Impugning Alleged Randomness

Yuri Gurevich

Guanajuato, Nov 13, 2014

impugn (ɪmˈpjuːn)

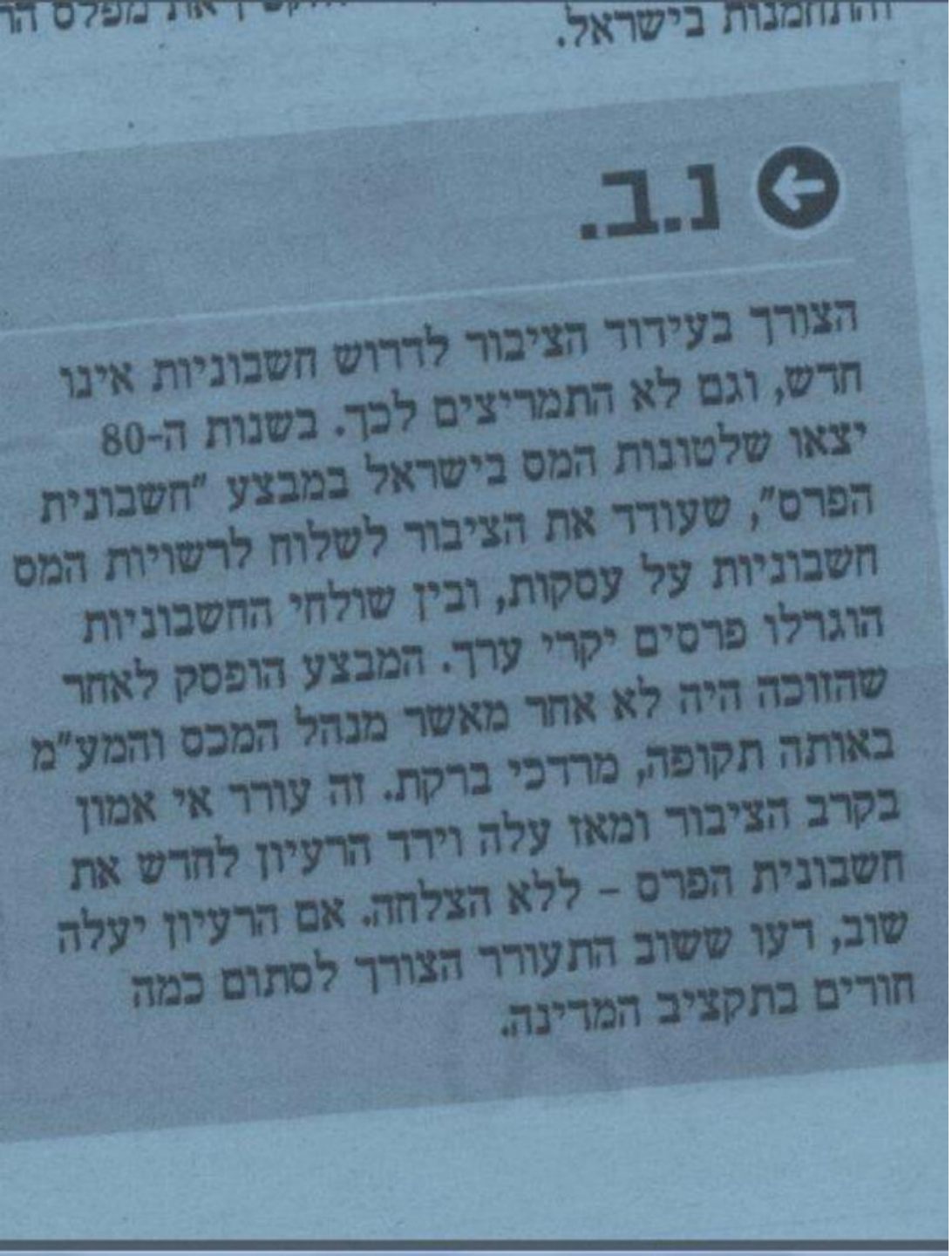
— ***vb***

(*tr*) to challenge or attack as false; assail; criticize

from Old French *impugner,*
from Latin *impugnāre* to fight against, attack,
from im- + *pugnāre* to fight

New York Times, 1985

TRENTON, July 22 – The New Jersey Supreme Court today caught up with the “man with the golden arm," Nicholas Caputo, the Essex County Clerk and a Democrat who has conducted drawings for decades that have given Democrats the top ballot line in the county 40 times out of 41 times. The court suggested – but did not order – changes in the way Mr. Caputo conducts the drawings to stem further loss of public confidence in the integrity of the electoral process."



נ.ב.

הצורך בעידוד הציבור לדרוש חשבוניות אינו
חדש, וגם לא התמריצים לכך. בשנות ה-80
יצאו שלטונות המס בישראל במבצע "חשבונית
הפרס", שעודד את הציבור לשלוח לרשויות המס
חשבוניות על עסקות, ובין שולחי החשבוניות
הוגרלו פרסים יקרי ערך. המבצע הופסק לאחר
שהווכה היה לא אחר מאשר מנהל המכס והמע"מ
באותה תקופה, מרדכי ברקת. זה עורר אי אמון
בקרב הציבור ומאז עלה וירד הרעיון לחדש את
חשבונית הפרס – ללא הצלחה. אם הרעיון יעלה
שוב, רעו ששוב התעורר הצורך לסתום כמה
חורים בתקציב המדינה.

The Marker of Dec. 16, 2011

www.news1.co.il/Archive/006-D-500-00.html:

מ-1980 ועד פרישתו ב-30 ביוני 1991, כיהן כמנהל אגף המכס והמע"מ.

Lottery

- John organized a state lottery. Every citizen was given one ticket, and his wife won the main prize.
- Is this a mere coincidence or was the lottery rigged?
- What is known about John? Not much. He is devoted to his family and close friends.

Cournot's principle

- How is probability theory related to the real world? Via the Cournot's principle:
- "A predicted event of sufficiently small probability does not happen".
- Known already to Jakob Bernoulli (1713 posthumous Art of Conjecturing).
  Concurred: Émile Borel, Ronald Fisher, Jacques Hadamard, Andrei Kolmogorov, Paul Lévy, ...

How small is sufficiently small?

- This is not a simple question. The answer depends on the application area and may evolve with time.
- *Simplifying Proviso*: There is an agreed and current probability threshold for the application area in question. Events of probability below the threshold are *negligible.*

Terminology and notation

- A *probabilistic scenario* (T, P, E) is given by
  - a trial T with a number of potential outcomes,
  - a probability distribution P, the *null hypothesis*, and
  - a *focal event* E (that will typically be negligible).
- Let's consider such a scenario.

Cournot's principle expounded

If the focal event E is specified
before the execution of trial T then

it is practically certain that

the focal event E does not happen.

Narrow Bridge Principle

If the focal event E is specified
(possibly after the trial T was executed but)
without any information about
the actual outcome of T

then it is practically certain that the focal event E does not happen.

Bridge Principle

If the focal event E is specified independently of the trial T execution then it is practically certain that the focal event E does not happen.

- But can a specification be a posteriori and yet independent?

ALGORITHMIC INFORMATION THEORY

Kolmogorov complexity

- $K(s)$ = length(shortest program for s)
  Here s is a binary string.
- What is the programming language?
  In a sense this is not too important because of the Invariance Theorem:
  $\forall P, Q \exists c \left(K_P(s) \leq K_Q(s) + c \right)$.

How is K(s) relevant?

- As $K(s)$ becomes smaller,
  s becomes less random,
  more objective and
  more independent of anything.
- Now think of s as
  the description of the focal event E.

Critique

- $K(s)$ is not computable.
- The lack of symmetry.
- Hard to reflect real-world scenarios.

The Kolmogorov centennial conference on Kolmogorov complexity in Dagstuhl at 2003.

TOWARD PRACTICAL SPECIFICATION COMPLEXITY

The idea

- Model the scenario in terms most natural to it. The background matters.
  - Some lottery organizers have been known to cheat.
  - Some clerks are too partisan.
- A succinct specification of a focal event in terms of such a natural model may be viewed to be independent of the actual outcome.

Logic models

- Logic models seem appropriate to the kind of scenarios we saw
- Other scenarios may use very different languages and modes.
  - Time series may be appropriate for analyzing stock market.

One-sorted relational structures

- Base set, relations, constants
- Example: directed graphs
- Example: trees
- Vocabulary

Multi-sorted relational structures

- Sorts
- Types of relations, variables, constants
- Example.
  - Sorts Person, Ticket
  - Relation Owns of type Person $\times$Ticket
  - Constant John of type Person
- By default relational structures will be multi-sorted

Logic

- Somewhat arbitrarily, we choose our logic to be first-order logic.
- The logic of textbooks. The most common logic.

Definitional complexity

- Let M be a relational structure and S one of the sorts of M.
- A set $X \subseteq S$ is definable in M if there is a first-order formula $\varphi(x)$ with
  $$X = \{x : \varphi(x)\}.$$
- Here φ is a definition of X.
- The definitional complexity of X in M is the length of a shortest definition of X in M.

Impugning randomness: the method

Given a probabilistic trial, a null hypothesis and a suspicious actual outcome, do:

1. Analyze the trial and establish what background information is relevant.
2. Model the trial and the relevant background info.
3. Propose a focal event E of low definitional complexity, negligible under the null hypothesis, that contains the actual outcome.

By the bridge principle, E is not supposed to happen during the execution. This is a reason to reject the null hypothesis.

Lottery

CloseRelative(John,w) or
CloseFriend(John,w)

In other words, the winner w is a close relative or close friend of John.

Man with golden arm

$$(\exists_{\leq 1} c)\ \text{nonDem}(o, c)$$

There is at most one election (out of 41) where the first candidate c is not a democrat.

THANK YOU

A BAYESIAN TAKE

BY ALEX ZOLOTOVITSKI

- A priori probability $P(F)$ of fraud is 0.01 (the percentage of incarcerated in the US). How relevant is this probability?
- $P(B) = 1 - P(F) = 0.99$. (B for "benign".)
- $P(W|F) = 1$. (W for the actual win.)
- $P(W|B) = 10^{-7}$. (She has 1 ticket out of 10^7.)
- $P(F|W) = \frac{P(W|F)P(F)}{P(W|F)P(F)+P(W|B)P(B)} \approx 0.99999$, a posteriori probability of F.
- $P(B|W) = \frac{P(W|B)P(B)}{P(W|F)P(F)+P(W|B)P(B)} \approx 10^{-5}$. a posteriori probability of B.

- Consider the costs CFP and CFN of a false positive and a false negative, and suppose that jailing one innocent is as bad as letting free 1000 fraudsters.
  Another judgment.
- If CFN = 1 then CFP = 1000.
- Then Cost (toJail) =

$$CFP \cdot P(B|W) \approx 1000 \cdot 10^{-5} = 0.01$$

- Cost(letFree) =

$$CFN \cdot P(F|W) \approx 0.99999$$

- So Cost(toJail) < Cost(letFree)
  Hence the decision: Guilty, go to Jail.
- We can't prove the guilt of the lottery organizer; we can only impugn the alleged probability distribution.