Order, Disorder, and Adaptivity for Stochasticity

Justin Miller
justin.d.miller@dartmouth.edu
Dartmouth College
Joint work with Liling Ko, ko.390@osu.edu

April 15th, 2023

NERDS 23.0

Flipping Coins

We have a sequence of countably many coins. Each coin exhibits a 1 (heads) or 0 (tails). The coins are identically p-biased, i.e. flipping one would result in 1 with probability p.

We think of this sequence as a set of natural numbers, with n being in the set if and only if the n-th coin shows a 1. While we generally associate coins with random outcomes via coin flips, nothing is stopping one from arranging a sequence of coins to have a specific, desired outcome.

Coin Sequences

Evens:

...

0 1 2 3 4 5 6 7

Odds:

...

0 1 2 3 4 5 6 7

Primes:

...

0 1 2 3 4 5 6 7

The Stochasticity Game

Each coin in the sequence has been hidden under a cup. We don't know which side is face up until we look under the corresponding cup.

Our opponent (the house) claims that the hidden sequence was obtained by flipping each coin randomly, with no modification. We are challenged to prove them wrong by finding a subsequence of coins which fails the law of large numbers.

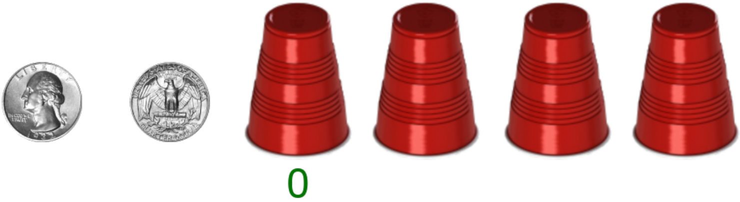
The sequence is said to be stochastic if our opponent wins the game, i.e. we cannot find such a subsequence. Depending on how we define legal plays for this game, we obtain different notions of stochasticity.

Hiding the Coins

Hiding the Coins

Hiding the Coins

Hiding the Coins



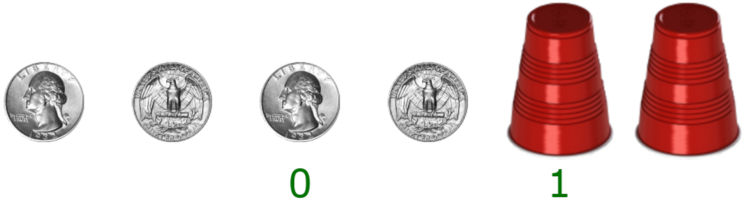
Hiding the Coins

0

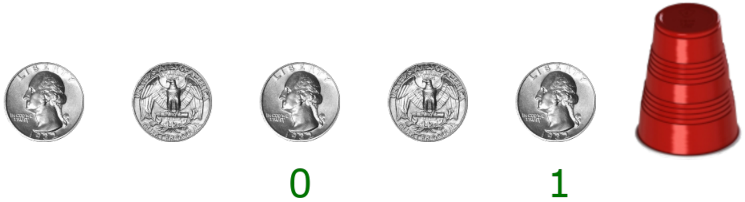
Hiding the Coins

0

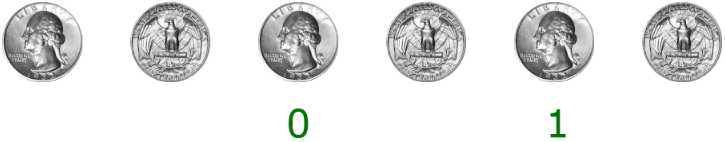
Hiding the Coins



Hiding the Coins



Hiding the Coins



Hiding the Coins

Adaptivity

A stochasticity notion is called **adaptive** if we are allowed to incorporate previously revealed information into our strategy. Stochastic notions under which we cannot use information revealed about the subsequence are called non-adaptive.

Order

A stochasticity notion is called **disorderly** if we are allowed to make decisions on coins regardless of their order in the original sequence. Stochastic notions which force us to make decisions in the order imposed by the original sequence are called **orderly**.

The Law of Large Numbers

Definition

The asymptotic density of A, if it exists, is

$$\rho(A) = \lim_{n\to\infty} \frac{|A \upharpoonright n|}{n}$$

The law of large numbers states that a countable sequence A generated by independent p-Bernoulli random variables satisfies $\rho(A) = p$ with probability 1.

Tools for Working with Density

Definition

Given

$$A = \{a_0 < a_1 < a_2 < \dots\}$$

and

$$B = \{b_0 < b_1 < b_2 < \dots\}$$

$B \triangleleft A$, "B within A", is defined via

$$B \triangleleft A = \{n : a_n \in B\}$$

This notation is useful because $B \triangleleft A$ takes $B \cap A$ and turns it into a sequence indexed by the natural numbers.

von Mises-Wald-Church Stochasticity

A **selection function** is a function $f : 2^{<\omega} \to \{\text{Yes}, \text{No}\}$. Given A,

$$f(A) = \{n : f(A \restriction n) = \text{Yes}\}$$

We say A is p-MWC stochastic if

$$\rho(A \triangleleft f(A)) = p$$

for all partial computable selection functions f. A is Church stochastic if it wins on all total computable selection functions.

Church and MWC stochasticity are both orderly and adaptive: decisions are made in order, and previous information can influence future decisions.

MWC Stochasticity

MWC Stochasticity

MWC Stochasticity

MWC Stochasticity

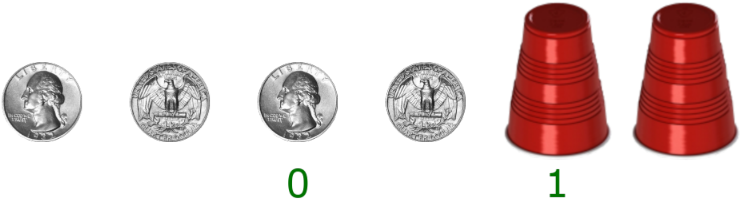
MWC Stochasticity

0

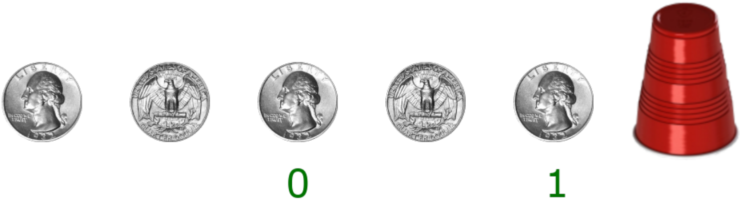
MWC Stochasticity

0

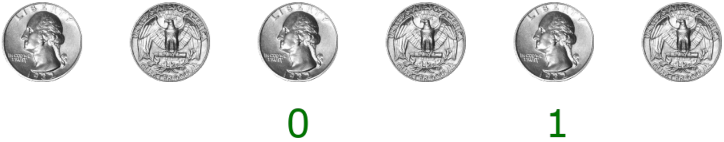
MWC Stochasticity



MWC Stochasticity



MWC Stochasticity



MWC Stochasticity

Kolmogorov-Loveland Stochasticity

KL-stochasticity is a stochasticity notion which is disorderly and adaptive. Much like MWC stochasticity, we can use the information from any previously visited coin in our decisions. We can also visit coins out of order: in addition to a selection function which determines whether or not we should include the next coin in our subsequence, we have a **scan rule** which determines what coin we will visit next.

KL-Stochasticity is disorderly and adaptive.

KL Stochasticity

KL Stochasticity

0

KL Stochasticity

KL Stochasticity

KL Stochasticity

KL Stochasticity

KL Stochasticity

KL Stochasticity

KL Stochasticity

KL Stochasticity

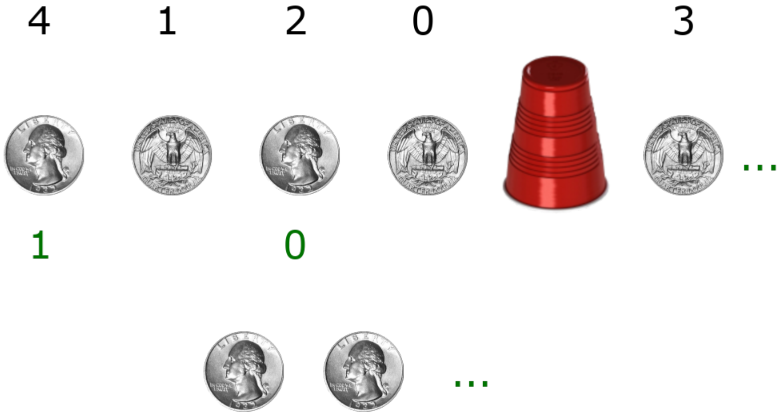
KL Stochasticity

KL Stochasticity

KL Stochasticity

KL Stochasticity

KL Stochasticity



Injection Stochasticity

A set A is p-injection stochastic if

$$\rho(f^{-1}(A)) = p$$

for all total computable injections f. This can be thought of as “uniform KL-stochasticity” because this is equivalent to having a scan rule and selection function ignore the provided information and just operate computably. This is a disorderly, non-adaptive stochasticity notion.

Astor [Ast15] proved that this is equal to intrinsic density, which is having density p under every computable permutation.

Injection Stochasticity

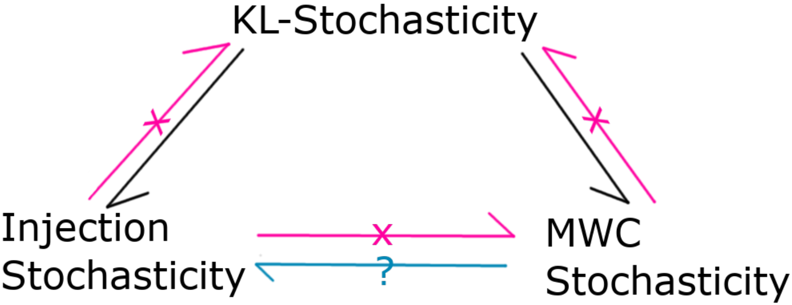
Injection Stochasticity

Injection Stochasticity

Injection Stochasticity

Injection Stochasticity

The Zoo thus Far



The implications (black arrows) are immediate from the definition of KL-Stochasticity being the most general.

The separation of MWC and KL (pink arrow) is due to Merkle, Joe Miller, Nies, Reimann, and Stephan. [MMN$^+$06].

[Mil21] proved that intrinsic density is preserved under joins, so the other two separations are immediate.

Disorderly, Non-Adaptive?

Our main motivation for this project was the following: If

$$\rho(A \triangleleft C) = p$$

for all infinite computable sets C, then does A have intrinsic density (injection stochasticity) p?

It turns out that this is equivalent to asking if “uniform MWC-stochasticity” implies “uniform KL-stochasticity.” As we’ve mentioned, the latter is injection stochasticity/intrinsic density, but to our knowledge the former has not been seriously studied.

Increasing Stochasticity/Computable Density

A set A is p-Increasing stochastic if

$$\rho(f^{-1}(A)) = p$$

for all total computable (strictly) increasing functions f. Alternatively, $\rho(A \triangleleft \mathit{range}(f)) = p$ for all such functions. This justifies the alternative name computable density.

This is “uniform MWC stochasticity,” and as such is disorderly and non-adaptive.

Order vs Disorder in the Non-Adaptive Setting

It is immediate that injection stochasticity implies increasing stochasticity, and equivalently, that intrinsic density p implies computable density p. But what about the converse?

It turns out that the converse fails, the proof of which gives rise to an interesting construction technique.

Theorem (Ko, M.)

There is a set A which has computable density 0 (a computably small set), but there is a permutation h of the naturals such that

$$\overline{\rho}(h^{-1}(A)) = 1$$

That is, A is as far as possible from having intrinsic density 0 (it is not intrinsically small).

Generalizing Beyond 0

An immediate objection one might have to this theorem is that density/stochasticity 0 and 1 are the least interesting variants from a randomness perspective. In terms of our randomness motivation, a 1-biased coin doesn't have any interesting meaning. However, the following lemma shows that the case for 0 is sufficient to obtain the more interesting ones.

Lemma

If there is a set A which has computable density 0 *but not intrinsic density* 0*, then for any* α *there is a set B with computable density* α *but not intrinsic density* α*.*

Proof

Proof.

Given such A, let B have intrinsic density α. Consider $B \setminus A$. Then it is straightforward to show that $B \setminus A$ has computable density α. If it does not have intrinsic density α, then we are done.

Then suppose it does. It then follows by a simple calculation that $B \cup A$ cannot have intrinsic density α since A doesn't have intrinsic density 0, but $A \cup B$ must also have computable density α. □

This style of proof actually tells us even more. We can show that all of these degrees are closed upwards using the methods from [Mil21], so these counterexamples are also closed upwards in the Turing degrees.

Proof Sketch of Main Theorem

Theorem (Ko, M.)

There is a set A which has computable density 0 (a computably small set), but there is a permutation h of the naturals such that

$$\overline{\rho}(h^{-1}(A)) = 1$$

That is, A is as far as possible from having intrinsic density 0 (it is not intrinsically small).

Our goal is to build a computable permutation h and a set A such that $\rho(f^{-1}(A)) = 0$ for all increasing computable functions f, but $\overline{\rho}(h^{-1}(A)) = 1$. We shall use the trick that we can enumerate the increasing functions/computable sets by enforcing that Turing machines never output a number less than or equal to those seen before. This gives us an effective enumeration f_i of the increasing computable functions. (Many of which are partial and have finite domain.) h will be built up via a series of finite permutations between intervals $[n, m]$.

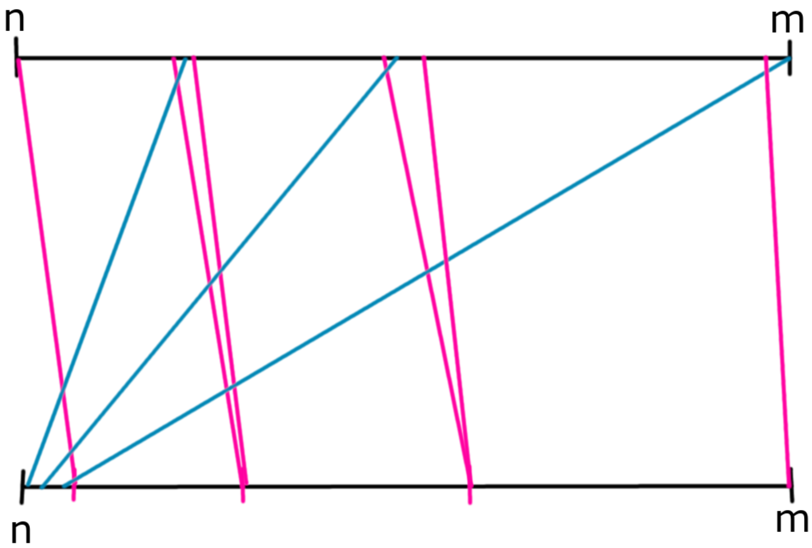
Proof Sketch

Let's first look at a single increasing function f. There are two ways we can win:

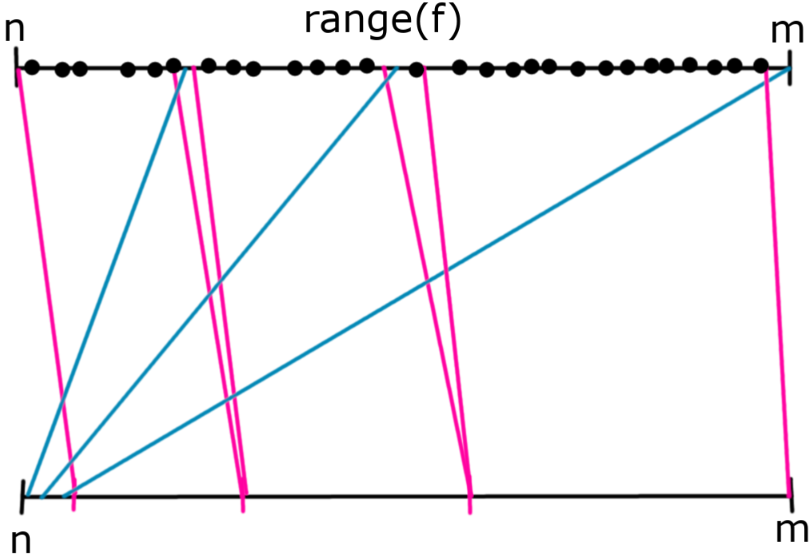
- *range*(f) contains many elements, and each "hit" (i.e. element of A) is preceded by many "misses." Then $f^{-1}(A)$ will have small density due to the misses, so we just need to ensure that $h^{-1}(A)$ is large.
- *range*(f) contains very few elements. In this case, we concentrate A away from the points of *range*(f) so that $h^{-1}(A)$ is big, but *range*(f) doesn't see any elements of A.

We construct h in such a way as to balance these two outcomes so that the construction of A is capable of leveraging either option to defeat f. This will be done by making h first increase very quickly on a block of numbers to beat the first case, then map larger blocks to the gaps between these outputs to beat the second case. The sizes of each block will increase quickly enough to ensure that a single block containing elements many of A will send the density towards 1.

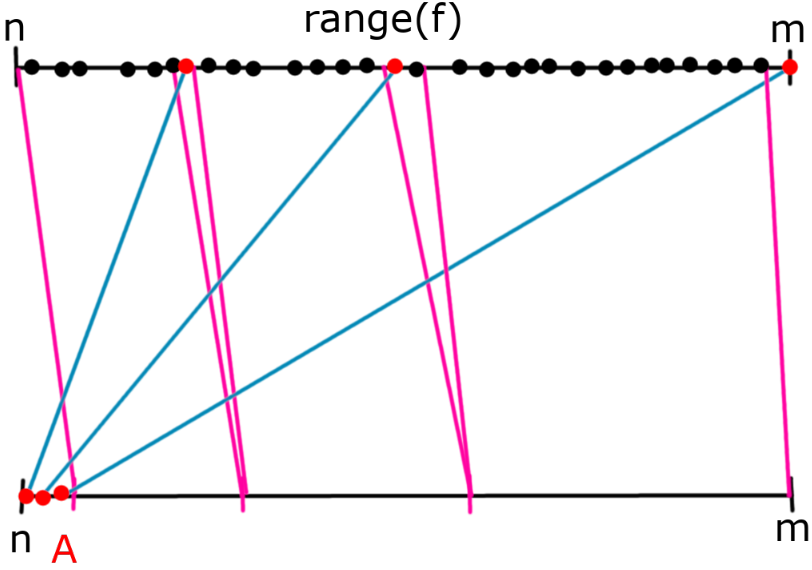
Examples



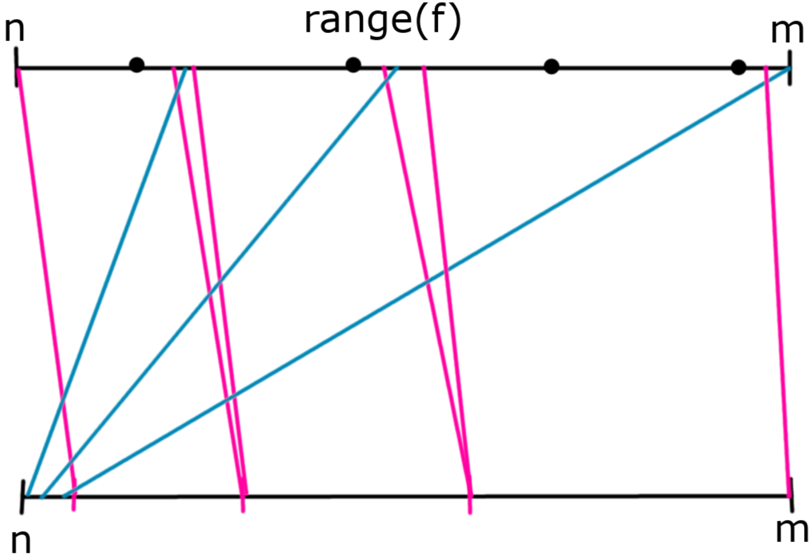
Examples



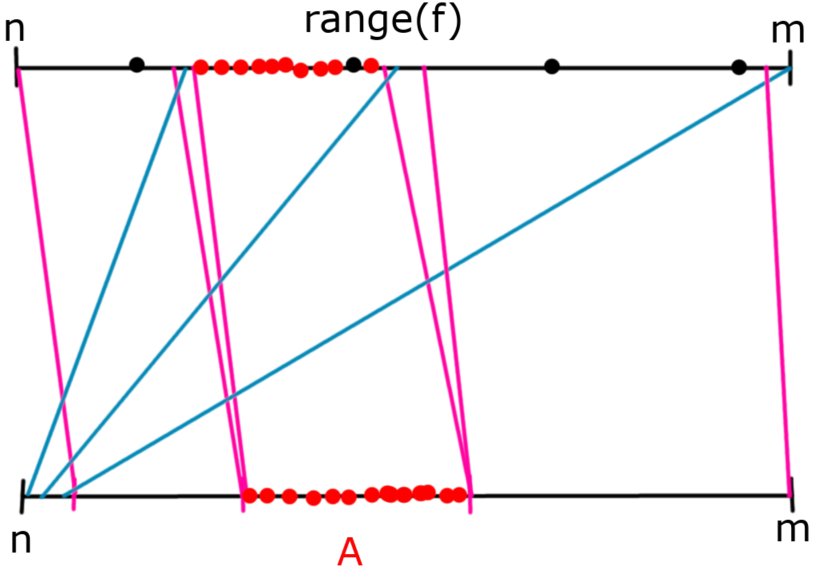
Examples



Examples



Examples



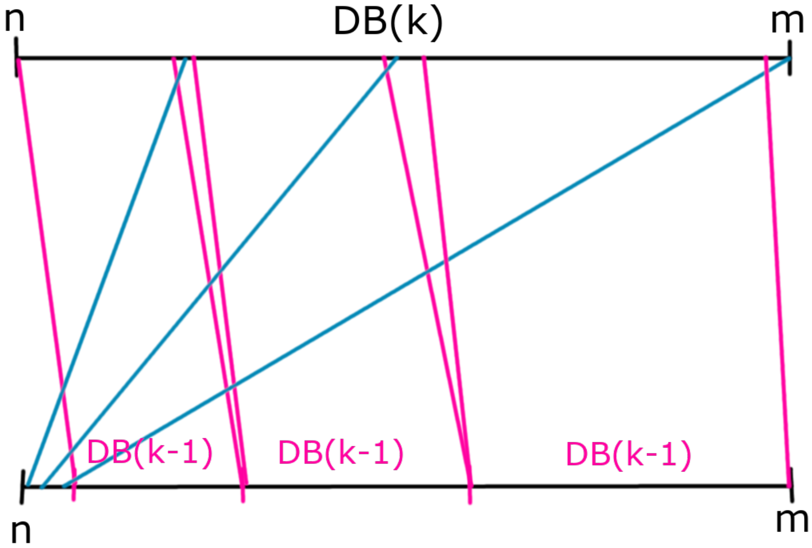
Proof Sketch

This proof method runs into difficulty when we have more than one f. In looking at a block of numbers, we might need to defeat f_0 via the first method, but defeat f_1 via the second method. How can we achieve both at once?

The solution is a recursive construction: the previous example, of a quick increase followed by filling in the gaps in order, will be the base level. Then for each subsequent level, the gaps will not be filled in order, but by a structure of the previous level in our recursion! These will be called *Disordered Blocks*, with a level n disordered block capable of defeating n increasing functions.

As an example: the first level will have each gap mimic the base level by having a large increase then filling the gaps in order. So the first level will start with an increase, then have periods of increasing followed by periods of gap filling.

Examples



Proof Sketch

We're now ready to define h. Let $h(0) = 0$.

Now having defined h up through k_n via a level n disordered block, define h from $k_n + 1$ to k_{n+1} using a level $n + 1$ disordered block. (The k_i's, the lengths of the blocks, are determined by k_n and $n + 1$.)

h will be a permutation because it is a sequence of finite permutations stacked on top of each other. It will be computable because the recursive structure we use to define disordered blocks is computable.

Building A

With our permutation in hand, we're ready to build A. With access to $\emptyset'$, we can know whether or not a given f_i converges up to some finite point, i.e. if its range is defined up through a given disordered block inside h.

Then $\emptyset'$ can determine for each of the first n f_i's which winning condition applies inside the disordered block of level n, and fill in A accordingly to satisfy the conditions that the density of $f^{-1}(A)$ is small but the density of $h^{-1}(A)$ is large.

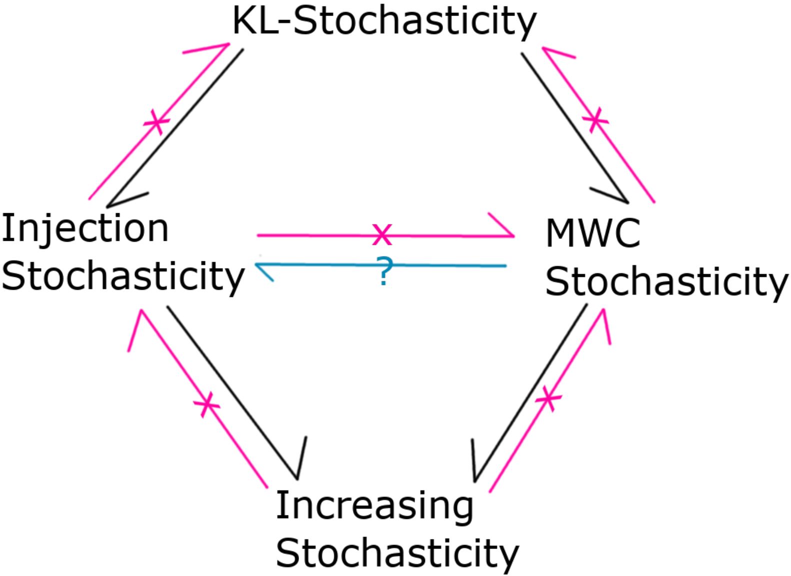
By continuing in this manner across every disordered block, $\emptyset'$ can compute A with the desired properties.

Applications of the Proof

The interesting corollary of this proof is that the specific h that we constructed can be re-used for different proofs: since h is computable, it cannot depend on A or knowing how the f_i's behave.

For example, we can prove that sets with computable density 0 not having defined intrinsic density satisfy cone avoidance, and h witnesses the failure to have intrinsic density no matter what cone we are avoiding.

The New Zoo


KL-Stochasticity
Injection Stochasticity
MWC Stochasticity
Increasing Stochasticity
?

Order vs Adaptivity

The major open question in this zoo is whether or not MWC stochasticity implies injection stochasticity, i.e. whether adaptive, orderly strategies outperform disorderly, non-adaptive strategies.

Our proof of the main theorem can be improved upon to make some progress on this question.

Skipping Ahead

A skip sequence is a finite sequence of ordered pairs $\langle n, b \rangle$ where the first coordinates are increasing natural numbers and the second coordinates are 0 or 1. Let S be the set of skip sequences.

A skip rule is a function $f : S \to \omega$ with $f(\sigma)$ being strictly larger than n, where $\langle n, b \rangle = \sigma(|\sigma| - 1)$. Given a set A, the sequence generated by f on A is defined recursively via:

- $f(A)(0) = A(f(\emptyset))$
- $f(A)(n+1) = A(f(f(A) \upharpoonright n))$

In other words, skip rules generate subsequences using only the information in A that is seen along the way.

Weak Stochasticity

A set A is p-weakly (adaptively) stochastic if

$$\rho(f(A)) = p$$

for all computable skip rules f.

One can think of this as an “honest” version of MWC stochasticity. MWC stochasticity allowed us to use the information from all previous coins when making decisions, regardless of if we included that coin in our subsequence or not. Weak stochasticity only allows us to look under a cup if we are going to include that coin in our subsequence.

Weak Stochasticity

Weak Stochasticity

0

Weak Stochasticity

Weak Stochasticity

Weak Stochasticity

Weak Stochasticity

Adaptivity vs Disorder

Theorem (Ko, M.)

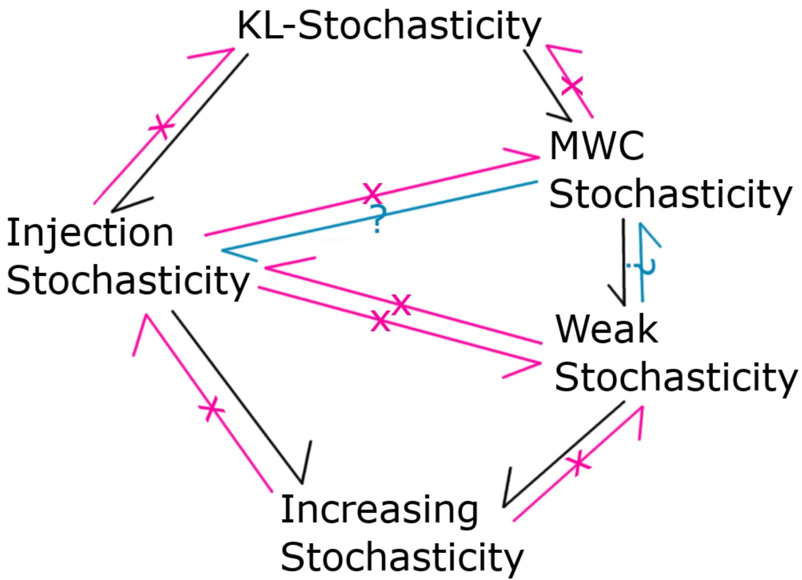
There is a set A which is 0-weakly stochastic, but

$$\rho(h^{-1}(A)) = 1$$

This provides a separation between orderly, adaptive stochasticity notions and disorderly, non-adaptive ones, which is a new result.

There is still work to be done: it is currently open whether or not weak stochasticity implies MWC stochasticity. We conjecture that it does not. If it does, then this is already a separation of injection and MWC.

The Current Zoo



References. Thank you.

Eric P. Astor.
Asymptotic density, immunity, and randomness.
Computability, 4(2):141–158, 2015.

Justin Miller.
Intrinsic Density, Asymptotic Computability, and Stochasticity.
PhD thesis, 2021.

Wolfgang Merkle, Joseph S Miller, André Nies, Jan Reimann, and Frank Stephan.
Kolmogorov–loveland randomness and stochasticity.
Annals of Pure and Applied Logic, 138(1-3):183–210, 2006.