Impugning Alleged Randomness

Yuri Gurevich

Tallinn, April 28, 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

MOTIVATING EXAMPLES

Baltimore Sun, Aug. 13, 1890

A gentleman of this city brings out some curious ... characteristics of lotteries ... “There are 100,000 tickets ... and 844 prizes ... Every month the numbers that draw prizes start at the lowest numeral and go up by regular gradation to the highest ... If the drawing begins at ... say 50, there will be several numbers up to 100, there will be numbers between 100 and 200, 200 and 300, ... up to 1,000; then between 1,000 and 2,000, up to 10,000; then between 10,000 and 20,000, and so on up to 100,000. Large prizes have frequently been drawn in the ninety thousands. Now, this may occur one or two month together or at intervals, but on the theory of probabilities it is impossible to so draw numbers every month.”

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."

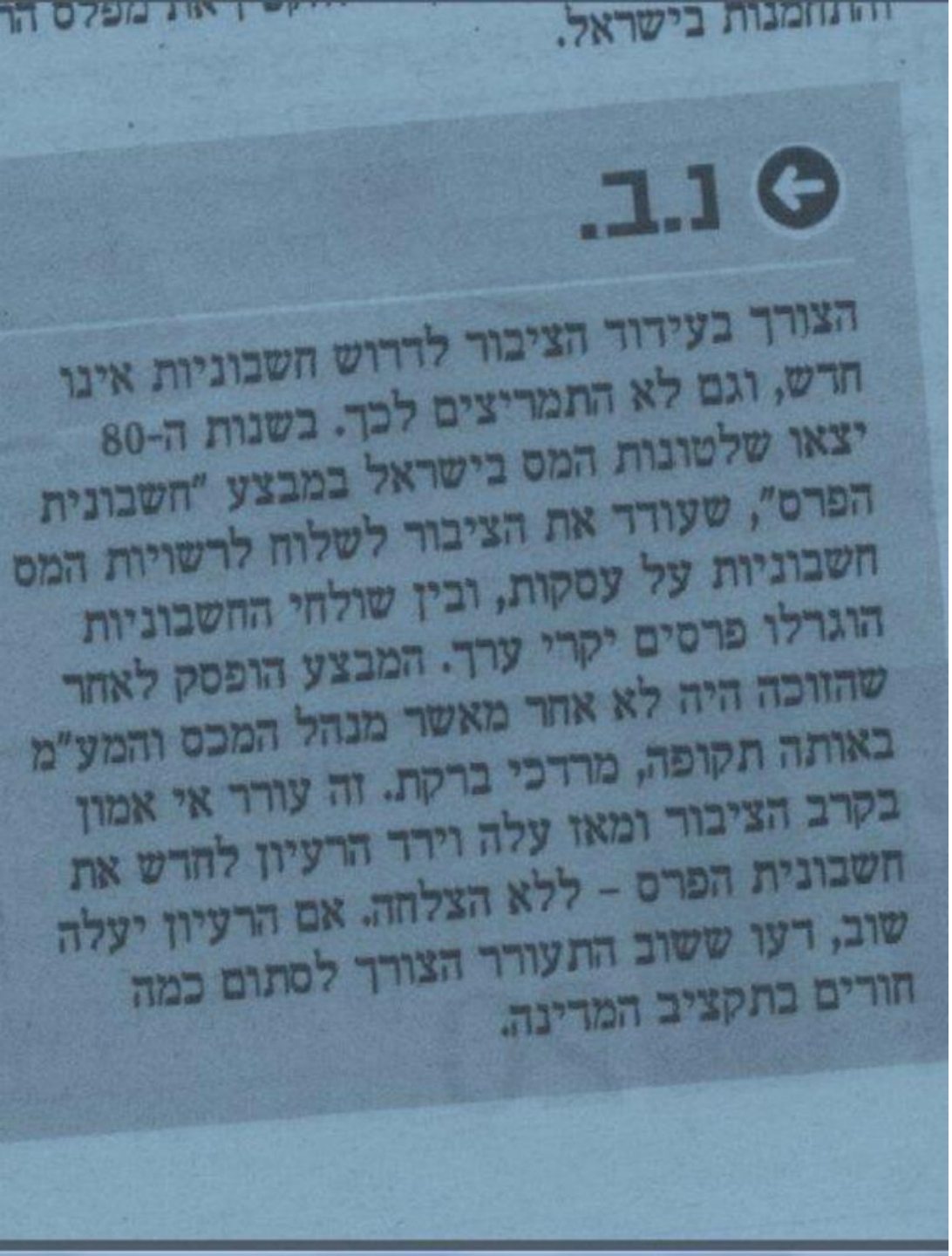
Forbes, July 21, 2011

The August issue of Harper's magazine contains a fascinating story about a woman named Joan R. Ginther, known in the press as the "luckiest woman in the world."

To earn that appellation, Ginther won the lottery four times. That's right, four times. And she didn't win no measly $20 and $30 payouts either—she hit multiple million dollar payouts each time.

First, she won $5.4 million; then a decade later, she won $2 million; then two years later $3 million; and finally, in the spring of 2008, she hit a $10 million jackpot.

The odds of this? One in eighteen septillion.



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.

ANCHORING PROBABILITY THEORY

Cournot's principle

- "A predicted event with very small probability will not happen".
- Known already to Jakob Bernoulli [1713 posthumous Art of Conjecturing]
- Antoine Cournot: The only connection between probability theory and physical world
- Concurred: Émile Borel, Ronald Fisher, Jacques Hadamard, Andrei Kolmogorov, Paul Lévy, ...

How small is sufficiently small?

- Depends on the application area.
  Evolves with time.

- *Proviso*: Given a probabilistic trial, we will always assume the existence of an agreed and current probability threshold for the application area of the trial.

Definitions and notation (slightly simplified)

- A trial T with outcomes and events
- An executed trial, the actual outcome, events that happened (or occurred)
- A probabilistic trial (T, P), the null hypothesis, negligible events, an executed probabilistic trial
- A probabilistic scenario (T, P, E) with focal event E

Remark. This is a slight simplification. Instead of a single probability distribution P one should consider (and the paper does) a family of perturbations of P.

Cournot's principle expounded

Consider a probabilistic scenario with a single negligible focal event.

If the focal event has been specified before the execution,

then it is practically certain that the focal event will not happen upon the execution.

The standard use of Cournot's principle.

If the focal event happens during the execution of the trial,
then reject the null hypothesis.

An aside on statistics

Cournot Principle is the foundation of the statistical theory of Ronald Fisher.

Discussion on Cournot Principle

1. How can we claim that something of positive probability will not happen? We certainly cannot prove the claim. Here's one relevant quote from Richard Feynman:
   - (Mathematics is not a science from our point of view, in the sense that it is not a natural science. The test of its validity is not experiment.) We must, incidentally, make it clear from the beginning that if a thing is not a science, it is not necessarily bad. For example, love is not a science. So, if something is said not to be a science, it does not mean that there is something wrong with it; it just means that it is not a science.
2. Could we make the absolutist position ("will not happen") relativistic?

THE BRIDGE PRINCIPLE

Narrow Bridge Principle

If the focal event is specified
without any information about
the actual outcome

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

Bridge Principle

- Let's strengthen Cournot's principle.
- Consider a probabilistic scenario with a negligible focal event.
- *Bridge Principle*. If the focal event is specified independently of the trial execution then it is practically certain that the focal event does not happen upon the execution.

What specifications are independent?

- Can a specification be a posteriori and yet independent?
- Can one write an independent specification after learning the actual outcome?

ALGORITHMIC INFORMATION THEORY

Algorithmic complexity

- Kolmogorov (or algorithmic information) complexity $K(s)$ of a string s is the length of a shortest program that produces s.
  - Given $|s|$, the smaller $K(s)$, the less random s is.
- What is the programming language?
- Invariance: $\forall P, Q \exists c \left(K_P(s) \leq K_Q(s) + c\right)$.
- Conditional Kolmogorov complexity $K(s|t)$.

Critique

- $K(s)$ is not computable.
- Hard to reflect real-world scenarios.
- The lack of symmetry.
- Is there a better way to deal with real-world scenarios?
- The Kolmogorov centennial conference on AIT, Dagstuhl, 2003.

TOWARD PRACTICAL SPECIFICATION COMPLEXITY

The idea

- Model the scenario in terms of the scenario.
- A succinct specification of a focal event in terms of such a natural model may be viewed to be independent of the actual outcome.

Naturality of a model

Q: How to judge the naturality of a model?

A: A natural model could have been written by an expert who knows all about the trial but about the actual outcome.

Q: Find such an expert and ask them to write the model?

A: If feasible then yes, one should do that.

Relevant background

The expert is given relevant background info including what a priori may go wrong.

- Some lottery organizers have been known to cheat.
- Some clerks are too partisan.

LOGIC MODELS

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.
- We use also quantifiers $\forall_n, \exists_n$ which do not increase expressivity but which make specifications shorter and more natural.

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 ALLEGED RANDOMNESS

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.

EXAMPLE DESCRIPTIONS

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.