Statistical Tests with Distributional Uncertainty: An Info-Gap Approach

Yakov Ben-Haim

Technion

Israel Institute of Technology

yakov@technion.ac.il

http://info-gap.com

[0]lectures\talks\lib\isipta2009poster01.tex 25.6.2009

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.**

Diseased Animal	Time to Expression
1	442 ± 16 (6/8)
2	> 594 (0/5)
3	463 ± 23 (2/3)
4	> 601 (0/6)

Table 1: Mice expression of deer prion protein from antler velvet of diseased animals. Angers *et al.*, 2009.

- **Given n nulls at t, test no-disease hypo.**

§ 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.

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

§ Shackle-Popper Indeterminism

- **"Prediction is always difficult, especially of the future." Scandinavian saying.**

- **Intelligence:**

  What people know, influences how they behave.

- **Discovery:**

  What will be discovered tomorrow cannot be known today.

- **Indeterminism:**

  Tomorrow's behavior cannot be modelled completely today.

- **Information-gaps, indeterminisms, sometimes cannot be modelled probabilistically.**

- **Ignorance is not probabilistic.**

§ Some info-gap models of distributional uncertainty.

- **Really severe distributional uncertainty:**
  - **Unbounded moments, fat tails, multi-modal, atoms.**
  - **Uniform-bound in the cdf:**

$$\mathcal{U}(h) = \left\{ F(y) : \; F(y) \in \mathcal{P}, \; |F(y) - \widetilde{F}_i(y)| \le h, \; \forall \; y \right\}, \quad h \ge 0 \tag{1}$$

- **Severe distributional uncertainty:**
  - **Unbounded moments, fat tails, multi-modal, no atoms.**
  - **Uniform-bound in pdf:**

$$\mathcal{U}(h) = \left\{ f(y) : \; f(y) \in \mathcal{D}, \; |f(y) - \widetilde{f}_i(y)| \le h f_i^{\star}, \; \forall \; y \right\}, \quad h \ge 0 \tag{2}$$

- **Moderate distributional uncertainty:**
  - **Bounded moments, ordinary tails, multi-modal, atoms.**
  - **Envelope-bound in cdf:**

$$\mathcal{U}(h) = \left\{ F(y) : \; F(y) \in \mathcal{P}, \; |F(y) - \widetilde{F}_i(y)| \le h\psi(y), \; \forall \; y \right\}, \quad h \ge 0 \tag{3}$$

- **Light distributional uncertainty:**
  - **Bounded moments, ordinary tails, multi-modal, no atoms.**
  - **Fractional-bound in pdf:**

$$\mathcal{U}(h) = \left\{ f(y) : \ f(y) \in \mathcal{D}, \ |f(y) - \widetilde{f}_i(y)| \le h \widetilde{f}_i(y), \ \forall \ y \right\}, \qquad h \ge 0 \tag{4}$$

- **Axioms:**
  - **Contraction:**

$$\mathcal{U}(0) = \{\widetilde{F}_i\}$$

  - **Nesting:**

$$h < h' \quad \textbf{implies} \quad \mathcal{U}_i(h) \subseteq \mathcal{U}(h')$$

  - $h =$ **unknown horizon of uncertainty.**

Statistical Tests with Distributional Uncertainty

§ Errors:

- **Type I: falsely reject H_0.**
- **Type II: falsely accept H_0.**

§ Threshold tests:

- **Test of (nominal) size $\alpha^\star$ rejects H_0 when:**

$$y \geq q_{\alpha^\star}(\widetilde{F}_0)$$

Falsely rejects H_0 with prob $\alpha^\star$.

- **Test of nominal power $1 - \beta^\star$**

correctly rejects H_0 with prob $1 - \beta^\star$:

$$1 - \beta^\star = 1 - \widetilde{F}_1[q_{\alpha^\star}(\widetilde{F}_0)]$$

- **$\alpha^\star$ small: low prob of type I error.**
- **$1 - \beta^\star$ large: low prob of type II error.**

§ Robustness of type I error:

maximum horizon of uncertainty at which the test at nominal size $\alpha^\star$ falsely rejects H_0 with probability no greater than α:

$$\widehat{h}_0(\alpha^\star, \alpha) = \max\left\{h : \left(\min_{F \in \mathcal{U}_0(h)} F[q_{\alpha^\star}(\widetilde{F}_0)]\right) \geq 1 - \alpha\right\}$$

§ Robustness of type II error:

maximum horizon of uncertainty at which the probability of falsely accepting H_0, with a test of nominal size $\alpha^\star$, is no greater than β:

$$\widehat{h}_1(\alpha^\star, \beta) = \max\left\{h : \left(\max_{F \in \mathcal{U}_1(h)} F[q_{\alpha^\star}(\widetilde{F}_0)]\right) \leq \beta\right\}$$

$\alpha^\star = 0.01$		$\alpha^\star = 0.05$	
n	$1-\beta^\star$	n	$1-\beta^\star$
5	0.1027	3	0.1784
7	0.3185	4	0.3736
9	0.5400	5	0.5390
12	0.7644	7	0.7457
31	0.9980	31	0.9997

Table 2: Size and power in the absence of distributional uncertainty.

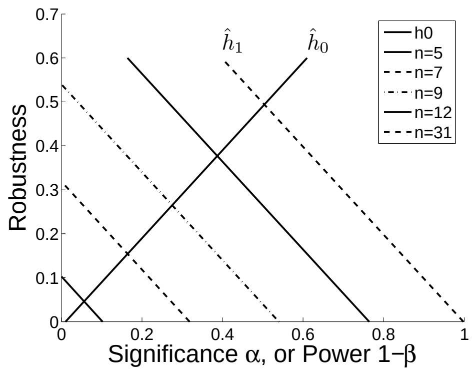

Figure 1: Robustness curves for the t test, $\widehat{h}_0(\alpha^\star, \alpha)$ for falsely rejecting H_0, and $\widehat{h}_1(\alpha^\star, \alpha)$ for falsely rejecting H_1. Nominal size is $\alpha^\star = 0.01$. $\widehat{h}_1(\alpha^\star, \alpha)$ calculated at 5 different sample sizes: $n = 5$, 7, 9, 12 and 31. $\delta = 1$.

§ Robustness curves:

- **Trade-off:**
  - **Positive slope of $\widehat{h}_0$:**

    Robustness trades-off with significance.
  - **Negative slope of $\widehat{h}_1$:**

    Robustness trades-off with power.
- **Zeroing:**

  Estimated significance or power has no robustness to distributional uncertainty.

§ Decisions and judgments.

- **Two decisions to**

  Determine the decision threshold $q_{\alpha^\star}(\widetilde{F}_0)$:

  - **Nominal test size $\alpha^\star$.**
  - **Sample size n.**

- **Two judgments:**

  - **Effective size α.**
  - **Effective power $1-\beta$**

  $\alpha =$ **prob of falsely rejecting H_0.**

  $1-\beta =$ **prob of correctly rejecting H_0.**

§ Trade-offs. positive robustness iff:

- $\alpha > \alpha^\star$.
- $1-\beta < 1-\beta^\star$.

Due to distributional uncertainty.

§ Use robustness functions $\widehat{h}_0(\alpha^\star, \alpha)$ and $\widehat{h}_1(\alpha^\star, \beta)$.

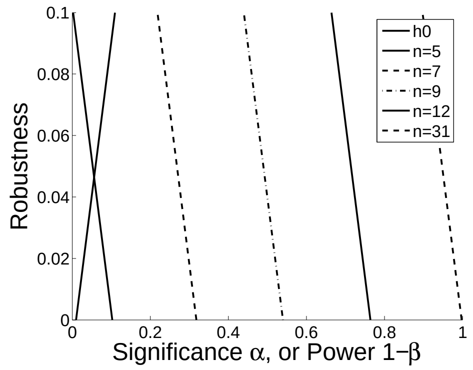

Figure 2: Expanded from fig. 1.

In fig. 2 consider nominal size $\alpha^\star = 0.01$. Consider the judgment that effective size $\alpha = 0.05$ is adequate and reliable because the robustness is $\widehat{h}_0(0.01, 0.05) = 0.04$. This judgment considers the robustness and the effective size together since they are linked through the trade-off between them. The judgment is that tails unlikely to err more than 4%, and the 5% risk of type I error is acceptable. Now apply this robustness to type II error by requiring $\widehat{h}_1(\alpha^\star, \beta) = 0.04$. From fig. 2: effective powers of 0.50, 0.72 and 0.96 for n = 9, 12 and 31. Judging that power of 0.50 is too small, we require $n > 9$. If power of 0.72 is adequate then adopt $n = 12$. Choosing $n = 31$ would result in power of 0.96.

§ 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:

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

§ Uncertainty model: fat tails:

$$\mathcal{U}(h) = \{ p : p \in \mathcal{P}, p(t) \le \widetilde{p}(t) + \frac{t_{\text{s}} h}{t^2} \forall t \ge t_{\text{s}} \} \quad (5)$$

§ Robustness function:

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

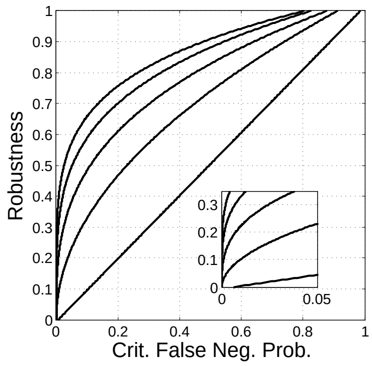

Figure 3: $\widehat{h}(n, P_{\text{fnc}})$ vs P_{fnc}, $n = 1$ to 5 (bottom to top).

Fig. 3 shows robustness curves for 5 sample sizes. Data: $t_i = $ 500, 530, 510, 520, 505 days. Bottom curve uses 1st datum; next curve uses 2 data; etc. Estimated distribution is normal; $\mu = 450$, $\sigma = 20$ days. $t_{\text{s}} = 490$.

Positive slopes express trade-off between robustness, $\widehat{h}$, and critical prob of false null, P_{fnc}. Large robustness entails large P_{fnc}. Zero robustness at estimated value of P_{fnc}.

Robustness increases substantially as n increases from 1 to 2. Marginal increase in robustness falls with increasing n. Slope increases dramatically with sample size. High slope means low cost of robustness: the robustness can be increased without significantly increasing the critical probability of false null, P_{fnc}.

§ Applications of info-gap theory:

- **Engineering design:**
  - **Off-road vehicles.**
  - **Automotive control systems.**
  - **UAV target-search strategies.**
  - **Flood control.**
- **Fault detection and diagnosis.**
- **Project management.**
- **Homeland security.**
- **Sampling, assay design.**
- **Statistical hypothesis testing.**
- **Monetary economics.**
- **Financial stability.**
- **Biological conservation.**
- **Medical decision making.**

§ Sources: http://info-gap.com