Hypothesis Testing with Interval Data: Case of Regulatory Constraints

Sa-aat Niwitpong[1], Hung T. Nguyen[2],
Vladik Kreinovich[3], and Ingo Neumann[4]

[1]Department of Applied Statistics
King Mongkut's University of Technology North Bangkok
Bangkok 10800, Thailand, snw@kmitnb.ac.th
[2]Mathematics, New Mexico State University
Las Cruces, NM 88003, USA, hunguyen@nmsu.edu
[3]Computer Science, University of Texas at El Paso
El Paso, TX 79968, USA, vladik@utep.edu
[4]Geodetic Institute, Leibniz University of Hannover
D-30167 Hannover, Germany, neumann@gih.uni-hannover.de

1. Hypothesis Testing: A General Problem

- It is often desirable to check whether a given object (or situation) satisfies a given property.
- *Examples:*
  - whether a patient has flu,
  - whether a building or a bridge is structurally stable.
- In statistics, this problem is called *hypothesis testing:*
  - we have a hypothesis – that a patient is healthy, that a building is structurally stable –
  - and we want to test this hypothesis based on the available data.
- This hypothesis H_0 is usually called a *null hypothesis*:
  - if H_0 is satisfied, no ("null") action is required,
  - if H_0 is not satisfied, action is needed: cure a patient, reinforce the building, etc.

2. Hypothesis Testing: Ideal Case of Complete Knowledge

- In the ideal case, we know the exact values of all the quantities $x_1, \ldots, x_n$ that characterize the object o.
- Since x_i are *all* the quantities characterize the object, they determine whether o satisfies the property.
- Thus, the set X of all possible values of the tuple $x = (x_1, \ldots, x_n)$ can be divided into:
  - the *acceptance region* A of all the tuples that satisfy the desired property; and
  - the *rejection region* R of all the tuples that do not satisfy the desired property.
- Thus, once we know the tuple x characterizing o, we:
  - accept the hypothesis if $x \in A$, and
  - reject the hypothesis if $x \in R$ (i.e., if $x \notin A$).

3. Hypothesis Testing: Realistic Case of Incomplete Knowledge

- In practice, we usually only have an incomplete knowledge about an object.
- Based on this partial information, we cannot always tell whether an object satisfies the given property.
- *Example:* H_0 is $x_1 + x_2 \le x_0$, and we only know x_1:
  - for some x_2 (when $x_2 \le x_0 - x_1$) we have $x_1 + x_2 \le x_0$ and thus, the hypothesis H_0 is satisfied;
  - for some x_2 (when $x_2 > x_0 - x_1$) H_0 is not satisfied.
- In such situations, the decision may be erroneous:
  - *false positive* (Type I error): the object o satisfies H_0, but we classify it as not satisfying H_0;
  - *false negative* (Type II error): the object o does not satisfy H_0, but we conclude that it does.

4. Traditional Statistical Approach to Hypothesis Testing

- We assume that we know the probability distribution of objects that satisfy the given hypothesis H_0.
- We are given the allowed probability p_0 of Type I error.
- *Idea:* we select the accept and reject regions A and R so as to minimize the probability p_{II} of Type II error.
- *Example:* in 1-D case, the distribution is usually Gaussian, with known mean a and standard deviation σ.
- Usually, situations are anomalous when the quantity (e.g., blood pressure or cholesterol level) is too high.
- In this case, we take $A = \{x_1 : x_1 \leq x_0\}$ for some x_0: $x_0 = a + 2\sigma$ for $p_0 = 5\%$, $x_0 = a + 3\sigma$ for $p_0 = 0.05\%$.
- To find p_{II}, we also need to know probability distribution for *all* objects (not necessarily satisfying H_0).

5. Limitations of the Traditional Statistical Approach to Decision Making

- *Main problem:* how to determine Type I error p_0.
- *Fact:* decreasing p_0 increases Type II probability p_{II}.
- *Example:* mass screening for breast cancer; when the result is suspicious, we apply a more complex text.
- *Consequences:* Type I error means missing cancer, Type II error means re-testing.
- If p_0 is *too low*, we apply the more complex test to too many people – so expenses are unrealistic.
- If p_0 is *too high*, we miss many cancers.
- To find desirable p_0, we must know the society's limitations and preferences.
- To determine p_0 from preferences, we must learn how to describe these preferences.

6. How to Describe Preferences: the Notion of Utility

- To get a scale, we select two alternatives: a very negative alternative A_0 and a very positive alternative A_1.
- For every $p \in [0,1]$, we consider an event $L(p)$ in which we get A_1 w/prob. p and A_0 w/prob. $1-p$.
- The larger p, the better $L(p)$: $L(0) < L(p) < L(1)$.
- $\forall$ event E, there exists a p for which E is equivalent to $L(p)$: $E \sim L(p)$; this p is called the *utility* $u(E)$ of E.
- Let an action $\mathcal{A}$ lead to alternatives $a_1, \ldots, a_m$ with utilities u_i and probabilities p_i.
- Since $a_i \sim L(u_i)$, $\mathcal{A}$ is equivalent to having $L(u_i)$ w/prob. p_i, i.e., to having A_1 w/prob. $p = p_1 \cdot u_1 + \ldots + p_n \cdot u_n$.
- Thus, the utility $u(\mathcal{A})$ of an action is equal to the expected value $E[u] = \sum p_i \cdot u_i$ of the utilities u_i.

7. Utility Is Defined Modulo Linear Transformations

- By definition $u(E)$ is the value for which E is equivalent to $L(u)$, i.e., to A_1 w/prob. u and A_0 w/prob. $1-u$.
- The numerical value of $u(E)$ depends on the choice of A_0 and A_1:
- Let $A'_0 < A_0 < A_1 < A'_1$, and let u' be utility based on A'_0 and A'_1.
- By definition, $A_0 \sim L'(u'(A_0))$ and $A_1 \sim L'(u'(A_1))$.
- Thus, E is equivalent to a composite event: $L'(u'(A_0))$ w/prob. u and $L'(u'(A_1))$ w/prob. $1-u$.
- In this composite event, we get A'_1 with probability $u \cdot u'(A_1) + (1-u) \cdot u'(A_0)$.
- Thus, in the new scale, $u' = u \cdot u'(A_1) + (1-u) \cdot u'(A_0)$, i.e., $u' = a \cdot u + b$ for $a > 0$.

8. Types of Uncertainty: Probabilistic, Interval, Fuzzy

- Uncertainty means that our estimate $\widetilde{x}$ differs from the actual (unknown) value x: $\Delta x \stackrel{\text{def}}{=} \widetilde{x} - x \neq 0$.
- *Ideal case:* we know the probabilities of different possible values of approximation error Δx.
- *Interval case:* often, we only know the upper bound Δ on the the approximation error: $|\Delta x| \le \Delta$.
- Based on $\widetilde{x}$, we conclude that the actual value of x is in the interval $\mathbf{x} \stackrel{\text{def}}{=} [\widetilde{x} - \Delta, \widetilde{x} + \Delta]$.
- In addition to Δ, experts can provide us with smaller bounds corr. to different degrees of uncertainty α.
- *Fuzzy uncertainty:* the resulting intervals can be viewed as α-cuts of a fuzzy set.

9. An Important New Case: Testing Whether an Object Satisfies Given Regulations

- *Known case:* when we know the probability distribution of objects that satisfy the null hypothesis.
- *New situation:* regulatory thresholds such as
  - "the speed limit is 75 miles",
  - "a concentration of certain chemicals in the car exhaust cannot exceed a certain level", etc.
- In general, we have:
  - the acceptance region A consisting of all the values that satisfy given regulations, and
  - the rejection region R consisting of all the values that do not satisfy the regulations.
- *Objective:* check whether the given object satisfies the corresponding regulations.

10. Testing Regulatory Thresholds: Discussion

- *Ideal case:* we know the exact value of the tested quantity x.
- *Solution:*
  - accept if $x \in A$,
  - reject if $x \notin A$.
- *1-D example:* if $A = \{x : x \le x_0\}$, we accept if $x \le x_0$ and reject if $x > x_0$.
- *In practice:* we only know the approximate value $\widetilde{x}$ of the tested quantity x.
- *Problem:* make an acceptance decision based on the estimate $\widetilde{x}$.
- *What we do:* we describe how to do it under probabilistic, interval, and fuzzy uncertainty.

11. Case of Probabilistic Uncertainty

- Let u_{++} be the utility of the situation in which x is acceptable (+), and we classify it as acceptable (+).
- Let u_{+-} be the utility of the situation in which x is acceptable (+), and we classify it as unacceptable (−).
- Similarly, we define u_{-+} and u_{--}.
- Let $p_A = \text{Prob}(\text{object w/estimate } \widetilde{x} \text{ is acceptable})$.
- We accept if $u_A > u_R$, where $u_A = p_A \cdot u_{++} + (1 - p_A) \cdot u_{-+}$ and $u_R = p_A \cdot u_{+-} + (1 - p_A) \cdot u_{--}$.
- Equivalent: accept if $p_A \geq p^{(0)} \stackrel{\text{def}}{=} \dfrac{u_{--} - u_{-+}}{u_{++} - u_{-+} - u_{+-} + u_{--}}$.
- *1-D example:* $A = \{x : x \leq x_0\}$, then p_A is the probability that $x = \widetilde{x} - \Delta x \leq x_0$, i.e., that $\Delta x \geq \widetilde{x} - x_0$.
- Here, $p_A = F(\widetilde{x} - x_0)$, so we accept if

$$\widetilde{x} \leq x_0 + F^{-1}(1 - p^{(0)}).$$

12. Example: Car Testing for Exhaust Pollution

- We test for CO, NO, and other pollutants.
- Measurement accuracy is 15–20%, so we assume Gaussian distribution with $\sigma = 0.175x_0$.
- Cost of tuning is $u_{--} = u_{+-} = -60$ (in US dollars), cost of polluting is $u_{-+} = -3000$; $u_{++} = 0$.
- In this case, $p^{(0)} = 2940/3000 \approx 0.98$, so $F^{-1}(1 - p^{(0)}) = F^{-1}(0.02) \approx -2.3\sigma \approx -0.4x_0$.
- Thus, we decide that the car passed the inspection if $\widetilde{x} \le x_0 + F^{-1}(1 - p^{(0)}) = x_0 + (-0.4x_0) = 0.6x_0$.
- Please note that here, the acceptance threshold is very low, 0.6 of the nominal value:
  - cost (Type I error) $\ll$ cost (Type II error);
  - hence, we err on the side of requiring good cars to be re-tuned.

13. Case of Interval Uncertainty

- Under interval uncertainty, we only know the interval $[\underline{u}, \overline{u}]$ of possible values of expected utility.
- We need to describe the equivalent utility $e(\underline{u}, \overline{u})$.
- *Reminder:* utility is defined modulo linear transformation $u' = a \cdot u + b$.
- It is reasonable to require that $e(\underline{u}, \overline{u})$ is invariant w.r.t. such re-scaling: $e(a \cdot \underline{u} + b, a \cdot \overline{u} + b) = a \cdot e(\underline{u}, \overline{u}) + b$.
- *Result:* $e(\underline{u}, \overline{u}) = \alpha \cdot \overline{u} + (1 - \alpha) \cdot \underline{u}$.
- When $\alpha = 1$, we base our decision on the most *optimistic* case $\overline{u}$.
- When $\alpha = 0$, we base our decision on the most *pessimistic* case $\underline{u}$.
- In general, we get an *optimism-pessimism* criterion proposed by the 2007 Nobelist L. Hurwicz.

14. Hurwicz Approach: Formula, Limitations, Alternative

- *Problem:* we know that $x \in \mathbf{x} = [\widetilde{x} - \Delta, \widetilde{x} + \Delta]$; a value x is acceptable if $x \le x_0$.
- *Simple cases:* accept if $\widetilde{x} + \Delta \le x_0$, reject if $x_0 < \widetilde{x} - \Delta$.
- *Hurwicz approach:* for $\widetilde{x} - \Delta \le x_0 < \widetilde{x} + \Delta$, accept if $\alpha \ge p^{(0)} = \dfrac{u_{--} - u_{-+}}{u_{++} - u_{-+} - u_{+-} + u_{--}}$.
- *Car inspection* (reminder): $p^{(0)} = 0.98$.
- *Conclusion:* reject (unless we are very optimistic).
- *Problem:* if $\widetilde{x} = 0.801 x_0$, then $\mathbf{x} = [0.601 x_0, 1.001 x_0]$; almost all values are acceptable, but we still reject.
- *Solution:* assume that there is a uniform distribution on $\mathbf{x}$; thus, accept if $p = \dfrac{|\mathbf{x} \cap A|}{|\mathbf{x}|} \ge p^{(0)}$.
- *If $\widetilde{x} = 0.801 x_0$:* $p = 0.9975 > 0.98$, so we accept.

15. Case of Fuzzy Uncertainty

- *We have:* a fuzzy set X.
- *We want:* to estimate $p_A = P(A \cap X \mid X) = \dfrac{P(A \cap X)}{P(X)}$.
- *Idea:* we can gauge $\mu_A(x)$ as the proportion of experts who believe that x satisfies the property A.
- So, $\mu_A(x)$ is the probability that, according to a randomly selected expert, x satisfies A.
- Every expert has a set of values that, according to this expert's belief, satisfy the property A.
- We consider the experts to be equally valuable, so these sets are equally probable.
- Thus, we have, in effect, a probability distribution on the class of all possible sets – a *random set.*

16. Case of Fuzzy Uncertainty (cont-d)

- *Reminder:* $\mu_A(x)$ can be interpreted as the probability that a given element x belongs to the random set.
- We know the probability $\mu_X(x) = P_X(x \in S)$ that a given element x belongs to the random set.
- Thus, $P(X) = \int p(x) \cdot P_X(x \in S)\, dx = \int p(x) \cdot \mu_X(x)\, dx$.
- We assume: all x are equally probable: $p(x) = c$.
- Similarly, $P(A \cap X) = c \cdot \int \mu_{A \cap X}(x)\, dx$, so

$$p_A = \frac{P(A \cap X)}{P(X)} = \frac{c \cdot \int \mu_{A \cap X}(x)\, dx}{c \cdot \int \mu_X(x)\, dx} = \frac{\int \mu_{A \cap X}(x)\, dx}{\int \mu_A(x)\, dx}.$$

- *Recommendation:* we accept the null hypothesis if $p_A \ge p^{(0)} = \dfrac{u_{--} - u_{-+}}{u_{++} - u_{-+} - u_{+-} + u_{--}}$, else reject.

17. Acknowledgments

This work was supported in part

- by the research project KU 1250/4-1 funded by the German Research Foundation (DFG),
- by NSF grant HRD-0734825,
- by Texas Department of Transportation Research Project No. 0-5453,
- by the Japan Advanced Institute of Science and Technology (JAIST) International Grant 2006-08, and
- by the Max Planck Institut für Mathematik.

The authors are thankful to the anonymous referees for valuable suggestions.