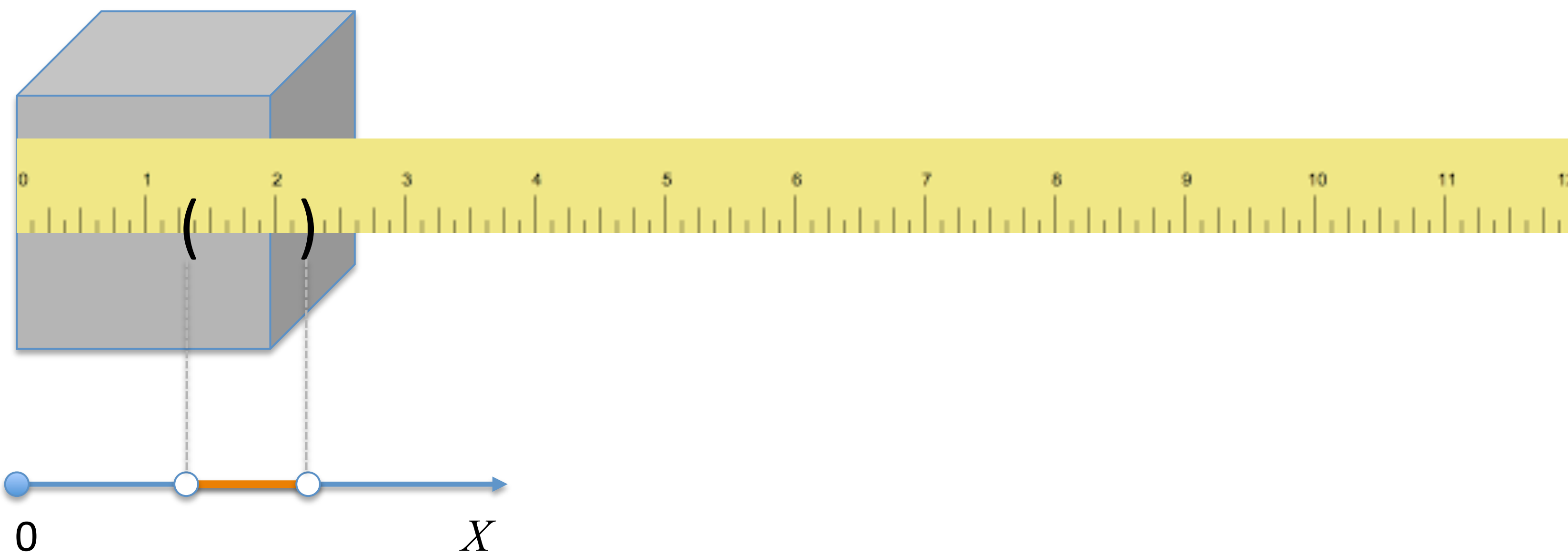
Worlds = infinite discrete sequences of outcomes.

Information states = cones of possible extensions:



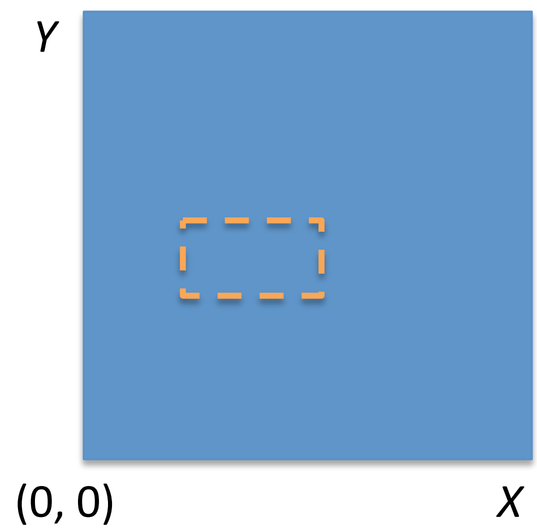
Example: Measurement of X

- **Worlds** = real numbers.
- **Information states** = open intervals.



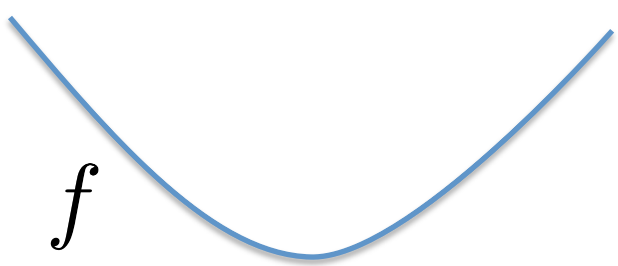
Example: Joint Measurement

- **Worlds** = points in real plane.
- **Information states** = open rectangles.



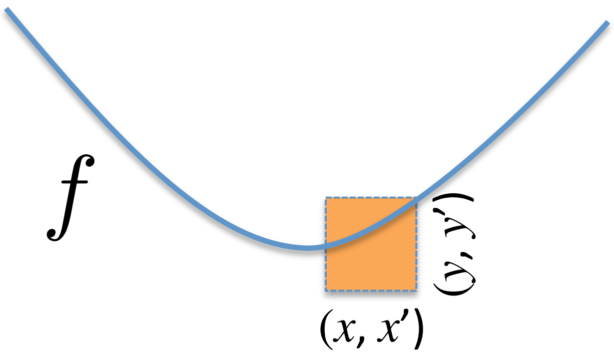
Example: Equations

- **Worlds** = functions $f : \mathbb{R} \to \mathbb{R}.$



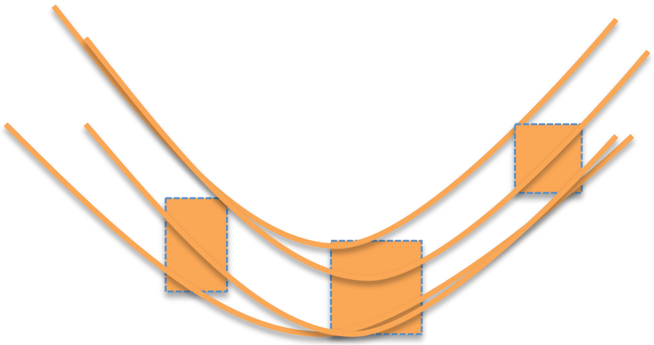
Example: Laws

- An **observation** is a joint measurement.



Example: Laws

- The **information state** is the set of all worlds that touch each observation.

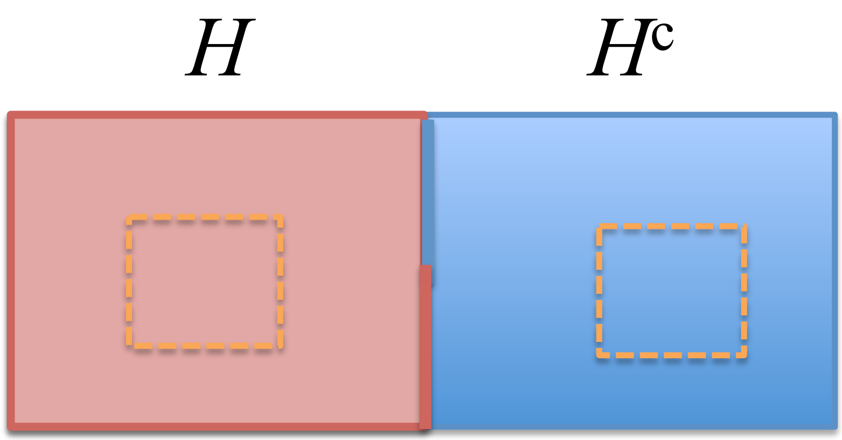


Deductive Verification and Refutation

H is **verified** by E iff $E \subseteq H$.

H is **refuted** by E iff $E \subseteq H^c$.

H is **decided** by E iff H is either verified or refuted by E.

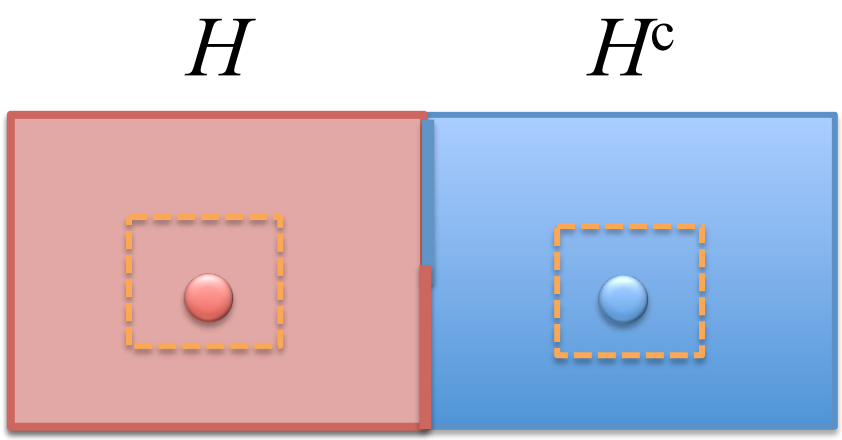


Will be Verified

w is an **interior [exterior] point** of H iff

iff H **will be** verified [refuted] in w

iff there is $E \in \mathcal{I}(w)$ s.t. H is verified [refuted] by E.

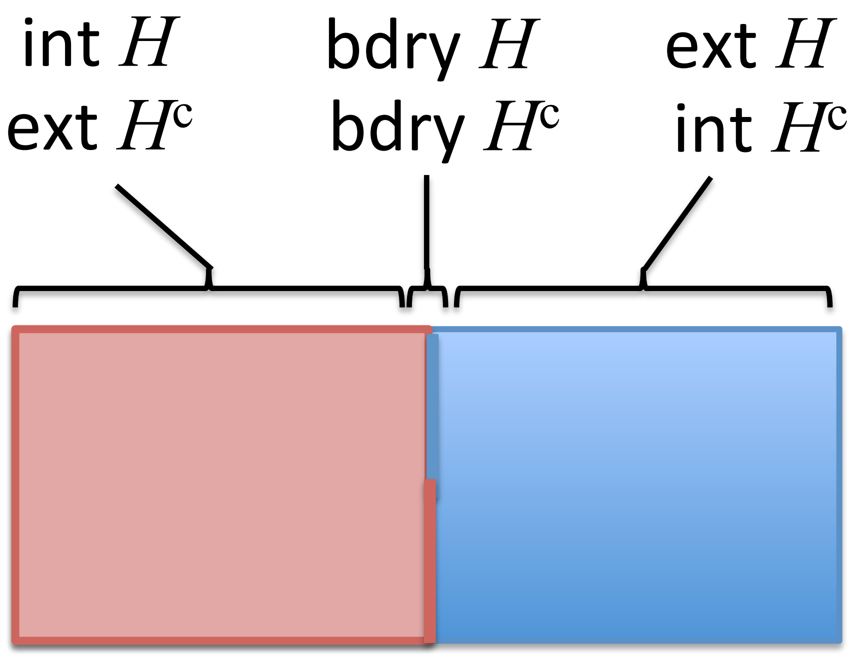


Will be Verified

int H := the proposition that H **will be verified.**

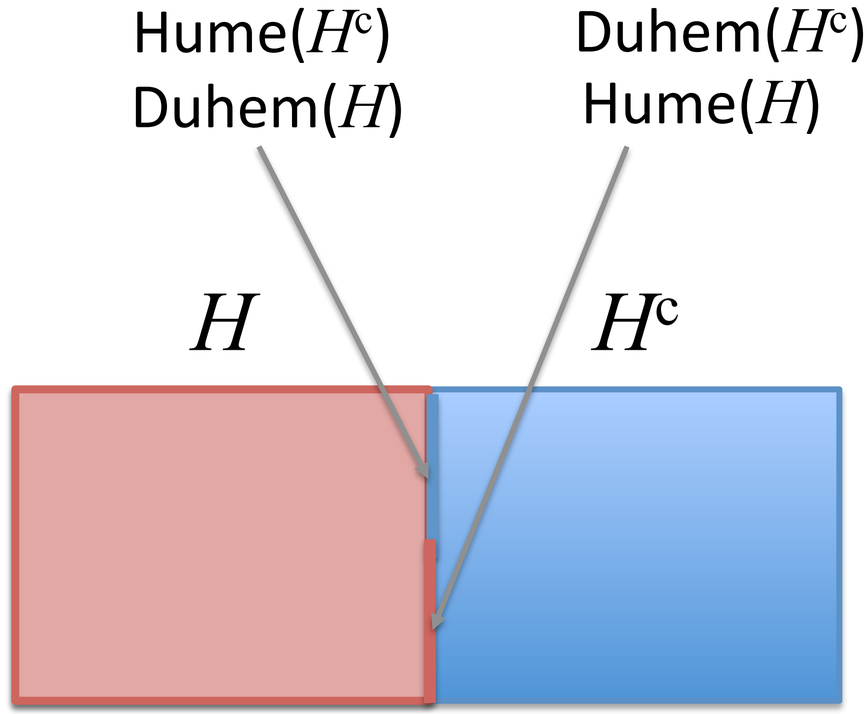
ext H := the proposition that H **will be refuted.**

bdry H := the proposition that H **will never be decided.**



Will be Verified

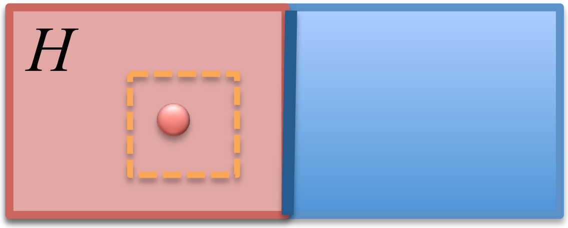
- bdry(H) ∩ H = "you face **Hume's problem** w.r.t. H";
- bdry(H) ∩ H^c = "you face **Duhem's problem** w.r.t. H"



Verifiability, Refutability, Decidability

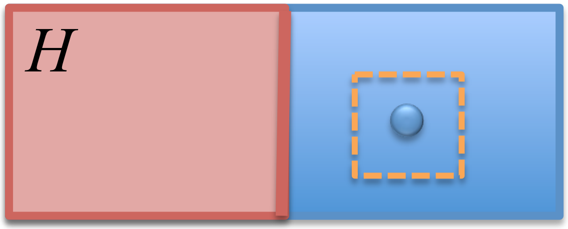
H is **verifiable** iff $H \subseteq \mathrm{int}(H)$.

i.e., iff H will be verified however H is true.

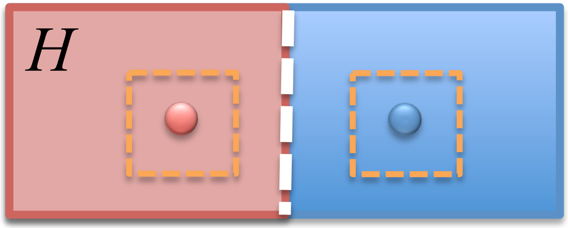


H is **refutable** iff $\mathrm{cl}(H) \subseteq H$.

i.e., iff H will be refuted however H is false.



H is **decidable** iff H is both verifiable and refutable.



Methods

- A **verification method** for H is an inference rule $V(E) = A$ such that in every world w:
  1. $w \in H$: V converges to H without error.
  2. $w \in H^c$: V always concludes W.

Methods

- A **verification method** for H is an inference rule $V(E) = A$ such that in every world w:
  1. $w \in H$: V converges to H without error.
  2. $w \in H^c$: V always concludes W.
- A **refutation method** for H is just a verification method for H^c.
- A **decision method** for H converges to H or to H^c without error.

Methods

- A **verification method** for H is an inference rule $V(E) = A$ such that in every world w:
  1. $w \in H$: V converges to H without error.
  2. $w \in H^c$: V always concludes W.
- A **refutation method** for H is just a verification method for H^c.
- A **decision method** for H converges to H or to H^c without error.
- H is **methodologically verifiable [refutable, decidable]** iff H has a method of the corresponding kind.

Verification is Deductive

Proposition (truth preservation and non-ampliativity).

If V is a verifier, refuter or decider for H and $V(E) = A$,

then $E \subseteq A$.

Proposition (monotonicity).

If there is a verifier, refuter or decider for H, then there is a monotonic one that never drops H or H^c after having concluded it.

Topology

Let $\mathcal{I}^*$ denote the closure of $\mathcal{I}$ under union.

Proposition:
If $\mathfrak{I} = (W, \mathcal{I})$ is an information basis
then $\mathfrak{I} = (W, \mathcal{I}^*)$ is a topological space.

- H is **open** iff $H \in \mathcal{I}^*$.
- H is **closed** iff H^c is open.
- H is **clopen** iff H is both closed and open.

Methodology = Topology

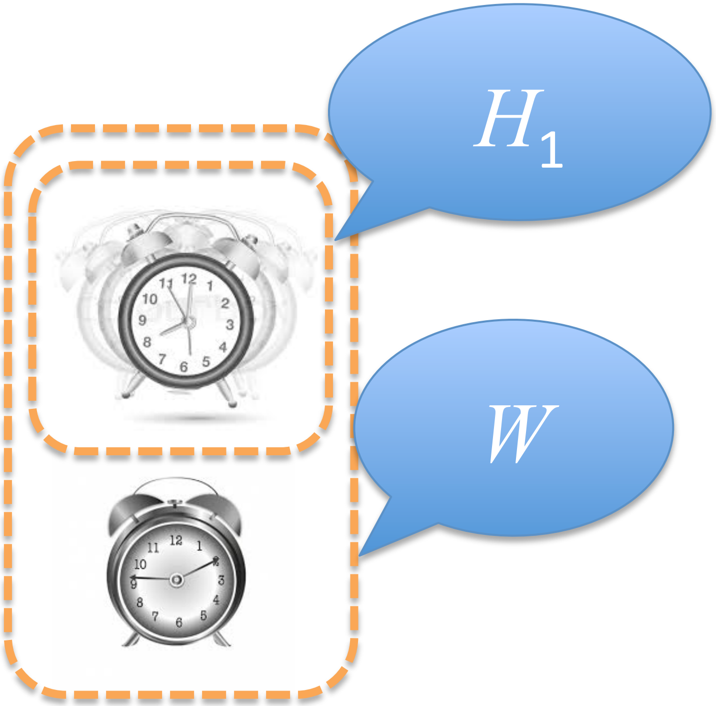
Proposition.

1. **open** = verifiable = methodologically verifiable.
2. **closed** = refutable = methodologically refutable.
3. **clopen** = decidable = methodologically decidable.

Simplest Example

H_0 = "I will never be awakened" is closed.

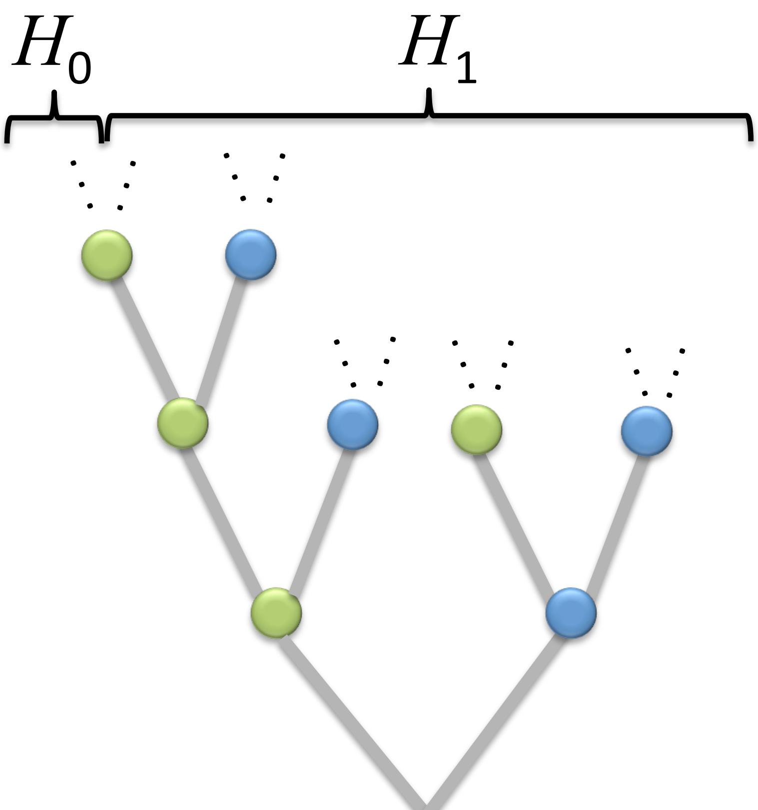
H_1 = "I will eventually be awakened" is open.



Sequential Examples

H_0 = "every outcome is green" is closed.

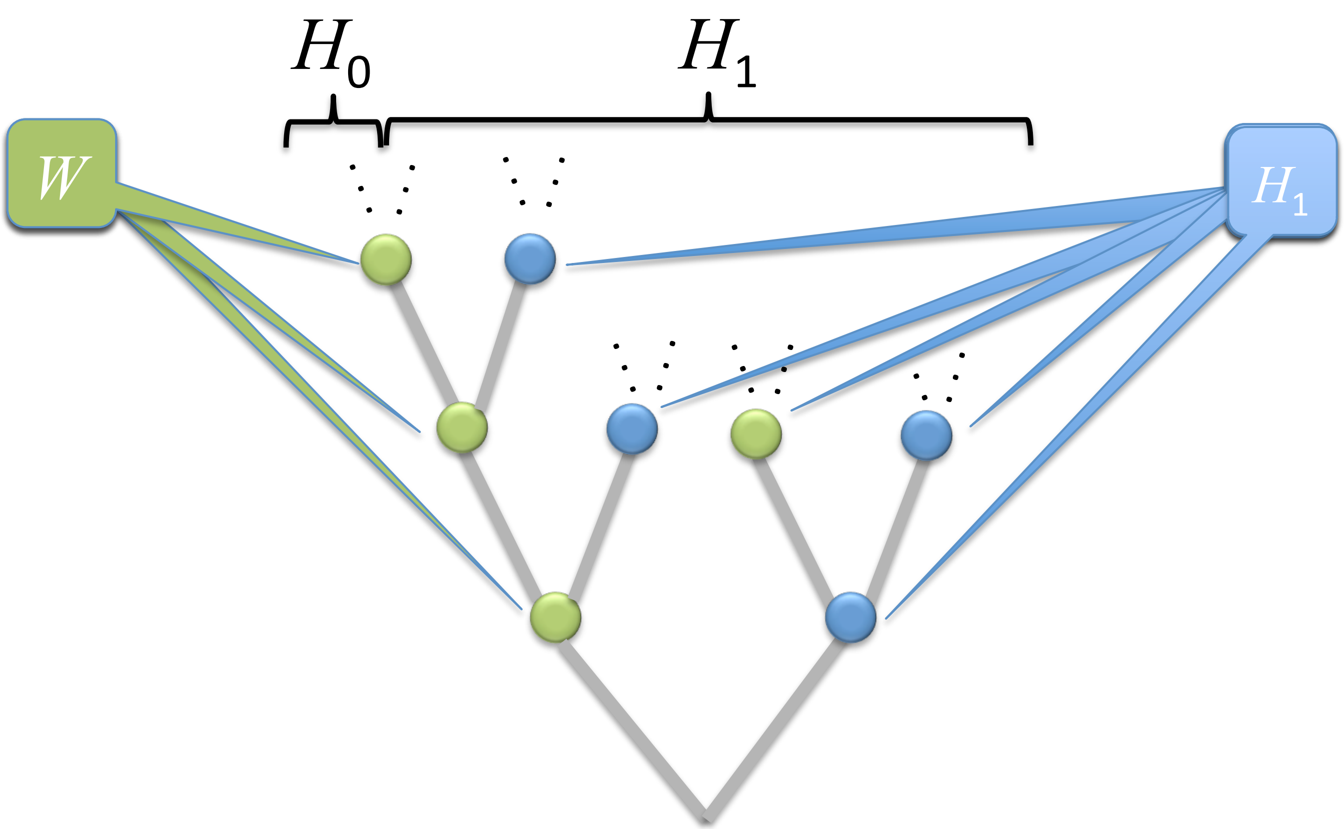
H_1 = "some outcome is blue" is open.



Sequential Examples

H_0 = "every outcome is green" is closed.

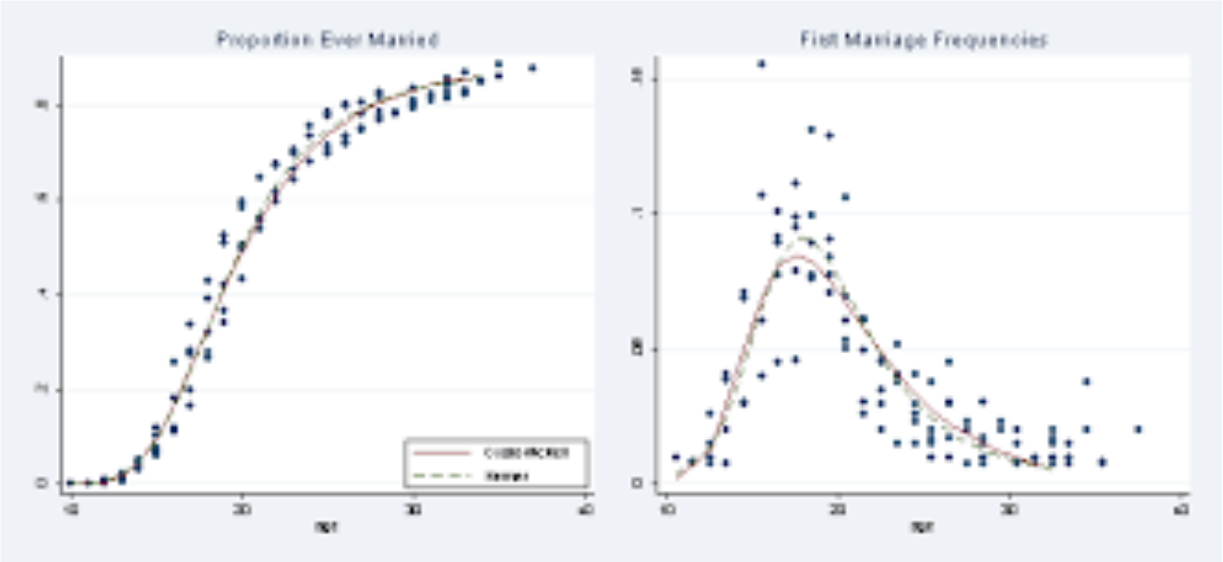
H_1 = "some outcome is blue" is open.



STATISTICAL DEDUCTION AND ITS TOPOLOGY

Statistical Methodology

- Does information topology also shed light on statistically deductive methods and problems?



Skepticism

The approach...

"may be okay if the candidate theories are **deductively related** to observations, but when the relationship is **probabilistic**, I am **skeptical** ...".

Eliott Sober, *Ockham's Razors*, 2015

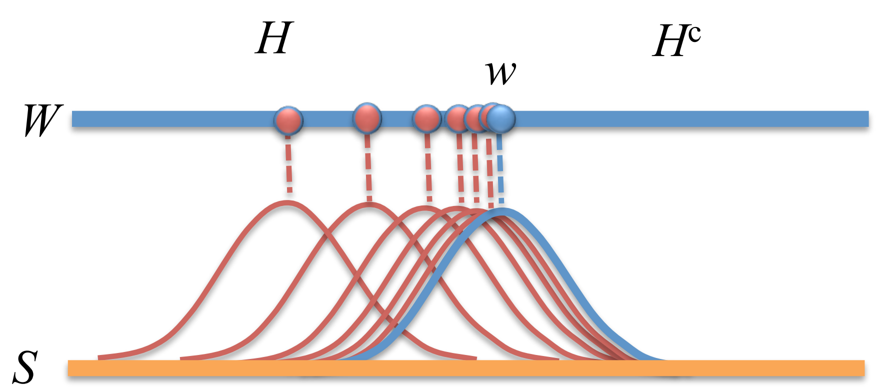
Skepticism

- If it's interesting to **guess** that something is **impossible**...
- then it's surely interesting to **demonstrate** that it is **necessary**.

Eliott Sober, *Ockham's Razors*, 2015

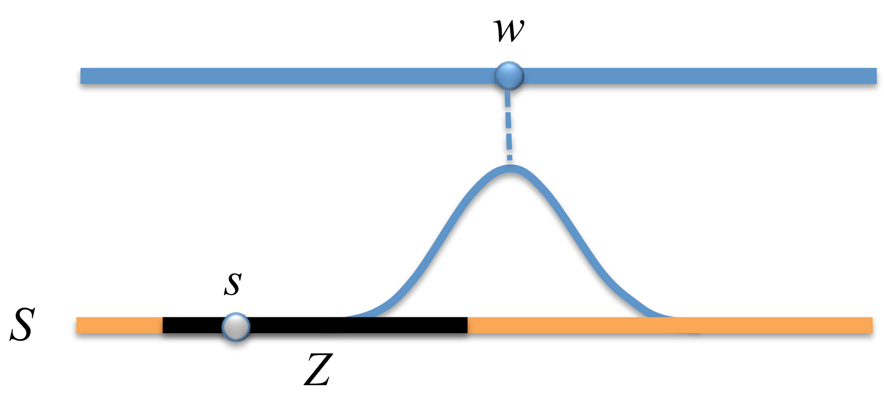
Statistical Information Topology

Possibilities nearer to the truth should be harder to rule out by statistical information.



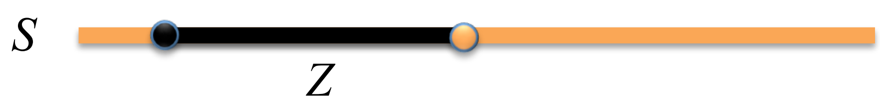
Gathering Statistical Information

1. The sample space S has its own (metrizable) topology.
2. Choose a sample event Z over S.
3. Obtain sample s.
4. Observe whether Z occurs.



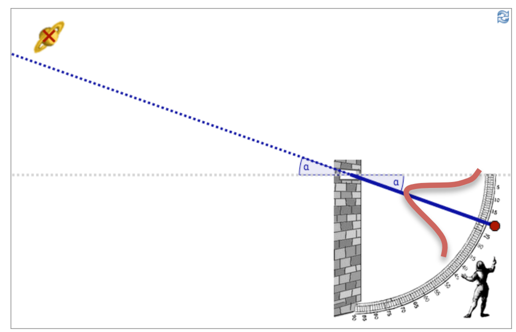
Feasible Sample Events

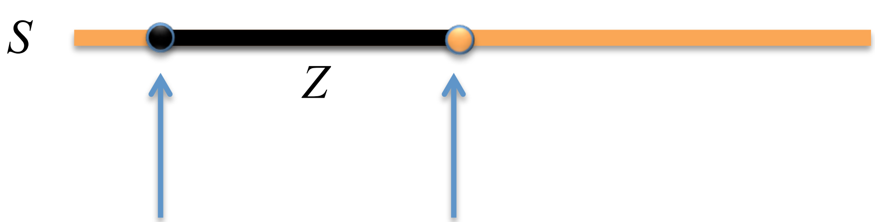
- But every non-trivial Z on the real line has boundary points.



Feasible Sample Events

- You can't really determine whether a sample hits exactly on the boundary.



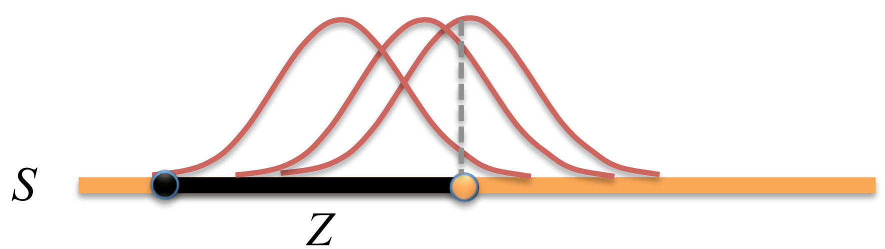


Feasible Sample Events

- That doesn't matter statistically as long as the boundary carries 0 probability.
- So Z is an observationally **feasible** sample event iff

$$p(\text{bdry } Z) = 0, \text{ for each } p \text{ in } W.$$

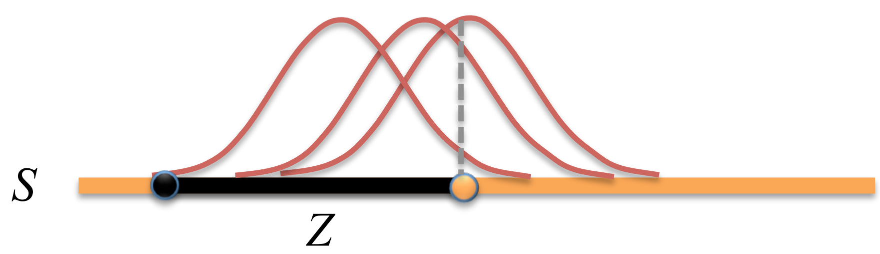
- I.e, feasible Z is **almost surely clopen** (decidable) in S.



Feasible Statistical Models

- S is **feasible** for W iff

  S has a countable topological basis of feasible zones.

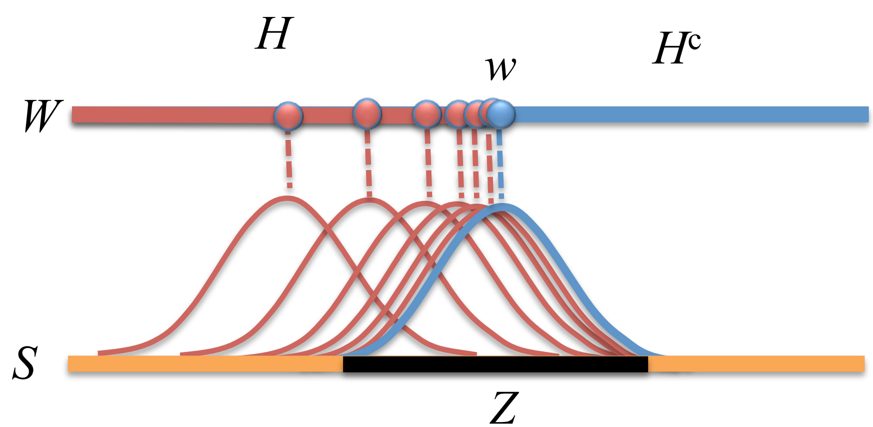


Statistical Information Topology

$w \in \mathrm{cl}(H)$ iff

H contains a sequence of worlds $w_0, \ldots, w_n, \ldots$ such that for every feasible sample event $Z \subseteq S$:

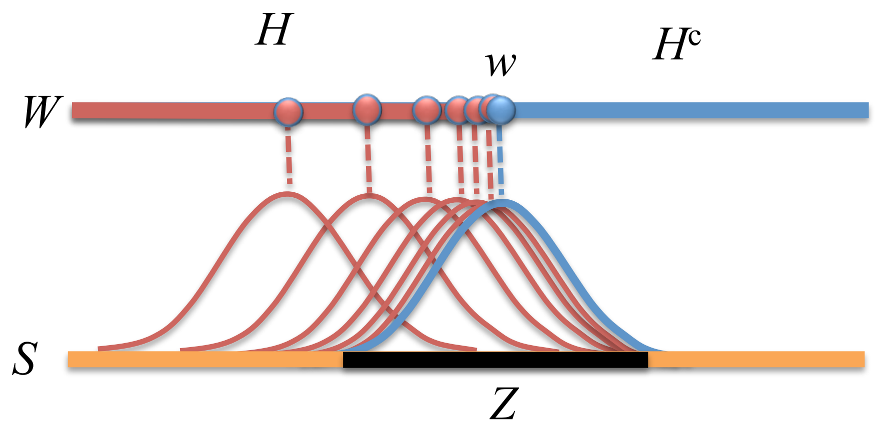
$$\mathrm{Lim}_{i \to \infty} p_{w_i}(Z) \;=\; p_w(Z).$$



For Those Who Care

Proposition: Assuming that S is feasible for W,

statistical information topology = weak topology.



IID Sampling

Proposition.

If S is feasible for W, then:

1. S^N is feasible for the IID product measures w^N such that $w \in W$;

2. The information topology on W^N is **homeomorphic** to the information topology on W.

Feasible Statistical Methods

A **feasible statistical method** at sample size N is a function M^N from sample events in S^N to propositions over W such that:

$$(M^N)^{-1}(H) \text{ is } \textbf{feasible}.$$

A **feasible statistical method** is a collection

$$\{M^N : N \in \mathbf{N}\}$$

of feasible statistical methods at each sample size.

Statistical Verification Methods

- A **statistical verification method** at level $\alpha > 0$ for H is a feasible statistical method $\{V^N : N \in \mathbf{N}\}$ with range $\{W, H\}$ such that:

  1. for $w \in H$: $\mathrm{Lim}_N p_w^N(V^N = H) = 1;$
  2. for $w \notin H$: $p_w^N(V^N = H) \leq \alpha,$ for all N.

- A statistical **refutation** method at level $\alpha > 0$ for H is a statistical verification method for H^c.
- A statistical **decision** method at level $\alpha > 0$ for H is both.

- H is **statistically verifiable [refutable, decidable]** iff H has a statistical verification [refutation, decision] method at each level $\alpha > 0$.

The Topology of Statistical Methodology

Proposition. Suppose that S is feasible for W. Then:

1. **open** = statistically verifiable.
2. **closed** = statistically refutable.
3. **clopen** = statistically decidable.

Stability

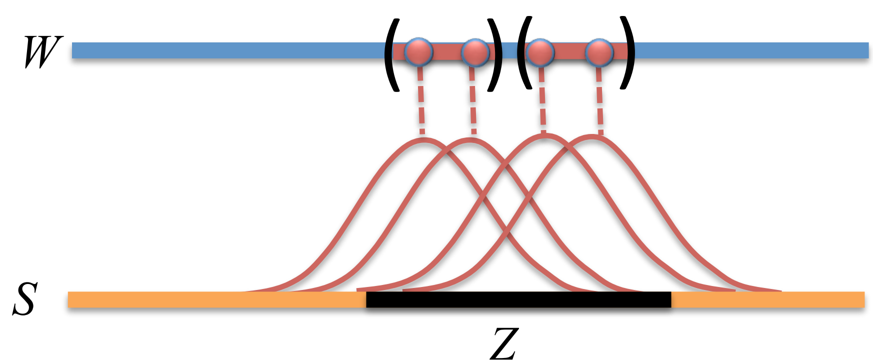
Conjecture: The methods can be constructed to be **monotonic** in chance of producing H.

Conjecture: the α level can be made to converge **monotonically** to 0.

Information Basis

Define **intervals** of worlds w.r.t. Z:

$$E_Z(a,b) \;=\; \{v \in W : a < p_v(Z) < b\}.$$



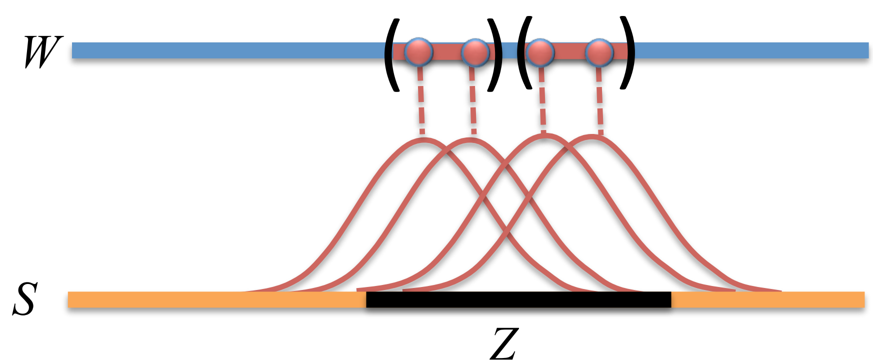
Information Basis

Proposition. Let $\mathcal{B}$ be a feasible, countable basis for S that (w.l.o.g) is closed under finite intersection.

Then:

$$\mathcal{I} \;=\; \{E_Z(a,b) : a, b \in \mathbb{Q} \;\wedge\; a < b \;\wedge\; Z \in \mathcal{B}\}$$

is a **countable basis** for the statistical information topology on W.



Ideal Statistical Information

Think of statistically verified basis elements as ideal statistical information available for other methods.

Let $\{E_0, \ldots, E_i, \ldots\}$ enumerate $\mathcal{I}$.

Let V_i statistically verify E_i at level $\alpha/2^i$.

Statistical Verification as Deduction

Let H be open (statistically verifiable).

Let $V_H(s) = H$ iff there exists $\boldsymbol{E_i} \subseteq \boldsymbol{H}$ such that $V_i(s) = E_i$.

Proposition. V statistically verifies H at level α.

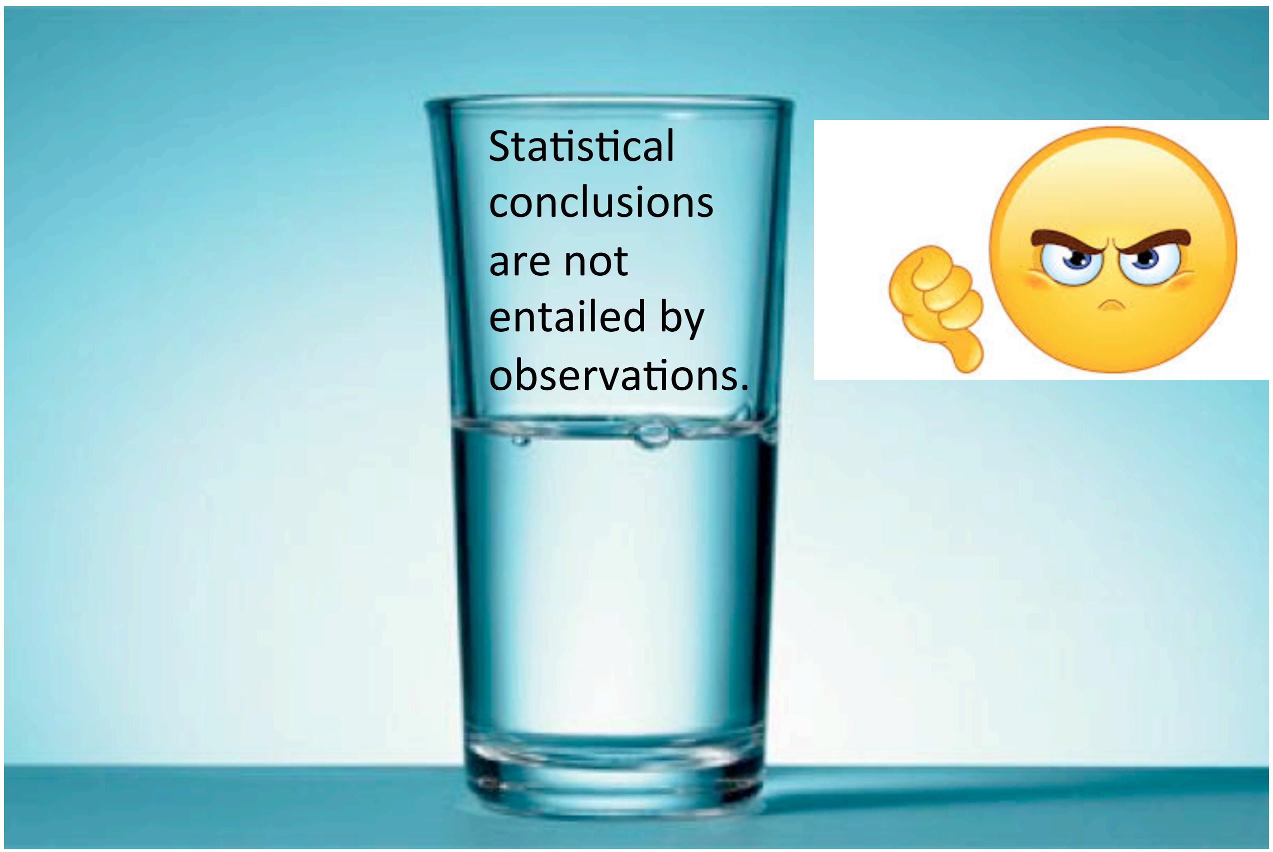
Moral. You can safely **think of** statistical verification of H as literal **deduction** of H from **ideal statistical information** E_i.

Similarly for refutation and decision.

Statistical Verification as Deduction

	Ideal	Statistical
closed	universal H_0	simple null H_0
open	existential H_1	composite alternative H_1
clopen	exhaustive universal H_0, H_1.	exhaustive simple H_0, H_1.

Statistical Verification as Deduction



Statistical Verification as Deduction

EXTENSION TO STATISTICAL INDUCTIVE INFERENCE

Inductive Inference

Ideal method D **converges** to H in w iff

there exists $E \in \mathcal{I}(w)$ such that

for all $F \in \mathcal{I}(w)$ for which $F \subseteq E$,

$$D(F) \subseteq H.$$

Statistical method D **converges** to H in w iff

$$\mathrm{Lim}_{N \to \infty}\, p_w^N(D = H) = 1.$$

Inductive Inference

- D **verifies** H **in the limit** iff:

$$w \in H \quad \Leftrightarrow \quad D \text{ converges to } H.$$

- D **refutes** H **in the limit** iff D verifies H^{c} in the limit.
- D **decides** H **in the limit** iff D both verifies and refutes H in the limit.

Borel Hierarchy*

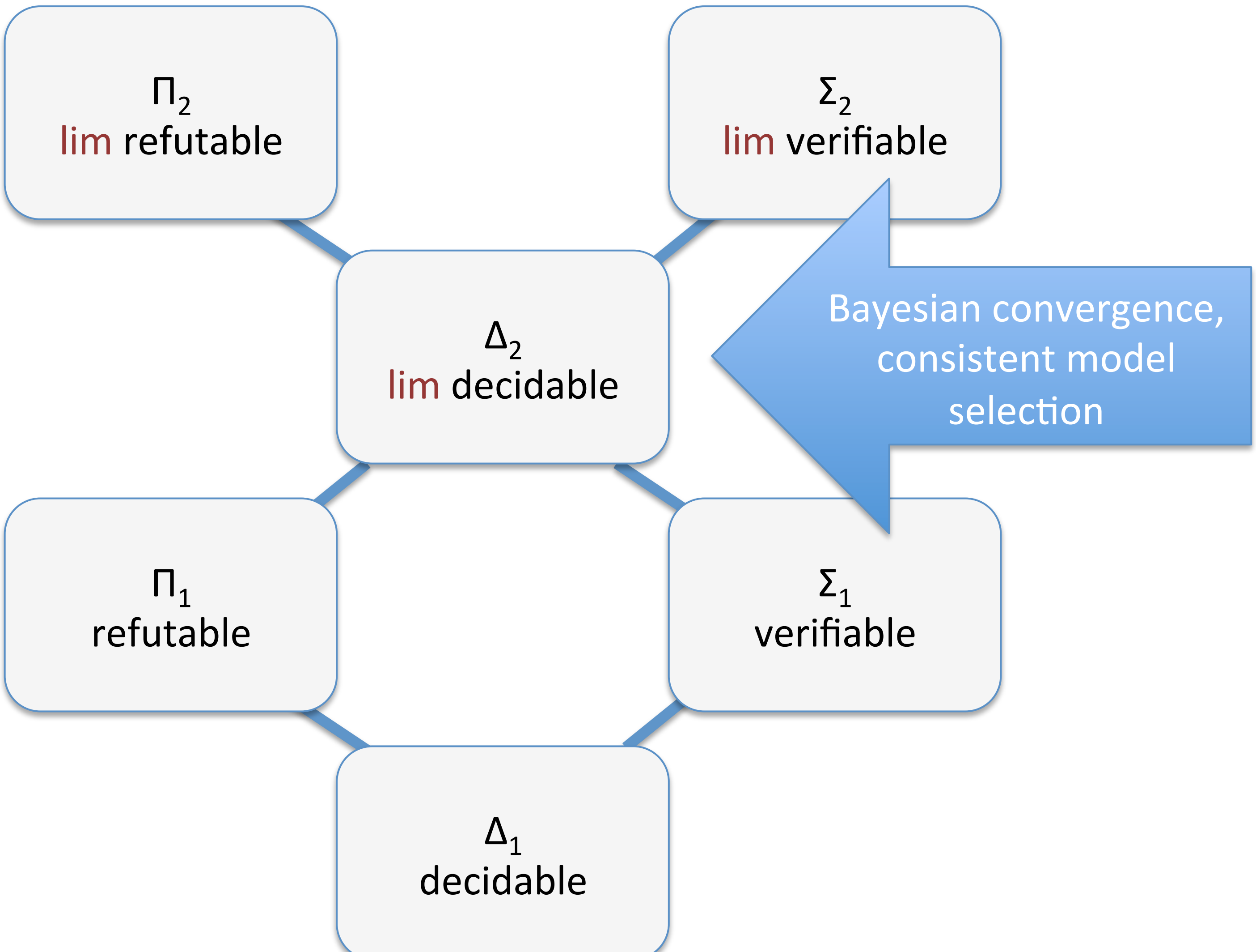
$$
\begin{tikzcd}
\begin{array}{c}\Pi_2 \\ \text{count. int. of } \Sigma_1\end{array} \arrow[dr, dash] & & \begin{array}{c}\Sigma_2 \\ \text{count. un. of } \Pi_1\end{array} \arrow[dl, dash] \\
& \begin{array}{c}\Delta_2 \\ \text{both } \Pi_2 \text{ and } \Sigma_2\end{array} \arrow[dl, dash] \arrow[dr, dash] & \\
\begin{array}{c}\Pi_1 \\ \text{closed}\end{array} \arrow[dr, dash] & & \begin{array}{c}\Sigma_1 \\ \text{open}\end{array} \arrow[dl, dash] \\
& \begin{array}{c}\Delta_1 \\ \text{clopen}\end{array} &
\end{tikzcd}
$$

*Assuming the topology is:

metrizable;

has countable basis.

Both Ideally and Statistically



EXTENSION TO OCKHAM'S RAZOR

Simplicity

- Simplicity can be defined topologically:

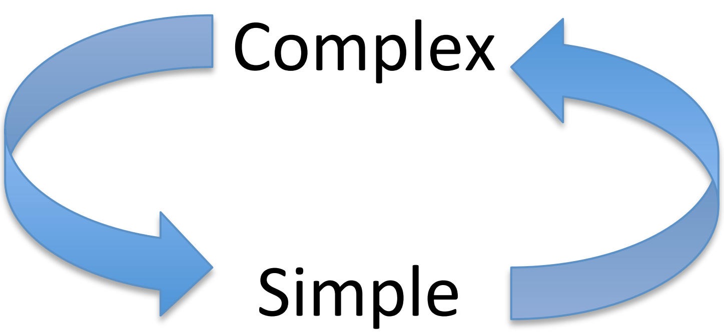
$$A < B \;\Leftrightarrow\; A \cap \mathsf{cl}(B) \cap B^c \neq \varnothing.$$

Epistemic Argument for the Razor

Ideal case:

- If you **violate Ockham's razor** then

1. either you **fail to converge** to the truth or
2. nature can force you into a **cycle of opinions** (complex-simple-complex), even though such cycles are avoidable.

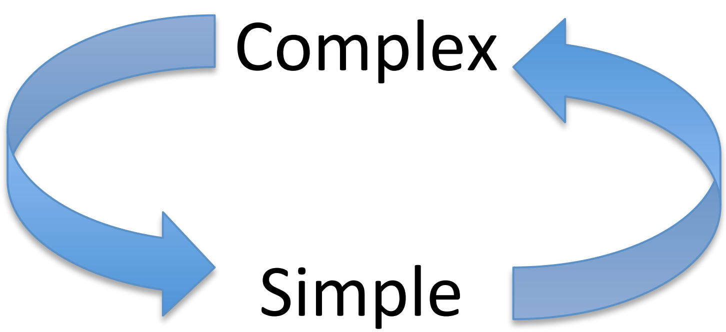


Epistemic Argument for the Razor

Statistical case:

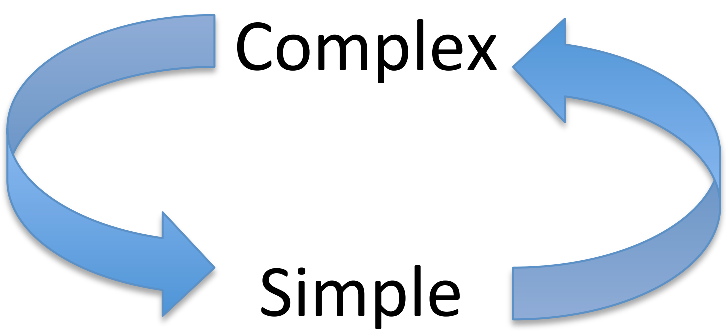
- If you **violate Ockham's razor** with chance α, then

1. either you **fail to converge** to the truth in chance or
2. nature can force you into an α-**cycle of opinions** (complex-simple-complex), even though such cycles are avoidable.



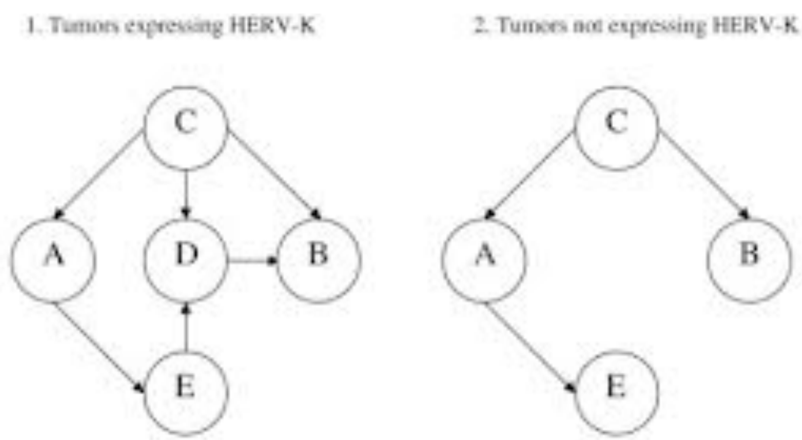
No Assumption that Reality s Simple

Indeed, by **favoring** a **complex** hypothesis, you incur the cycle in a **complex** world!



Application: Causal Inference from Non-experimental Data

- **Causal network** inference from **retrospective** data.
- That is an **inductive** problem.
- The search is strongly guided by **Ockham's razor**.
- We have the **only non-Bayesian foundation** for it.



Symbols: "A" = cause (YFV); "B" = outcome (cancer); "C" = confounders (recreational solar exposure and high social class); "D" and "E" = mediators (HERV-K antigen and immune response).

Application: Science

- All scientific conclusions are supposed to be counterfactual.
- Scientific inference is strongly simplicity biased.
- Standard ML accounts of Ockham's razor do not apply to such inferences (J. Pearl).
- Our account does.

A New Objective Bayesianism?

How much **prior bias toward simple models** is necessary to avoid α-cycles?

A Method for Methodology

1. **Develop** basic methodological ideas in **topology**.
2. **Port** them to **statistics** via the **statistical information topology**.

Some Concluding Remarks

1. Information topology is the **structure** of the scientist's **problem context**.
2. The apparent **analogy** between statistical and ideal verifiability reflects **shared topological structure**.
3. Indeed, one can **think of** basis elements as **propositional statistical information** from which statistical conclusions can be **literally deduced**.
4. Thereby, **ideal logical/topological ideas** can be **ported** in a direct and uniform fashion to statistics.