THE LOGIC IN COMPUTER SCIENCE COLUMN

BY

YURI GUREVICH

Microsoft Research
One Microsoft Way, Redmond WA 98052, USA
gurevich@microsoft.com

IMPUGNING RANDOMNESS, CONVINCINGLY

Yuri Gurevich   Grant Olney Passmore*

Abstract

John organized a state lottery and his wife won the main prize. You may feel that the event of her winning wasn't particularly random, but how would you argue that in a fair court of law? Traditional probability theory does not even have the notion of random events. Algorithmic information theory does, but it is not applicable to real-world scenarios like the lottery one. We attempt to rectify that.

1   Introduction

To motivate our study, we begin with four examples. In each of the four cases, a probabilistic trial produces a suspicious outcome, and the question arises whether the outcome is the result of pure chance. The first case is a thought experiment inspired by a remark of Leonid Levin in article [32]. The second and third cases are also thought experiments, arguably more realistic. The fourth case is real.

---

*Clare Hall, University of Cambridge and LFCS, University of Edinburgh, UK. grant.passmore@cl.cam.ac.uk

1.1 A Lottery Case

John and Donna live in Georgia, a state of about 10,000,000 inhabitants. John is Donna's husband and the president of the Georgia State Lottery. Anybody may enter into the lottery by buying as many tickets as (s)he wishes. Every ticket is uniquely numbered. A winner is chosen at random by the selection of a ticket number. This year Donna happens to win. Over 10,000,000 tickets were purchased in total, spread among about 4,000,000 people. Donna purchased three. John bought none.

Soon after this win is announced, the local media begins to echo claims of corruption against John. How could it be that of all of the about 4,000,000 participants, the president's wife won? Surely something must be amiss. John, faced with allegations of unfairness, argues as follows:

> Someone had to win the lottery. The process of choosing the winner was fair. Almost every ticket owner (the only exception being a handful of people who bought many tickets) had a small chance of winning. If a stranger to me who also bought a small number of tickets had won, no one would be crying foul. But, such a stranger would have roughly the same small probability of winning as Donna did. Given that someone had to win, nothing strange has happened. In particular, there are no grounds to claim the lottery was rigged.

Would you believe him?

1.2 A Jury Case

Thomas seemed to be a common criminal but there was something uncommon about his case. At least the prosecutor thought so. As prospective jurors were questioned, she realized that some of them were unusually informed. She investigated. It turned out that seven out of 50 prospective jurors belonged to a Facebook group of about 100 people that discussed Thomas's case. This was the only Facebook group that discussed the case.

Prospective jurors had been chosen at random from a population of about one million adults available for the purpose. Can it be mere chance that so many prospective jurors belong to the one and relatively small Facebook group that have discussed the case?

1.3 A Stalking Case

Alice and Bob are a couple living in New York City. They don't have a real kitchen and usually dine out. Chris is Alice's unstable ex-boyfriend, and Alice believes

that Chris is stalking her. Too often she has seen him at restaurants. Alice has wanted to obtain a restraining order but Bob has argued that they didn't have enough evidence to convince the authorities. After all, Chris and Alice used to live together and may naturally frequent the same restaurants. So Bob suggested an experiment: "There are at least 100 reasonable restaurants within walking distance from our place. For the next ten nights, let's pick a restaurant at random except that it should be a new restaurant each time". They performed the proposed experiment. In 5 out of the 10 cases, Chris showed up. Is this evidence sufficient for Alice to obtain a restraining order?

1.4   The Case of the Man with the Golden Arm

The story appeared in New York Times on July 23, 1985, on page B1. We learned of it from the book [18].

> 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 felt that something was wrong.

> 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."

Caputo wasn't punished. A question arises whether the circumstantial evidence was sufficient to justify punishing him.

1.5   Overview

The four cases above have something in common. In each case, there is a strong suspicion that a presumably random event is not random at all. But how can one justify the suspicion?

The purpose of this paper is to build a practical framework in which the desired chance-elimination arguments can be formalized and defended. We start, in §2, with a classical principle, often called Cournot's principle, according to which it is a practical certainty that an event with very small probability will not happen. We expound Cournot's principle. In particular, we make explicit that the event of interest is supposed to be specified in advance. Then we generalize Cournot's

principle to a more liberal principle, called the bridge principle, that requires only that the event of interest be specified independently from the execution of the probabilistic trial in question. At the end of §2, we address the question how an after-the-fact specification can be independent. The inspiration comes from algorithmic information theory, and the intuitive idea is this: some specifications are so succinct that they could have been naturally written ahead of time (and maybe have been written in similar cases in the past).

§3 is auxiliary. First we recall some basic notions of algorithmic information theory, in particular the notion of Kolmogorov complexity (or information complexity) of events. In algorithmic information theory, events are represented by binary strings. The Kolmogorov complexity of an event is the length of a shortest program for a fixed universal Turing machine that outputs the string presentation of the event. This approach does not work for our purposes. In each probabilistic case in question, we need specifications formulated in terms pertinent to the case and abstracted from irrelevant information. To this end we use logic, and the rest of the section is devoted to logic. We recall and illustrate logic structures and the notion of logical definability. Then we introduce and discuss the notion of the description complexity of events.

In §4, we explain how we intend to impugn randomness. In §5 we illustrate our approach on the cases described above. There are many discussions with our old friend Quisani throughout the article. The final discussion is in §7.

Acknowledgments

2 Bridging Probabilities and the Physical World

We explicate and broaden the well known principle according to which it is a practical certainty that an event with very small probability will not happen.

2.1 Cournot's Principle

Probability theory is applied widely and successfully. But what makes this mathematical theory relevant to the physical world? In The Art of Conjecturing, published posthumously in 1713 [2], Jakob Bernoulli related mathematical probabilities to practical (or "moral") certainty:

Something is morally certain if its probability is so close to certainty that the shortfall is imperceptible. Something is morally impossible if its probability is no more than the amount by which moral certainty falls short of complete certainty. Because it is only rarely possible to obtain full certainty, necessity and custom demand that what is merely morally certain be taken as certain. It would therefore be useful if fixed limits were set for moral certainty by the authority of the magistracy — if it were determined, that is to say, whether 99/100 certainty is sufficient or 999/1000 is required.

In other words, it is a practical certainty that an event with very small probability will not happen. Antoine Cournot seems to be the first to suggest, in the book [15], that this principle is the only way of connecting mathematical probabilities to the world. Accordingly the principle is often ascribed to Cournot.

Cournot's Principle It is a practical certainty that an event with very small probability will not happen.

The principle was supported by many heavyweights of probability theory and statistics, in particular Émile Borel [3], Ronald A. Fisher [21], Andrei N. Kolmogorov [29] and Paul Lévy [33]. Borel called Cournot's principle The Single Law of Chance.

More information on Cournot's principle is found in the Shafer and Vovk book [42] and Shafer's lecture [41] that is available online.

2.2 How Small is Sufficiently Small?

Quisani: How small is sufficiently small? I presume that there is some cut-off point, a threshold value, the least upper bound for the sufficiently small values.

Authors: Yes, that's the idea. Let us quote Émile Borel in this connection. In his 1943 book [3] for the non-scientist he wrote the following.

When we stated The Single Law of Chance, "events whose probability is sufficiently small never occur," we did not conceal the lack of precision of the statement. There are cases where no doubt is possible; such is that of the complete works of Goethe being reproduced by a typist who does not know German and is typing at random. Between this somewhat extreme case and ones in which the probabilities are very small but nevertheless such that the occurrence of the corresponding event is not incredible, there are many intermediate cases. We shall attempt to determine as precisely as possible which values of probability must be regarded as negligible under certain circumstances.

Note "under certain circumstances". Different application areas may use different threshold values. You may want to reject the null hypothesis "by preponderance of evidence" or "beyond a reasonable doubt". If the courts of law were to use probabilistic thresholds (tailored to specialized circumstances), these distinct judicial criteria would give rise to distinct threshold values. Different criteria and different threshold values may be used in testing scientific hypotheses.

Q: I suppose much experience is needed to just propose a reasonable threshold for a fixed application area. And new developments, new technologies may require that the accepted value be revised. I am thinking of the use of DNA evidence in courts. A number of convictions have been overturned. If the courts used thresholds, some of them should have been adjusted down. I can imagine also the necessity to adjust a threshold value up. Think of cases when people have not been convicted but later confessed to crimes.

A: The issue of appropriate threshold values has been much discussed, especially in connection with statistical hypothesis testing. In fact there are many discussions, in various applications domains, e.g. clinical trials [17, 25, 36, 39], psychology [14, 24]. Statistical hypothesis testing is often misunderstood and abused [1, 14, 24, 38]. In the rest of the paper, we will avoid the issue of appropriate threshold values.

Proviso 1. *Given a probabilistic trial, we will always assume the existence of an agreed and probability threshold for the application domain of the trial.*

2.3   Cournot's Principle Expounded

Our formulation of Cournot's principle is rather common. It is also aphoristic. As stated, Cournot's principle does not hold water: events of very small probability happen all the time in the physical world. Several aspects of the principle are implicit. One aspect, for example, is that the probabilistic experiment in question is performed only once. (This particular aspect is made explicit in [29].) In this subsection we explicate Cournot's principle. Later we will broaden it and call it the bridge principle to emphasize that it bridges between probability theory and the physical world.

As in probability theory, a trial T is a real or imaginary experiment with a well defined set Ω_T of possible outcomes and with events of T as subsets of Ω_T. In general (when Ω_T is uncountable) there may be some subsets of Ω_T that are not events. The case of most interest to us is when Ω_T is finite; in that case every subset of Ω_T is an event.

We introduce a few nonstandard terms that are useful for our purposes. An *executed trial* is a trial together with a particular execution of the trial; the execution results in a particular outcome called the *actual outcome* of the executed trial. An

event *E happens* or *occurs* at the executed trial if and only if *E* contains the actual outcome.

A *probabilistic trial* $(T, \mathcal{F})$ is a trial *T* together with a hypothesis, called the *null hypothesis*, that the probability distribution that governs the trial *T* belongs to $\mathcal{F}$. The probability distributions of $\mathcal{F}$ are the *innate probability distributions* of the probabilistic trial $(T, \mathcal{F})$. (We are not going to define what it means for a trial *T* to be governed by a probability distribution P; the connection to the physical world is given by the expounded Cournot principle below.) An *executed probabilistic trial* is a probabilistic trial $(T, \mathcal{F})$ together with a particular execution of *T* (that produces the actual outcome of the executed probabilistic trial).

An event *E* of a probabilistic trial $(T, \mathcal{F})$ is *negligible* if, for every innate probability distribution P, the probability P(*E*) is less than the current probability threshold in the application area of the trial.

Example 1. View the lottery of §1.1 as an executed probabilistic trial with possible outcomes of the form "*o* wins the lottery" where *o* ranges over the lottery ticket owners. The null hypothesis says that the trial is governed by the probability distribution where the probability of outcome "*o* wins the lottery" is proportional to the number of lottery tickets that *o* owns. The actual outcome is "Donna wins the lottery". Since Donna bought three tickets, the event "the winner bought three tickets" occurs during the execution.

A *probabilistic scenario* $(T, \mathcal{F}, E)$ is a probabilistic trial $(T, \mathcal{F})$ together with an event $E \subseteq \Omega_T$ called the *focal event* of the probabilistic scenario. An *executed probabilistic scenario* is a probabilistic scenario $(T, \mathcal{F}, E)$ together with a particular execution of *T* (that produces the actual outcome of the executed probabilistic scenario).

Cournot's Principle Expounded   Consider a probabilistic scenario with a negligible focal event. If the focal event is specified before the execution of the trial then it is practically certain that the focal event will not happen upon the execution.

Q: What is the point in fixing an event and execution?

A: If too many small-probability events are specified then it may become likely that at least one of them happens even if we restrict attention to one execution of the trial. Similarly any event of positive probability becomes likely to happen if the trial is executed too many times.

Q: Given an informal description of a trial, it may be not obvious what the possible outcomes are. Consider the lottery case. The way the story is told in §1.1, every outcome is associated with the winning person. This is natural. But it is also natural, maybe even more natural, to associate outcomes with the winning tickets.

A: You are right. An informal description of a trial may be somewhat vague about precisely what possible outcomes are. But the definition of a probabilistic trial requires that the set of possible outcomes be indicated explicitly. In the lottery case, there are indeed these two natural ways to view possible outcomes. It does not really matter which way to go. We picked the first way because it is a tiny bit more convenient for our purposes.

2.4   Cournot's Principle and Statistical Hypothesis Testing

Q: You started with the problem of connecting probabilities to the physical world. But do you know the true probabilities of real world events? I think not. All you have is a mathematical model. It surely is at best an approximation of reality. For example, people speak about tossing a fair coin but no real coin is perfectly fair and no tossing is perfect. More importantly, a real world trial may be rigged, so that your mathematical model may be far from reality. I would not be surprised if some magicians can produce any desired sequence of heads and tails by repeatedly tossing a coin.

A: Turn Cournot's principle upside down. Consider an executed probabilistic scenario with a negligible focal event specified before executing the trial. If the focal event occurs during the execution of the trial then reject the null hypothesis.

Q: Is this related to statistical hypothesis testing?

A: This is, as far as we understand, the basic idea of Ronald A. Fisher's method of statistical hypothesis testing [21, 22, 23]. The term "null hypothesis" is borrowed from Fisher. In Fisher's approach, the focal event is specified by means of the *p*-value of some statistics.

Q: Hmm, I don't know anything about *p*-values. What is your best reference on Fisher's approach?

A: The Cox and Hinkley book [16].

2.5   The Bridge Principle

Now we are ready to broaden Cournot's principle. We use the definitions of §2.3.

The Bridge Principle Consider a probabilistic scenario with a negligible focal event. If the focal event is specified *independently* of the execution of the trial then it is practically certain that the focal event does not happen upon the execution.

We will use the bridge principle as a ground for rejection of (or at least as a significant argument against) the null hypothesis.

Q: How can an after-the-fact specification be independent? I think that I understand the intent of prior specifications. They are predictions. If I give Grant a deck

of cards and, without examining the cards, he draws the king of spades, there is no surprise. But if he announces in advance that he is going to draw the king of spades and then indeed he does that, then there is a surprise. I'd think that Grant is a magician. But after-the-fact "predictions" do not make sense to me.

A: Suppose that Grant did not announce a card in advance. Instead Yuri announces the card, after the fact and apparently without any communication with Grant. Is there an element of surprise?

Q: Yes, I suppose there is. I would suspect that Grant is a magician or that there was some communication between you two. I guess the point is not that the focal event is specified in advance but that it is specified — possibly after the fact — without any knowledge about the outcome of the trial, not even partial knowledge.

A: Let us give a name to the principle that you propose:

The Narrow Bridge Principle Consider a probabilistic scenario with a negligible focal event. If the focal event is specified without any information about the actual outcome of the trial then it is practically certain that the focal event does not happen upon the execution.

Unfortunately the narrow bridge principle is too narrow for our purposes. To illustrate the broader principle, consider a trial that consists of tossing a fair coin 41 times. Would the outcome

$$\text{HTTTTHTHHHTHTHTHHTTHTTHHHTTTHHTHHTTTTTHHHH} \tag{1}$$

be surprising?

Q: I do not think so.

A: And what about the outcome where all tosses but the last one came up heads?

Q: Yes, it would be surprising. I would suspect cheating.

A: But why? Both outcomes have exactly the same probability, 2^{-41}.

Q: I begin to see your point. The second outcome is special. It is surprising even though it has not been predicted. I guess it is surprising by its very nature.

A: Yes. But what makes the second outcome special?

Q: The particularly simple specification?

A: That is exactly it. The particularly simple specification makes the outcome surprising and independent from the execution of the trial.

Q: But how do you measure the simplicity of specifications? There are 41 characters, including blanks, in the phrase "all tosses but the last one came up heads", and 41 binary symbols in 1. Since the Latin alphabet (with the blank symbols) is

richer than the binary alphabet $\{H, T\}$, one can reasonably argue that the specification 1 is simpler, much simpler.

A: This is a good question. The whole next section is devoted to it.

3 Random Events and their Specification Complexity

3.1 Algorithmic Information Theory

The introductory examples illustrate the possibility that some presumably random events may be not random at all.

Q: What does it mean that an event is random?

A: Classical probability theory does not address the question but algorithmic information theory (AIT) does. The basic ideas of AIT were discovered in the 1960s independently by Ray Solomonoff [44], Andrei N. Kolmogorov [30] and Gregory J. Chaitin [8]. The history of these discoveries is described in §1.13 of the book [35] by Li and Vitányi that we use as our main reference on the subject. Chaitin's book [9] on AIT is available online.

A key notion of AIT is the Kolmogorov (or information) complexity of strings. Intuitively, the Kolmogorov complexity $C(s)$ of a string s is the length of a shortest program that outputs s. The larger $C(s)$ is, comparative to the length of s, the more random s is.

Q: Programs in what programming language?

A: Programs for a universal Turing machine. AIT was influenced by the computation theory of the time. Traditionally, in AIT, one restricts attention to Turing machines with the binary alphabet, and the universality of a Turing machine U means that U faithfully simulates any other Turing machine T on any input x given T (in one form or another) and given exactly that same input x. View a universal Turing machine U as a programming language, so that programs are binary strings. The Kolmogorov complexity $C_U(s)$ of a binary string s is the length of a shortest U program that outputs s.

Q: But this depends on the choice of a universal Turing machine U.

A: It does. But, by the Invariance Theorem [35, §2.1], for any two universal Turing machines U_1 and U_2, there is a constant k such that $C_U(s) \leq C_V(s) + k$ for all binary strings s. In that sense, the dependence on the choice of universal Turing machine U is limited.

There is also a conditional version $C_U(s|t)$ of Kolmogorov complexity, that is the complexity of string s given a string t.

Q: I wonder how would one use Kolmogorov complexity to show that the suspicious outcomes of the introductory examples are not random.

A: Unfortunately Kolmogorov complexity does not seem to work well for our purposes. Whatever universal Turing machine U is fixed, the function $C_U(s)$ is not computable [35, Theorem 2.3.2]. And the machine U does not know anything about the scenarios. Consider the lottery scenario for example. Intuitively the event of Donna winning the lottery should have smaller description complexity than the event of some stranger to John winning the lottery. But this is most probably not the case, precisely because the machine U does not know anything about the scenario.

Q: Maybe one can use the conditional version $C_U(s|t_0)$ of Kolmogorov complexity where t_0 is a particular string that describes the given scenario. I suspect that the function $f(s) = C_U(s|t_0)$ is still uncomputable. But maybe one can approximate it.

A: Maybe. But it seems to us that there are simpler and more natural ways to deal with scenarios like those in our introductory examples.

3.2 Description Complexity

We assume that the reader is familiar with the basics of first-order logic though we recall some notions. A (one-sorted) relational structure A consists of

- a nonempty set, the *base set* of structure A, whose elements are called the elements of A;

- several relations over the base set including equality, the *basic relations* of structure A; each basic relation has its own arity (the number of arguments);

- several distinguished elements of A known as *constants*.

The ever present equality is a *logic relation*, in contrast to other basic relations. The names of the basic non-logic relations and constants form the *vocabulary* of A. Equality is typically omitted when a structure is described. For example, a directed graph is typically described as a relational structure with one binary relation and no constants.

A multi-sorted relational structure is defined similarly except that the base set is split into several nonempty subsets called *sorts*. Each argument position i of every basic relation R is assigned a particular sort S_i; the *type* of R is the direct product $S_1 \times \cdots \times S_r$ where r is the arity of R. Equality, the only logic relation, is an exception. Its type can be described as $\bigcup_S (S \times S)$ where S ranges over the sorts. The vocabulary of a multi-sorted relational structure contains the names of

sorts, relations and constants. Besides, the vocabulary indicates the types of basic relations, individual variables and constants.

Example 2. Here is a structure related to the lottery scenario. It has two sorts. One sort, called Person, consists of people, namely all lottery ticket owners as well as John, the lottery organizer. The other sort, called Ticket, consists of all the lottery tickets that have been sold. The structure has a binary relation Owns of type Person × Ticket, with the obvious interpretation. It also has a constant John that denotes the lottery organizer.

If A is a relational structure, S is a sort of A and $X \subseteq S$, we say that X is *definable* in A if there is a first order formula $\varphi(x)$ with a single free variable x of type S such that X is the set of elements a of sort S satisfying the proposition $\varphi(a)$ in A, that is if

$$X = \{a : A \models \varphi(a)\}.$$

The formula $\varphi(x)$ is a *definition* of X in A. The description complexity of X in A is the length of the shortest definition of X in A.

Example 3. The event "the winner owns just one ticket" consists of the outcomes "p wins the lottery" where p ranges over the people owning exactly one ticket. The event is thus definable by formula

$$\exists t_1(\text{Owns}(p, t_1) \land (\forall t_2(\text{Owns}(p, t_2) \rightarrow t_1 = t_2))).$$

in the structure of Example 2.

Q: How do you measure the length of a formula?

A: View the names of relations, variables and constants as single symbols, and count the number of symbols in the formula. Recall that the vocabulary specifies the type of every variable and every constant.

Q: As far as the lottery case is concerned, the structure of Example 2 is poor. For example, it does not distinguish between people that own the same amount of tickets. In particular, it does not distinguish between Donna and anybody who owns exactly three tickets. You can extend it with a constant for Donna. If you want that the structure reflects a broader suspicion that John may cheat, you can add a constant for every person such that there is a reasonable suspicion that John will make him a winner. Much depends of course on what is known about John. For example, you can add a constant for every close relative and every close friend of John.

A: Alteratively we may introduce binary relations CloseRelative(p, q) and CloseFriend(p, q).

Example 4. Extend the structure of Example 2 with binary relations CloseRelative(p, q) and CloseFriend(p, q) of type Person × Person, with the obvious interpretations: q is a close relative of p, and q is a close friend of p respectively; in either case q is distinct from p.

3.3 Alternatives

Q: There is something simplistic about Example 4. Both relations seem to play equal roles. In reality, one of them may be more important. For example, John may be more willing to make a close relative, rather than a close friend, to win. People often put different weights on different relations. For a recent example see [40]. You should do the same.

A: You are right of course but, for the time being, we keep things simple.

Q: And why do you use relational first-order logic? There are many logics in the literature.

A: In this first paper on the issue, it is beneficial for us to use relational first-order logic as our specification logic. It the best known and most popular logic, and it works reasonably well. As (and if) the subject develops, it may be discovered that the best specification logic for one application domain may not be the best for another. At this point, our experience is very limited.

Q: First-order logic isn't the best logic for all purposes.

A: It is not. And there are two distinct issues here. One issue is expressivity. If, for example, you need recursion, first-order logic may be not for you. It lacks recursion. The other issue is succinctness. It is possible to increase the succinctness of relational first-order specifications without increasing the expressive power of the logic. For example, one may want to use function symbols. One very modest extension of relational first-order logic which is nevertheless useful in making specifications shorter is to introduce quantifiers $\dot{\exists}x\varphi$ (note the dot over $\exists$) saying that there exists x different from all free variables of the formula φ under quantification. If $y_1, \ldots, y_k$ are the free variables of φ then $\dot{\exists}x\varphi$ is equivalent to

$$\exists x(x \neq y_1 \wedge \cdots \wedge x \neq y_k \wedge \varphi)$$

but it is shorter. It would be natural of course to introduce the $\dot{\exists}$ quantifier together with its dual $\dot{\forall}$ quantifier.

Q: Instead of logic, one can use computation models, especially restricted computation models, e.g. finite state automata, for specification.

A: Yes, you are right. Note though that, for every common computation model, there is a logic with equivalent expressivity. For example, in the case of finite automata over strings, it is existential second-order logic [5].

4 Impugning Randomness

Now we are ready to explain our method of impugning the null hypothesis in executed probabilistic scenarios with suspicious outcomes. Given

- a trial such that some of its outcomes arouse suspicion and

- a null hypothesis about the probability distribution that governs the trial,

one has several tasks to do.

1: Background Information   Analyze the probabilistic trial and establish what background information is relevant.

2: Logic Model   Model the trial and relevant background information as a logic structure.

3: Focal Event Propose a focal event that is

- negligible under the null hypothesis and

- has a short description in the logic model.

By the bridge principle, the focal event is not supposed to happen, under the null hypothesis, during the execution of the trial. If the focal event contains the actual outcome of the trial, then the focal event has happened. This gives us a reason to reject the null hypothesis.

What background information is relevant?

Relevant background information reflects various ways that suspicious outcomes occur. In this connection historical data is important. In the lottery case, for example, it is relevant that some lottery organizers have been known to cheat.

What does the Model Builder Know about the Actual Outcome?

The less the model builder knows about the actual outcome the better. Ideally the model builder has no information about the actual outcome, so that we can use the narrow bridge principle. We may not have a model builder with no information about the actual outcome; it may even happen that the actual outcome has been advertised so widely that everybody knows it. In the absence of blissfully unaware model builder, we should try to put ourselves into his/her shoes.

The Desired Logic Model

One may be lucky to find an existing logic model that has been successfully used in similar scenarios. If not, construct the most natural and frugal model you can.

Q: "Natural" is a positive word. Surely it is beneficial that the desired model is natural. But why should the model be frugal?

A: If the model is too rich (like in the case of classical algorithmic information theory), too many events have short specifications. Imagine for example that, in the lottery case, the model allows you to specify shortly various people that have nothing to do with the lottery organizer.

Q: But how do you know that those people have nothing to do with the lottery organizer? Maybe one of them is a secret lover of the lottery organizer.

A: Indeed, the background information deemed relevant may be deficient. But, at the model building stage, we want to reflect only the background information deemed relevant.

The Desired Focal Event

The desired focal event contains the suspicious outcomes of the trial.

5 Examples

We return to the four cases of §1.

5.1 The Case of Lottery

Trial For an informal description of the trial see §1.1. Recall that John is the lottery organizer, and Donna is his wife. As mentioned in Example 1, we view the lottery as a trial with potential outcomes of the form "o wins the lottery" where o ranges over the lottery ticket owners.

Null Hypothesis There is only one innate probability distribution P, and the probability $P(x)$ of any person x to win is proportional to the number of lottery tickets that x owns.

Background Information We assume that the following is known about John, the lottery organizer. He is a family man, with a few close friends that he has known for a long time. He bought no lottery tickets.

Actual Outcome "Donna wins the lottery".

Logic Model Our model is a simplification of the structure of Example 4. We don't need the sort Ticket introduced originally in Example 2 to illustrate the notion of multi-sorted model. And we do not need the full extent of relations CloseRelative and CloseFriend, only the sections of them related to John. Our model is one-sorted. The one sort, called Person, consists of people, namely all lottery ticket owners as well as John, the lottery organizer. The structure has one constant and two unary relations. The constant is John; it denotes the lottery organizer. The two unary relations are CloseRelativeOfJohn(p) and CloseFriendOfJohn(p), both of type Person. The interpretations of the two relations are obvious: p is a close relative of John, and p is a close friend of respectively; in both cases p is distinct from John.

Focal Event: The winner is a close relative or friend of John, the lottery organizer, in other words, the winner belongs to the set

$$\{x : \text{ CloseRelativeOfJohn}(x) \lor \text{CloseFriendOfJohn}(x)\}.$$

5.2 The Case of Jury Selection

Trial For an informal description of the trial see §1.2. The trial in question selects a pool of 50 prospective jurors from about a 1,000,000 people available for the purpose.

Null Hypothesis There is only one innate probability distribution, and the one innate probability distribution is uniform so that all possible pools of 50 prospective jurors are equally probable.

Background Information There is a unique Facebook group of about 100 people that discusses the criminal case.

Actual Outcome A pool with seven prospective jurors from the Facebook group.

Logic Model The model has two sorts and one relation.

- Sort Pool consists of all possible pools of 50 prospective jurors.

- Sort Member consists of the members of the Facebook group that discussed Thomas's case.

- The relation In(m, p) of type Member $\times$ Pool holds if and only if member m of the Facebook group belongs to pool p.

Focal Event

$$\{p :\ \exists m_1 \exists m_2 (m_1 \neq m_2\ \wedge\ \text{In}(m_1, p)\ \wedge\ \text{In}(m_2, p))\}$$

Q: If the null hypothesis is impugned then some rules have been violated. Who is the guilty party? In the lottery case, it was clear more or less that John was the guilty party. In this case, it is not obvious who the guilty party is.

A: We do not pretend to know the guilty party. Our only goal is to impugn randomness. There may be more than one guilty party as far as we know. Their actions may or may not have been coordinated.

Q: As far as I see, plenty of randomness might have remained. You did not impugn all randomness.

A: You are right. Let us express our goal more precisely: it is to impugn the null hypothesis, no more no less.

5.3 A Stalking Case

Trial   An informal description of the case is given in §1.3. The 10 nights of the trial may be represented by numbers $1, \ldots, 10$. The outcomes of the trial may be represented by functions f from $\{1, \ldots, 10\}$ to $\{0, 1\}$ where the meaning of $f(n) = 1$ (resp. $f(n) = 0$) is that Alice and Bob met (resp. did not met) Chris at the restaurant on night n.

Null Hypothesis   Intentionally, the null hypothesis says that Chris does not stalk Alice. Formally, the null hypothesis says that a probability distribution P on the outcomes is innate if and only if it satisfies the following two requirements for every outcome f.

1. Events $f(n_1) = 1$ and $f(n_2) = 1$ are independent for any nights $n_1 \neq n_2$.

2. $P(f(n) = 1) \leq 1/100$ for every night n.

Requirement 2 says that $P(f(n) = 1)$ is less than (rather than equal to) $1/100$ rather than $P(f(n) = 1) = 1/100$ because, at night n, Chris may not to be present at all in any of the 100 restaurants at the time when Alice and Bob dine. If he is in one of the 100 restaurants when Alice and Bob dine then $P(f(n) = 1) = 1/100$.

Background Information   Chris is suspected of stalking Alice in restaurants.

Actual Outcome   Five times out of ten times Alice and Bob meet Chris at the chosen restaurant.

Logic Model   There are three sorts of elements and one relation.

- Sort Night consists of numbers $1, \ldots, 10$.

- Sort Outcome consists of all possible functions from Night to $\{0, 1\}$.

- Relation $R(f, n)$ holds if and only if, on night n, Alice and Bob meet Chris at the chosen restaurant.

Focal Event $\{f : \exists n_1 \exists n_2 (n_1 \neq n_2 \wedge R(f, n_1) \wedge R(f, n_2))\}$.

Q: The focal event is that there are two distinct nights when Chris dines at the restaurant where Alice and Bob dine. What if the current probability threshold is lower than you presume, and the focal event turns out to be non-negligible?

A: In this particular case, it is natural to consider the focal event that there are three distinct nights when Chris dines at the restaurant where Alice and Bob dine.

Q: It is rather expensive to say that there are k distinct elements; the description complexity is $O(k^2)$. Now I see why you mentioned those dotted existential quantifiers $\dot\exists$ in §3.3.

5.4   The case of Nicholas Caputo

Trial   An informal description of the case is given in §1.4. The 41 elections may be represented by numbers $1, \ldots, 41$. The possible outcomes of the trial can be seen as functions f from $\{1, \ldots, 41\}$ to $\{D,N\}$ where the letters D, N indicate whether the top ballot line went to a Democrat or not.

Null Hypothesis   Intentionally the null hypothesis is that the drawings were fair. Formally, the null hypothesis says that there is a unique innate probability distribution P on the outcomes, and that P satisfies the following two requirements for every outcome f.

1. Events $f(e_1) = D$ and $f(e_2) = D$ are independent for any elections $e_1 \neq e_2$.

2. For every election e, $P(f(e) = D) = r_e$ where r_e is the fraction of Democrats on the ballot.

We assume that every $r_e \geq 0.4$.

Background Information   The county clerk, who conducted the drawings, was a Democrat.

Actual Outcome   40 times out of 41 times the top ballot line went to Democrats.

Logic Model   The model has three sorts, two constants and one relation.

- Sort Election consists of numbers $1, \ldots, 41$ representing the 41 elections.

- Sort Party consists of two elements. The elements of Party will be called parties.

- Sort Outcome consists of all f from Election to Party.

- The constants D and N of type Party denote distinct elements of type Party.

- The relation $R(f, e, p)$ of type Outcome $\times$ Election $\times$ Party holds if and only if, according to outcome f, the top ballot line went to party p at elections e.

Focal Event:
$$\{f : \; \forall i \forall j ((R(f, i, N) \wedge R(f, j, N)) \rightarrow i = j)\}.$$

6   Related Work

Our paper touches upon diverse areas of science. We restrict attention to a few key issues: Cournot's principle, algebraic information complexity, and social network analysis.

6.1   Cournot's Principle

The idea that specified events of small probability do not happen seems to be fundamental to our human experience. And it has been much discussed, applied and misapplied. We don't — and couldn't — survey here the ocean of related literature. In §2 we gave already quite a number of references in support of Cournot's principle. On the topic of misapplication of Cournot's principle, let us now turn to the work of William Dembski. Dembski is an intelligent design theorist who has written at least two books, that are influential in creationist circles, on applications of "The Law of Small Probability" to proving intelligent design [18, 19].

We single out Dembski because it is the only approach that we know which is, at least on the surface, similar to ours. Both approaches generalize Cournot's principle and speak of independent specifications. And both approaches use the

information complexity of an event as a basis to argue that it was implicitly specified. We discovered Dembski's books rather late, when this paper was in an advanced stage, and our first impression, mostly from the introductory part of book [18], was that he ate our lunch so to speak. But then we realized how different the two approaches really were. And then we found good mathematical examinations of the fundamental flaws of Dembski's work: [46] and [4].

Our approach is much more narrow. In each of our scenarios, there is a particular trial T with well defined set Ω_T of possible outcomes, a fixed family $\mathcal{F}$ of probability distributions — the innate probability distributions — on Ω_T, and a particular event — the focal event — of sufficiently small probability with respect to every innate probability distribution. The null conjecture is that the trial is governed by one of the innate probability distributions. Here events are subsets of Ω_T, the trial is supposed to be executed only once, and the focal event is supposed to be specified independently from the actual outcome. By impugning randomness we mean impugning the null hypothesis.

Dembski's introductory examples look similar. In fact we borrowed one of his examples, about "the man with a golden arm". But Dembski applies his theory to vastly broader scenarios where an event may be e.g. the emergence of life. And he wants to impugn any chance whatsoever. That seems hopeless to us.

Consider the emergence of life case for example. What would the probabilistic trial be in that case? If one takes the creationist point of view then there is no probabilistic trial. Let's take the mainstream scientific point of view, the one that Dembski intends to impugn. It is not clear at all what the trial is, when it starts and when it is finished, what the possible outcomes are, and what probability distributions need to be rejected.

The most liberal part of our approach is the definition of independent specification. But even in that aspect, our approach is super narrow comparative to Dembski's.

There are other issues with Dembski's work; see [46, 4].

6.2 Algorithmic Information Theory

The idea of basing the intrinsic randomness of an event upon its description in a fixed language is fundamental to algorithmic information theory (in short AIT) [9, 35] originated by Ray Solomonoff [44], Andrei N. Kolmogorov [30] and Gregory J. Chaitin [8].

In §3.1, we sketched the basic ideas of the theory. In the classical AIT, the theoretical power is gained by basing the information complexity measure on universal Turing machines. This becomes an impediment to practical applications; the classical information complexity of (the string representation of) events is not

computable. For practical applications, it is thus natural to look at restricted variants of AIT which "impoverish" the event description language even though the classical theorems of AIT may no longer hold.

The influential Lempel-Ziv compression theory of strings [51, 52] can be viewed as such a restriction of AIT. However Lempel and Ziv developed their theory without any direct connection with AIT. One recent and even more restrictive theory [7] was inspired by AIT: "we develop a version of Algorithmic Information Theory (AIT) by replacing Turing machines with finite transducers".

One useful application of AIT to real-world phenomena has been through the Universal Similarity Metric and its uses in genetics and bioinformatics [31, 34, 20, 26], plagiarism detection [11] and even analysis of music [13]. In [12], the authors combine a restricted variant of Kolmogorov complexity with results obtained from Google searches to derive a metric for the similarity of the meaning of words and phrases. In doing so, they are able to automatically distinguish between colors and numbers, perform rudimentary automatic English to Spanish translation, and even distinguish works of art by properties of the artists. In such lines of research, practitioners often replace the Kolmogorov complexity measure with measures based on string-compression algorithms [47, 6, 49] more efficient than the original Lempel-Ziv algorithms.

In cognitive science, the simplicity theory of Chater, Vitányi, Dessalles and Schmidhuber offers an explanation as to why human beings tend to find certain events "interesting" [10]. The explanation correlates the interest of an event with its simplicity (i.e., the lowness of its Kolmogorov complexity) .

Our logic-based definition of description complexity in §3.2 fits this mold of restricted algorithmic information theories. We note, however, that the logic approach is rather general and can handle the classical information complexity and its restricted versions and even its more powerful (e.g. hyper-arithmetical) versions.

6.3   Social Network Analysis

The idea of modeling real-world scenarios using relational structures dates back at least to the 1950s [37]. The primary scientific developers of this idea were for many years sociologists and social anthropologists working in the field of social network analysis. As a field of mathematical sociology, social network analysis has put forth a network-theory oriented view of social relationships and used it to quantitatively analyze social phenomena.

Even in the days before massive social network data was available, social network analysts obtained fascinating results. For example, in a 1973 paper "The Strength of Weak Ties", Mark Granovetter put forth the idea that most jobs in the United States are found through "weak ties", that is acquaintances the job seeker

knows only slightly. Granovetter obtained his relational data by interviewing only dozens of people, yet his conclusions held up experimentally and are widely influential today in sociology. With the advent of large-scale digitized repositories of relational social network data such as Facebook (according to a recent estimate, more than 40% of the US population have Facebook accounts [48]), the applicability of social network analysis techniques grew tremendously. The relational algebra of social network analysis tends to be simple. Typically, analysis is done with rudimentary graph theory: members of a population (called actors) are nodes in a graph and the relationships of interest between actors are modeled as edges. Multiple binary relations are combined into composite relations so that core social network analysis calculations are done over a single graph's adjacency matrix [45].

In the case that our models are graphs, there is much machinery of social network analysis which could be of use to us. For instance, social network analysts have developed robust and scalable methods for determining the central nodes of interest of social networks, based upon things like weighted connectivity. We can imagine this being useful for impugning randomness. For instance, if one does not know which members of the population should be distinguished and named by constant symbols, the very structure of a social network may force certain nodes upfront. There are many other techniques from social network analysis (and available high-performance software) which have the potential to be useful for our goals.

7   Final Discussion

Q: I have been thinking about algorithmic information theory and its applications, and I also did some reading, e.g. [28, 43]. In general your logic approach appeals to me but I have some reservations.

A: Let's start with the positive part. What do you like about the logic approach?

Q: The situation at hand is described directly and rather naturally. I also like that some outcomes and events are not definable at all. Consider for example the lottery-related model in §5.1. Unless John, the lottery organizer, has a single close relative or a single close friend, no particular outcome is definable in the model. And the model does not distinguish at all between any two persons outside the circle that contains John, his close relatives and his close friends. This simplicity may be naive but it is certainly appealing.

A: Indistinguishability is important. It is rather surprising in a sense that, in the application of probability theory, so often one is able to compute or at least approximate probabilities. Think about it. The number of possible outcomes of a

trial may be rather large and may even be infinite. And these are just outcomes; most events contain multiple outcomes. A probabilistic measure on the event space is a function from the events to real numbers between 0 and 1 that satisfies some slight conditions.

Q: Most probability measures are useless I guess. Which of them are useful?

A: Those few that allow us feasible — though possibly approximate — computations of the probabilities of interesting events. Typically useful measures heavily exploit the symmetries inherent in the trial and the independence of various parts of the trial.

Q: I think I see the connection to indistinguishability. But let me go to my reservations. It is basically about the annoying freedom in fixing the probability threshold, in choosing the appropriate logic, in figuring out what background information is relevant and what should the focal event be, in constructing the logical model, and in deciding whether a proposed logical specification of the focal event is short enough.

A: The "annoying freedom" is inherent in the impugning-randomness problem.

Q: Kolmogorov complexity is objective, due to the Invariance Theorem mentioned in §3.1.

A: It is objective only to a point. Recall that the Invariance Theorem involves an unspecified additive constant. So Kolmogorov complexity also suffers from the nagging question "is it short enough". Besides, one may be interested in the length of a shortest specification of a given string in first-order arithmetic or Zermelo-Fraenkel set theory for example. The resulting specification complexity measures are rather objective. They are undecidable of course, but so is Kolmogorov complexity.

Q: So how do you intend to deal with the annoying freedom?

A: We believe that the annoying-freedom problem cannot be solved by theorists. It can be solved, better and better, by experimentation, trial and error, accumulation of historical records, standardization, etc.

Q: Allow me one other question before we finish. You mentioned in §2.4 that, in Fisher's approach, the focal event is specified by means of the *p*-value of some statistics. "In statistical significance testing", says Wikipedia [50], "the *p*-value is the probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true". Note the closure under the at-least-as-extreme values. If a focal event is not specified by means of a *p*-value, is there any kind of closure that the focal event should satisfy?

A: Yes, in our examples, the focal event contained not only the actually observed outcome but also other suspicious outcomes. In fact, the focal-event approach is rather flexible. Consider the lottery scenario for example. The actual outcome —

that Donna won the lottery — may be judged to be the most suspicious and, from that point of view, the most extreme, so that there are no other outcomes at least as extreme. But the focal event contains additional suspicious outcomes.

References

[1] Carmen Batanero. Controversies around the role of statistical tests in experimental research. *Mathematical Thinking and Learning*, 2((1-2)): 75–98, 2000.

[2] Jakob Bernoulli. *Ars Conjectandi*. Thurneysen Brothers, Basel, 1713.

[3] Émile Borel. *Les Probabilités et la Vie*. Presses Universitaires de France, Paris, 1943. English: *Probabilities and Life*, Dover, 1962.

[4] James Bradley. Why Dembski's Design Inference Doesn't Work. The Biologos Foundation, 2009.

[5] J. Richard Büchi. Weak second-order arithmetic and finite automata. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 6: 66–92, 1960.

[6] bzip2. http://bzip.org/

[7] Cristian S. Calude, Kai Salomaa and Tania K. Roblot. Finite State Complexity. *Theoretical Computer Science*, to appear.

[8] Gregory J. Chaitin. On the length of programs for computing finite binary sequences. *Journal of ACM*, 13, 547–569, 1966.

[9] Gregory J. Chaitin. *Algorithmic Information Theory*. Cambridge University Press, New York, NY, USA, 1987.

[10] Nick Chater. The search for simplicity: a fundamental cognitive principle? *The Quarterly Journal of Experimental Psychology Section A*, 52(2): 273–302, 1999.

[11] Xin Chen, B. Francia, Ming Li, B. McKinnon, and A. Seker. Shared information and program plagiarism detection. *IEEE Transactions on Information Theory*, 50(7): 1545 – 1551, 2004.

[12] Rudi Cilibrasi and Paul Vitányi. The Google Similarity Distance. *IEEE Transactions on Knowledge and Data Engineering*, 19(3): 370–383, March, 2007.

[13] Rudi Cilibrasi, Paul Vitányi, and Ronald De Wolf. Algorithmic clustering of music based on string compression. *Comput. Music J.*, 28: 49–67, December 2004.

[14] Jacob Cohen. The Earth Is Round (p < .05). *American Psychologist* 49(12): 997–1003, 1994.

[15] Antoine Augustin Cournot. *Exposition de la Théorie des Chances et des Probabilités*. Hachette, Paris, 1843.

[16] D. R. Cox and D. V Hinkley. *Theoretical Statistics*. Chapman and Hall London, 1974.

[17] B.R. Davis and R.J. Hardy, Upper bounds for type I and type II error rates in conditional power calculations *Communications in Statistics*, 19(10): 3571–3584, 1990.

[18] William A. Dembski. *The Design Inference: Eliminating Chance Through Small Probabilities*. Cambridge University Press, 1998.

[19] William A. Dembski. *No Free Lunch: Why Specified Complexity Cannot Be Purchased Without Intelligence*. Rowman and Littlefield, 2006.

[20] Paolo Ferragina, Raffaele Giancarlo, Valentina Greco, Giovanni Manzini, and Gabriel Valiente. Compression-based classification of biological sequences and structures via the universal similarity metric: experimental assessment. *BMC Bioinformatics*, 8(1): 252, 2007.

[21] Ronald A. Fisher. *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh, 1925.

[22] Ronald A. Fisher. *Statistical Methods and Scientific Inference*. Hafner, 1926.

[23] Ronald A. Fisher. The arrangement of field experiments. *Journal of the Ministry of Agriculture of Great Britain*, 33: 503–513, 1956.

[24] R. Chris Fraley. The Statistical Significance Testing Controversy: A Critical Analysis. `http://www.uic.edu/classes/psych/psych548/fraley/`.

[25] L. Friedman, C. Furberg and D. DeMets. *Fundamentals of Clinical Trials*, 3rd edition. Mosby, 1996.

[26] D. Gilbert, F. Rossello, G. Valiente, and M. Veeramalai. Alignment-free comparison of tops strings. *London Algorithmics and Stringology 2006*, 8: 177–197, 2007.

[27] Mark S. Granovetter. The strength of weak ties. *The American Journal of Sociology*, 78(6): 1360–1380, 1973.

[28] Walter Kirchherr, Ming Li and Paul Vitányi. The miraculous universal distribution. *The Mathematical Intelligencer*, 19(4): 7–15, 1997.

[29] Andrei N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, Berlin, 1933.

[30] Andrei N. Kolmogorov. Three approaches to the quantitative definition of information. *Problems Inform. Transmission*, 1(1): 1–7, 1965.

[31] N. Krasnogor and D. A. Pelta. Measuring the similarity of protein structures by means of the universal similarity metric. *Bioinformatics*, 20: 1015–1021, May 2004.

[32] Leonid A. Levin. Randomness Conservation Inequalities; Information and Independence in Mathematical Theories. Information and Control 61: 15–37, 1984.

[33] Paul Lévy. *Calcul de probabilités*. Gauthier-Villars, Paris, 1925

[34] Ming Li, Jonathan H. Badger, Xin Chen, Sam Kwong, Paul Kearney, and Haoyong Zhang. An information-based sequence distance and its application to whole mitochondrial genome phylogeny. *Bioinformatics*, 17(2): 149–154, 2001.

[35] Ming Li and Paul Vitányi. *An Introduction to Kolmogorov Complexity and its Applications*. 2nd edition, Springer, 1997.

[36] C.L. Meinert. *Clinical Trials Design, Conduct, and Analysis*. Oxford University Press, 1986.

[37] Robert King Merton. *Social Theory and Social Structure*. Free Press, (Glencoe, Il.), 1957.

[38] Denton E. Morrison and Ramon E. Henkel, editors. *The Significance Test Controversy*. Aldine. Oxford, England, 1970.

[39] L.A. Moye. P-value interpretation and alpha allocation in clinical trials. *Annals of Epidemiology*, 8(6): 351–357, 1998.

[40] Matthew Richardson and Pedro Domingos. Markov logic networks. *Machine Learning*, 62(1-2): 107–136, 2006.

[41] Glenn Shafer. Why did Cournot's Principle disappear? Lecture at École des Hautes Études en Sciences Sociales, May 2006. `http://www.glennshafer.com/assets/downloads/disappear.pdf`.

[42] Glenn Shafer and Vladimir Vovk. Probability and Finance: It's Only a Game. John Wiley and Sons, 2001.

[43] Alexander Shen. Algorithmic Information Theory and Foundations of Probability. *arXiv* 0906.4411, 2009.

[44] Ray Solomonoff. A formal theory of inductive inference, part 1 and part 2. *Information and Control*, 7(1): 1–22, and 7(2): 224–254, 1964.

[45] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge Univ Pr, 1994.

[46] Richard Wein. Not a Free Lunch But a Box of Chocolates: A critique of William Dembski's book "No Free Lunch". `http://www.talkorigins.org/design/faqs/nfl`.

[47] Terry Welch. A Technique for High-Performance Data Compression. *IEEE Computer*, 17(6): 8-Ű19, 1984.

[48] Roy Wells. 41.6% the US Population Has a Facebook Account. Posted August 8, 2010 at `http://socialmediatoday.com/index.php?q=roywells1/158020/416-us-population-has-facebook-account`.

[49] Wikipedia. Prediction by partial matching. `http://en.wikipedia.org/wiki/PPM_compression_algorithm`.

[50] Wikipedia. P-value. `http://en.wikipedia.org/wiki/P-value`.

[51] Jacob Ziv and Abraham Lempel. A universal algorithm for sequential data compression. *IEEE Transactions on Information Theory*, 23(3): 337–343, 1977.

[52] Jacob Ziv and Abraham Lempel. Compression of individual sequences via variable-rate coding. *IEEE Transactions on Information Theory*, 24(5): 530–536, 1978.