# Towards Algebraic Descriptive Randomness

Yuri Gurevich[1] Grant Olney Passmore[2]

[1]FSE, Microsoft Research

[2]MRG&ACG, LFCS, University of Edinburgh

GTP Workshop, 2010

# Outline of Topics

There will be a review of algorithmic randomness later today, so we will keep our usage of Kolmogorov-Solomonoff-Chaitin to be largely intuitive and non-technical. Really, the whole talk will be intuitive and non-technical save for a simple model and the application of some basic machinery from algebraic topology at the end (time permitting).

# Goal

- An *easily applicable* theory of randomness useful for justifying accusations of unfairness in certain games of chance.
- To be *easily applicable*, we require:
  - not only computable, but also *feasibly* computable
  - models should be well-motivated and easy to understand even from the perspective of a (mathematical) layperson: We're especially interested in a theory which could be easily applied in the court of law
- To be *feasibly* computable, we require:
  - *symmetry*, uniformity, independence
- Result: A modest, very simple special case of Kolmogorov complexity which is feasibly computable and whose models are highly symmetric

# Enter Cournot's Principle

Cournot's Principle

> *"an event of small or zero probability singled out in advance will not happen."*

– "From Cournot's Principle to Market Efficiency" (Shafer)

# Example I: Statement

Let us imagine...

- John and Donna both live in Columbus, Georgia, a town of 500,000 inhabitants.
- John is Donna's husband.
- John is the president of the Columbus Town Lottery.
- Every inhabitant of Columbus is automatically entered into the lottery, and a winner is chosen "at random."
- This year, Donna happens to win.

# Example I: A town up in arms

Soon after this win is announced, the local media begins to echo claims of corruption against John. How could it be that of all 500,000 inhabitants, the president's wife won? Surely something must be amiss.

# Example I: John's response

John, however, argues as follows:

> *"Someone had to win the lottery. There are 500,000 people in our town. Thus, assuming the process of choosing the winner was fair, as I claim it was, every member of the town had a* $\frac{1}{500{,}000}$ *chance of winning. If a stranger to me had won, no one would be crying foul. But a stranger to me has the exact same odds of winning as Donna does:* $\frac{1}{500{,}000}$. *Thus, it is no less probable for Donna to win than for anyone else, and given that someone had to win, nothing strange has happened. In particular, there are no grounds to claim the lottery was unfair."*

Something doesn't feel right. But why?

# Example I: Is John right?

What does classical probability theory have to say about this argument? Would John's argument stand up in court?

- Well, if the probability distribution is indeed defined as John claims, e.g., with $\mathcal{I}$ as the set of inhabitants of Columbus we have $\forall i \in \mathcal{I}(Pr(X = i) = \frac{1}{500{,}000})$, then his argument is probabilistically sound.
- But surely something is not right with this situation.

If classical probability takes no issue with John's argument, then perhaps we are in need of a sharper tool.

# Enter Kolmogorov Complexity...

Let us recap and sketch the basic idea:

- John claims that Donna's win is a random event, as random as anybody's else win.
- Intuitively Donna's win is not random at all.
- We smell a rat.
- What does probability theory say about that?
- Classical probability theory says nothing: **It does not have the notion of random events, only the notion of random variables**.

This was the problem that attracted the attention of Kolmogorov as well as Chaitin and Solomonoff, the three founding fathers of information complexity.

# An Intuitive look at Kolmogorov Complexity I...

- Flip a fair coin 50 times and write 1 if it comes heads and 0 if it comes tails.
- Let $s$ be the resulting binary string.
- Compare $s$ with another possible outcome $t$ which is 01 repeated 25 times.
- Obviously $s$ looks "more random" than $t$.
- But why? What makes it look more random?

Kolmogorov complexity has a great answer: string $t$ is easier to generate than string $s$. The harder it is generate a string, the more random the string is. Can this idea be made quantitative?

# An Intuitive look at Kolmogorov Complexity II...

Can this idea be made quantitative? Yes!

- Fix a universal Turing machine, and
- define the Kolmogorov complexity $K(x)$ of a string $x$ as the length of the shortest program generating $x$.
- The larger $K(x)$ is, the more random $x$ is.

There is, however, a problem with this approach. The information complexity $K(x)$ is **not computable**. Besides, how do you apply classical information complexity to the lottery scenario?

# An Intuitive look at Kolmogorov Complexity IV...

Kolmogorov complexity of binary strings gives rise to a probability distribution $\mathcal{P}(s)$ proportional to $2^{-K(s)}$ on binary strings known as the *a priori probability distribution*.

# An Intuitive look at Kolmogorov Complexity III...

It is easy to be a skeptic on Kolmogorov complexity. It is not computable and in that sense unknowable. Besides, it is not obvious what the fixed universal Turing machine should be. As a result $K(x)$ is known only up to an additive constant.

(Aside: How big is this constant typically? Kolmogorov thought that it length is hundreds of bits but not tens of thousands of bits. –Levin).

# How might KC apply to the Lottery?

- Inhabitants in Columbus have names. These names can be considered as strings of characters. Some names are then shorter than others.
- John knows Donna by at least two "short" names: 'Donna' and 'my wife.'
- There is a small number of inhabitants in Columbus for which John posesses "short" names. Let us call these inhabitants *special.* Almost all inhabitants are not special.
- Therefore, all else equal, the probability of a special inhabitant winning is small.
- According to Cournot's principle, (a priori named) small events do not happen.
- Thus, Donna should not have won.

Of course, to make this argument precise, one must provide a mathematical formalisation for the notion of "short names" and construct a probability distribution, $\mathcal{P}^*$, that takes this naming structure soundly into account. With the use of Kolmogorov's algorithmic complexity theory, such a feat may indeed be accomplished.

But, ...

Computability issues aside, there is still a problem. Consider the following addendum:

- Before the lottery, John met Amy at the grocery store and committed her name to memory.
- Thus, Amy is special.
- John knows Amy by a name which is, as a character string, even shorter than any name by which he knows Donna.
- Hence, Amy is (w.r.t. $\mathcal{P}^*$) even *more special* than Donna.
- One would then expect the probability of Amy winning to be, according to $\mathcal{P}^*$ , an event of smaller probability than that of Donna winning.
- But, no one would argue that Amy winning would be a sign of unfairness. Thus, Donna winning should not be a sign of unfairness and John should not be suspected of having done anything wrong.

# Moral crux: Closeness of relationship, not shortness of name!

In its reliance upon names as character strings, the above approach misses the point. Donna winning is suspect not because of the short character length of her name known to John, but because of her *close relationship* to John. In this way, we should look for a variant of Kolmogorov complexity, and hence Kolmogorov randomness, that relies upon a notion of naming that is essentially *algebraic*, not characteristic.

Returning to Example 1, let us attempt to construct an argument that John most likely cheated which takes his relationship to Donna into account. Informally, it might go as follows.

- Donna is John's wife. She is thus very close to him.
- If you are very close to someone, your lives are likely intertwined in complex ways, and you have myriad incentive for wanting them to win a game that you believe could improve their life.
- John believes that Donna winning the lottery could improve her life.
- There are few, if any, other inhabitants who are this close to John.
- There are few, if any, other inhabitants for which John would have as much incentive to cheat and manipulate the game in their favour.
- Donna winning is then a small event.
- According to Cournot's principle, small events do not happen.
- Thus, Donna should not have won.

# How does this fix the Amy lacuna?

- Recall that John and Amy met only briefly in the grocery store, enough for him to know her name.
- Amy is thus not close to John, and he has no remarkable incentive for wanting her to win.
- There are many other inhabitants who are not close to John. In fact, almost all of them are like Amy in this way.
- Thus, Amy winning is not a small event, and should not arouse suspicion.

Let us recap Cournot's principle in a slightly different form, depending more on *independence* of event specification than on *prior* specification:

> *Independently specified rare events do not happen.*

- Every event can be specified after it occurs but such specification depends on event happening.
- We are talking about independent specifications.
- For example, the specification could be given ahead of time.
- What's important for us is that a specification can be so natural that it is accepted as a legitimate independent specification even after the event happened.

Thus, we would like to argue that *descriptive complexity* provides a natural way to apply Cournot's Principle to a whole slew of interesting settings:

> *the rare events that we define w.r.t. our models are naturally described and this allows us to apply Cournot's Principle to them.*

That is, by having "short, natural descriptions," these events (such as Donna winning the lottery) have been a priori *independently specified*.

We are almost ready to construct a simple model, but let's look at one more example first.

Denis is playing a strange but intriguing game involving a large freely-moving globe and a needle. His game is played as follows.

1. He is blind-folded.
2. He spins the globe aggressively along all three spatial axes.
3. When the globe stops moving, he places the needle in his right hand, positions his hand over his head, and proceeds to bring his hand straight forward until the pin punctures the globe.
4. He then removes his blind-fold and smiles with revelry. Denis has placed the pin directly upon the equator.

It seems very unlikely that Denis accomplished this task without cheating. But why? Surely the landing upon the equator is as least as likely as landing upon any other line of latitude.

In fact, because the equator is a *great circle*, it has a greater area than any line of lattitude that doesn't run through the center. As Denis must have landed upon some line of latitude, isn't it most likely that he would land upon the equator?

Nevertheless, the feeling that Denis's game was probably unfair persists.

An argument, similar to the character-based Kolmogorov approach to Example 1, might take the following form. By "subsets of the globe" we mean "subsets of the surface of the globe."

- Denis only knows names for a small number of subsets of the globe.
- The equator is a subset for which Denis knows a name.
- Almost all subsets of the globe do not contain the equator.
- Thus, landing on the equator is a small event.
- By Cournot's principle, Denis must have cheated.

This argument suffers from an obvious defect. Consider the mechanism of naming points on the globe by their longitudinal and latitudinal coordintes.

Because Denis is in the real world and constrained by physical (e.g., rational) measurements, we can then identify every point on the globe with a longitude-latitude pair $\langle x, y\rangle \in \{-180, \ldots, 180\} \times \{0, \ldots, 90\} \subset \mathbb{Q}^2$. In fact, Denis learned this naming scheme in school.

Thus, Denis knows a name for every point on the globe, and the claim that Denis only knows names for a "small" number of subsets of the globe no longer seems to hold water. Again we see that the property of "knowing a (short) name for" is not enough to capture the phenomena we seek.

The problem here is the lack of *naturality* in the naming system put forth. Sure, Denis may have a notation system for naming arbitrary subsets of the globe, but of those subsets, only a small number of them are in some sense *naturally* named with respect to common practices of human discourse.

Examples of natural names in this situation might be "Equator," "North Pole," "South Pole," "Paris," "St. Petersberg," "The Atlantic Ocean" for specific places (e.g., *constants* which do not vary depending on who is playing the game), or functions of the game configuration such as "CapitolOfHomeCountry," or "CityWhereMotherWasBorn," whose values depend on the player(s) of the game.

A probability distribution (for events of a game taking place w.r.t. a sufficiently large population)based upon a *natural naming system* should have the property that events with small, naturally named descriptions are *small* with respect to the probability measure.

We present a very restrictive mathematical model. It can be easily extended in various directions. But here we want just a proof of concept. In other words, all we want to do is just to make a point that the algebraic approach to the problem of descriptive complexity may be useful. We develop a model for "games" in the spirit of Example 1. In particular we make the following simplifying assumptions.

- There is a central player in our game (e.g., John in Example 1).
- We take into account only binary relations between players.
- We consider all these binary relations equally binding.

# Arenas and our Probability Distribution

- In general, the arena of a game in question is an arbitrary structure in the sense of mathematical logic.
- To simplify things, we presume that the universe of the structure is finite and that its vocabulary consists of one constant $c$ (naming the central player) and finitely many binary relations.
- Note that a binary relation $R(x, y)$ gives rise to multi-valued functions $x \mapsto \{y : \ xRy\}$ and $y \mapsto \{x : \ xRy\}$.

### Definition (Arena)

An *arena* $\langle D, c, f_1, f_2, \ldots \rangle$ is a finite set $D$ (the *population domain*) equipped with a distinguished element $c$ and multi-valued functions $f_1, f_2, \ldots$.

## Definition ($F_i$ Liftings)

Given an arena $\mathcal{S} = \langle \mathcal{D}, c, f_1, f_2, \ldots, \rangle$, it is convenient to associate with each $f_i$ a polymorphic *lifted marking function* $F_i : (\mathcal{D} + 2^{\mathcal{D}}) \rightarrow 2^{\mathcal{D}}$ as follows:

$$F_i(x) = \begin{cases} f_i(x) & \text{if } x \in \mathcal{D}, \\ \bigcup_{d \in x} f_i(d) & \text{if } x \in 2^{\mathcal{D}}. \end{cases}$$

From now on, we use *marking function* to mean a *lifted marking function.* Let $c \in \mathcal{D}$ be known as the *distinguished center* of the marking system $\mathcal{S}$.

### Definition (Marked member)

We say $d \in \mathcal{D}$ is *marked* if there exists some finite sequence of marking functions $F_{\alpha(1)}, \ldots, F_{\alpha(k)}$ s.t.

$$d \in F_{\alpha(1)}(F_{\alpha(2)}(\ldots(F_{\alpha(k)}(\{c\}))\ldots)),$$

where $1 \leq \alpha(i) \leq n$.

Definition (Markset)

A *markset* is a subset of the population domain $\mathcal{D}$ consisting of members which are marked.

Observe that a markset is a subset of the population domain (e.g., it is not a collection of marks but rather their denotations).

**Definition (Maximal Markset)**

Let the *maximal markset* of the marking system $\mathcal{S}$ be the collection of all marked members of $\mathcal{D}$ w.r.t. $F_1, F_2, \ldots$. We use the symbol $\mathcal{N}$ to represent this set.

## Definition (Level Hierarchy)

We obtain a *levelled hierarchy of marksets* from $\mathcal{S}$ w.r.t. $c$ as follows:

$$\mathcal{L}_0 = \{c\},$$

$$\mathcal{L}_{i+1} = \bigcup_{j=1}^{n} F_j(\mathcal{L}_i).$$

As $\mathcal{D}$ is finite, there must exist a least fixed point for the construction of this hierarchy. As we imagine the hierarchy of names as consisting of concentric circles built around the distinguished center $c$, we also refer to the index of this least fixed point as the radius of our arena.

Definition (Radius of $\mathcal{L}$)

Let the radius of our arena, $|\mathcal{L}| \in \mathbb{N}$, be defined as follows:

$$|\mathcal{L}| = r \text{ s.t. } \bigcup_{i=0}^{r} \mathcal{L}_i = \bigcup_{i=0}^{r+1} \mathcal{L}_i \ \wedge \ \forall 0 \leq s < r \ \bigcup_{i=0}^{s} \mathcal{L}_i \neq \bigcup_{i=0}^{s+1} \mathcal{L}_i.$$

It is important to note that a given member of the population domain may appear in more than one level of the markset hierarchy. For the construction of a probability distribution on suitable events in our model, it is important to have

- a refinement of the markset hierarchy which is in fact a *partitioning* of the set of marked members of $\mathcal{D}$, and
- access to the least level of the hierarchy of marksets in which a given member of the population appears.

Definition (Minimal Level Hierarchy)

$$\min \mathcal{L}_0 = \{c\},$$

$$\min \mathcal{L}_{i+1} = \mathcal{L}_{i+1} \setminus \bigcup_{j=1}^{i} \mathcal{L}_j.$$

Proposition ($\mathcal{N}$ Partitioning Theorem)

*The family of sets arising from the minimal level hierarchy*

$$\mathcal{M} = \left\{min\ \mathcal{L}_0, min\ \mathcal{L}_1, \ldots, min\ \mathcal{L}_{|\mathcal{L}|}\right\}$$

*is a partitioning of* $\mathcal{N}$*.*

Definition (Minimal Level of Named Member)

Given a named member $d \in \mathcal{N}$, we let

$$\mu(d) = i \text{ s.t. } d \in \min \mathcal{L}_i.$$

Proposition ($\mu$ uniqueness)

$\forall d \in \mathcal{N} \exists! 1 \leq i \leq |\mathcal{L}| \;\; s.t. \;\; d \in min\ \mathcal{L}_i.$

Let us reflect a bit upon what we have done.

- By constructing the minimal level hierarchy $\mathcal{M}$, we have performed a drastic symmetry reduction upon our marking system.
- Indeed, Proposition 1 guarantees us that we have placed each marked member of the population into an equivalence class over $\mathcal{N}$.
- Thus, from an algebraic perspective, we have taken a free algebra of marks whose set of denotations consists of $\mathcal{N}$ and quotiented by a set of equations of the form $\{(d_1 = d_2) \mid \mu(d_1) = \mu(d_2)\}$.
- That is, two marked members of the population are seen as equal when their shortest marks with both have the same length. It is precisely this form of symmetry reduction, alluded to in the introduction, which will make the probabilities associated with our models computable in practice.

Armed with this setup, we are now ready to define an appropriate probability distribution. Random variables will range over our set of levels, $\{0, \ldots, |\mathcal{L}|\}$.

## Definition (Probability Mass Function)

Given $k \in \{0, \ldots, |\mathcal{L}|\}$,

$$Pr(X = k) = \frac{|\text{min } \mathcal{L}_k|}{|\mathcal{D}|}.$$

It is clear that this defines a distribution if every member of the population is named, e.g. precisely when $\mathcal{N} = \mathcal{D}$. But, we will need our machinery to be more general:

- it is possible to build natural marking systems in which the entire population will not be marked.
- even in cases where the entire population is marked, it will become computationally infeasible to compute the entire markset hierarchy upon the population.

Thus, we will proceed by introducing an extra parameter into our machinery: $w$, the *wall* of the arena.

We will place all members of the population which are *not marked within $w$ steps* into the same equivalence class.

From the perspective of the model, members of the population $d_1$ which appear for the first time in some $\mathcal{L}_i$ s.t. $i > w$ (e.g., $d_1 \in min\ \mathcal{L}_i$) will be symmetric to members of the population $d_2$ which are unmarked (e.g., $d_2 \in (\mathcal{D} \setminus \mathcal{N})$).

Thus, we must introduce an extension of the minimal level hierarchy which takes this parameter into account.

### Definition (Walled arena)

Given an arena $\langle D, c, f_1, f_2, \ldots \rangle$ and some $w \in \mathbb{N}$, $\langle D, c, w, f_1, f_2, \ldots \rangle$ will be called a *walled arena* and $w$ its *wall*.

From now on, let us assume we are working with a *walled arena* $\langle D, c, w, f_1, f_2, \ldots \rangle$.

**Definition (Walled off subset: $\mathcal{L}_\infty$)**

We define the *walled off subset of the population* $\mathcal{L}_\infty \subseteq \mathcal{D}$ as follows:

$$\mathcal{L}_\infty = \min \mathcal{L}_\infty = (\mathcal{D} \setminus \bigcup_{i=0}^{w} \mathcal{L}_i).$$

Observe that if $w \geq |\mathcal{L}|$ then $\mathcal{L}_\infty = (\mathcal{D} \setminus \mathcal{N})$.

Theorem ($\mathcal{D}$ Partitioning Theorem)

*The family of sets arising from the extended minimal level hierarchy*

$$\mathcal{M} = \{min\ \mathcal{L}_0, min\ \mathcal{L}_1, \ldots, min\ \mathcal{L}_{|\mathcal{L}|}, \mathcal{L}_\infty\}$$

*is a partitioning of* $\mathcal{D}$.

Now, we may extend our probability mass function so that random variables take on values over $\{0, \ldots, |\mathcal{L}|, \infty\}$.

Definition (Probability Mass Function, revised)

Given $\alpha \in \{0, \ldots, |\mathcal{L}|, \infty\}$,

$$Pr(X = \alpha) = \frac{|\min \mathcal{L}_\alpha|}{|\mathcal{D}|}.$$

Theorem (Distribution Theorem)

$$\left( \sum_{\alpha \in \{0, \ldots, |\mathcal{L}|, \infty\}} Pr(X = \alpha) \right) = 1.$$

And we at last see how our logico-algebraic framework may be used to address our examples.

Let us suppose we are given a walled arena and are asked to compute the probability that the member of the population $d \in \mathcal{D}$ was chosen at random. Observing our symmetry reductions, we can answer this question by computing the probability that someone with a name of length at most as long as the shortest name of $d$ was chosen w.r.t. the wall $w$. That is, we compute this probability as

$$\mathcal{P}(d) = \begin{cases} Pr(X \leq \mu(d)), & \text{if } d \in \mathcal{N} \\ Pr(X = \infty) & \textit{otherwise}. \end{cases}$$

Let $\mathcal{D}$ be the population of Columbus, GA with $|\mathcal{D}| = 500,000$. Let us propose the following as natural lifted marking functions: $\mathit{Lover}, \mathit{Sibling}, \mathit{Parent}, \mathit{Child}, \mathit{CloseFriend}$. As *John* is the distinguished center in this example, we have $\mathcal{L}_0 = \{\mathit{John}\}$. To keep the example a reasonable size, let us set $w = 3$.

So, we build the *markset hierarchy* as follows:

$$\begin{aligned}
\mathcal{L}_0 &= \{John\}, \\
\mathcal{L}_1 &= Lover(John) \cup Sibling(John) \cup Parent(John) \\
&\quad \cup Child(John) \cup CloseFriend(John) \\
&= \{Donna\} \cup \{Dianne\} \cup \{June, Greg\} \cup \{Grant\} \cup \{Bill\} \\
&= \{Donna, Dianne, June, Greg, Grant, Bill\}, \\
\mathcal{L}_2 &= Lover(\mathcal{L}_1) \cup Sibling(\mathcal{L}_1) \cup Parent(\mathcal{L}_1) \cup Child(\mathcal{L}_1) \cup Close \\
&= \{John, Ron, Greg, June, Erika, Lolly, Sue, Hannah, Brett, Brant, B \\
\mathcal{L}_3 &= Lover(\mathcal{L}_2) \cup Sibling(\mathcal{L}_2) \cup Parent(\mathcal{L}_2) \cup Child(\mathcal{L}_2) \cup Close \\
&= \{Donna, Dianne, Horace, Jan, Lyle, Ron, Greg, June, Fernando, Er \\
&\quad Carly, Hanna, Brett, Brant, Chun, Grant, Barry, Thomas, Terry, E \\
&\quad Iulia, Dominicka, Meg, Amy, Christine, Ben, David, Charles, Jim,
\end{aligned}$$

Now, we refine the above hierarchy into its minimal counterpart:

$$
\begin{aligned}
min\ \mathcal{L}_0 &= \{John\}, \\
min\ \mathcal{L}_1 &= \{Donna, Dianne, June, Greg, Grant, Bill\}, \\
min\ \mathcal{L}_2 &= \{Ron, Erika, Lolly, Sue, Hannah, Brett, Brant, Barry, Beth, Iu \\
min\ \mathcal{L}_3 &= \{Horace, Jan, Lyle, Fernando, Chun, Barry, Thomas, Terry, B \\
&\quad Dominicka, Meg, Amy, Christine, Ben, David, Charles, Jim, V
\end{aligned}
$$

So, 36 members of the population are within the arena walls. Let us say we're interested in the probability of *Erika* being selected. Then, we have

$$\mu(Erika) = 2 \quad \text{and} \quad |\bigcup_{i=0}^{2} \mathcal{L}_i| = 17$$

thus

$$Pr(X = `Erika') = \frac{17}{500,000} \approx 3.4 \times 10^{-5}.$$

And if *Stranger* is someone not *w*-close to *John*, then

$$Pr(X = `Stranger') = \frac{499964}{500,000} \approx 0.99.$$

# Extending the model

There are many ways to generalise this machinery:

- Multiple centers,
- Ternary, 4-ary, 5-ary, ... relations giving rise to MVFs,
- Incorporation of weights upon relationships,
- ... .

For multiple centers, could place them together into $\mathcal{L}_0$ – perfect symmetry between them – or, consider more robust notions: a sheaf-theoretic construction (sketch on board if time).

Finally, can a mathematical analysis of *naturality* of naming systems be accomplished?

Consider measures of *biodiversity* in ecosystems – that is, the *effective cardinality* of metric spaces.
This is very natural – it's just the Euler characteristic of a category enriched in $\mathbb{R}$!

In this way, we can use a lot of computational machinery from stable homotopy theory to analyse the structure of our naming systems.

# In conclusion

- Motivated by a few examples, we've given a very simple theory for modelling and obtaining probability distributions justifying accusations of unfairness.
- We've argued that Cournot's Principle can be applied directly to events with small probabilities in suitable models constructed in this way, as the shortness of a natural name gives in a sense an independent specification.
- The resulting machinery is nothing more than a *very simple* relative (special case?) of Kolmogorov Complexity.
- Some neat possibilities: think about FB!

Thank you!