# Statistical Tests with Distributional Uncertainty: An Info-Gap Approach

**Yakov Ben-Haim**

**Technion**

**Israel Institute of Technology**

[0]lectures\talks\lib\isipta2009talk01.tex 25.6.2009

# 1 *INFO-GAP THEORY*

**In the beginning,**

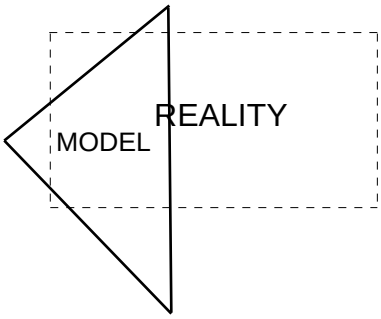
**God created the heavens and the earth.**

**And the earth was total confusion ...**

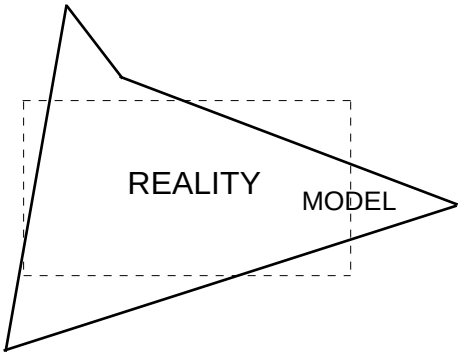
# . . . so humans started making models . . .

# Out there is a . . .

# We build models which, well, . . .



# . . . but over time . . .



## § Scientific optimism, philosophical positivism:

$$\lim_{t\to\infty} \textbf{MODEL} = \textbf{REALITY}$$

## § Scientific optimism, philosophical positivism:

$$\lim_{t\to\infty} \textbf{MODEL} = \textbf{REALITY}$$

## § Motivations:

- **Truth.**
- **Utility.**

## But...

§ **The art of designing, deciding, planning:**

**Use the wrong model**

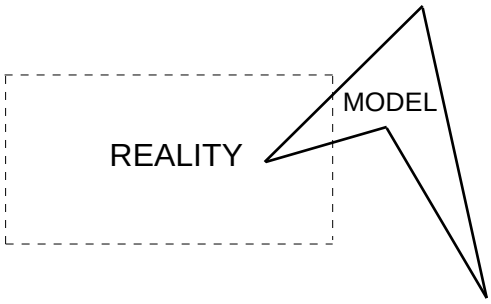
**to make the right decision**

**(when the right model is unknown).**

## § Robert Burns, Ode to a Mouse:

**The best laid schemes o' mice and men,**
**gang oft agley,**
**And leave us nought but grief and pain,**
**for promised joy.**

**Still, thou are blest, compared wi' me.**
**The present only toucheth thee.**
**But Och! I backward cast my eye**
**on prospects drear.**
**And forward, though I cannay see,**
**I hope and fear.**




§ **Evaluate and select a design under severe uncertainty.**

§ **Info-gaps:**

- **Incomplete understanding.**
- **Erroneous data.**
- **Changing conditions.**
- **Sur$_{\mathbf{p}}$rises.**

## § Info-gap decision strategies:

- **Robust-satisficing:**

  **protecting against uncertainty.**

- **Opportune-windfalling:**

  **exploiting uncertainty.**

§ Sources: http://info-gap.com

§ Other issues:

- Robustness and opportuneness.
- Robustness as a proxy for probability.

§ Applications of info-gap theory:

- Engineering design.
- Fault detection and diagnosis.
- Project management.
- Homeland security.
- Sampling, assay design.
- Statistical hypothesis testing.
- Monetary economics.
- Financial stability.
- Biological conservation.
- Medical decision making.

## 2 *DISTRIBUTIONAL UNCERTAINTY*

§ **Uncertainty, two foci:**

- **Randomness: structured uncertainty.**
- **Info-gaps:**

  **Surprise, ignorance, indeterminism.**

§ Distributional Uncertainty:

Unknown sampling distribution due to:

- Non-independence of observations.

  E.g. unknown causal pathways.

- Non-stationarity of population.

  E.g. unknown evolution over time.

- Variability of observer.

  E.g. professional/non-professional.

- Non-asymptotic data.

§ The challenge:

Design (statistical) test of hypothesis.

## § Example: Chronic Wasting Disease.

- **Antler extract from diseased deer induces disease in mice.**
- **Time to expression: uncertain pdf.**
- **Given $n$ nulls at $t$, test no-disease hypo.**

## § Example: Chronic Wasting Disease.

- **Antler extract from diseased deer induces disease in mice.**
- **Time to expression: uncertain pdf.**

## § Question:

- $n$ **innoculated mice.**
- **No PrP expression after incubation times** $t_1, \ldots, t_n$**.**
- **How confident that CWD is not present?**

§ **System model: probability of false null:**

$$P_{\rm fn}(t_1, \ldots, t_n) = \prod_{i=1}^{n} [1 - P(t_i)]$$

§ **Uncertainty model: fat tails:**

$$\mathcal{U}(h) = \left\{ p : p \in \mathcal{P}, p(t) \leq \widetilde{p}(t) + \frac{t_{\rm s} h}{t^2} \forall t \geq t_{\rm s} \right\}, \quad h \geq 0$$

- **Unbounded family.**
- **No worst case.**

§ **Robustness: Max tolerable uncertainty.**

$$\widehat{h}(n, P_{\rm fnc}) = \max \left\{ h : \left( \max_{p \in \mathcal{U}(h)} P_{\rm fn} \right) \leq P_{\rm fnc} \right\}$$

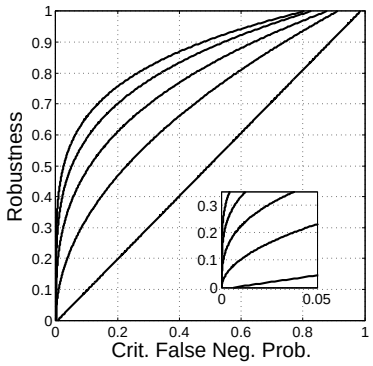



Figure 1: $\widehat{h}(n, P_{\text{fnc}})$ vs $P_{\text{fnc}}$, $n = 1$ to 5 (bottom to top).

§ **Trade-off: Robustness vs performance.**

§ **Zeroing: No rbs of estim. performance.**

§ **Cost of robustness: slope.**

§ **Choose sample size.**

## § Example: Long-term bio-monitoring.

- Given 200 ys of data, test no-change hypo.
- Data:
  - Naturalists' logs.
  - Museum collections.
- Uncertainty:
  - Museum policy changes over time.
  - Observers' habits are variable.
  - Variable observers: pros, amateurs.
  - Protocol and purpose of observation.

## § Example: Detect invasive species.

- **Uncertainties:**
  - **Transport mechanisms.**
  - **Entry mechanisms.**
  - **Habitat suitability.**
- **Decisions:**
  - **Choose traps and deployment.**
  - **Allocate resources:**
    - **Professional vs non-professional.**
    - **Detection vs irradication.**
  - **Interpret finds (e.g. nulls).**