

CAS II: Symmetric Partitions as Kolmogorov Models

Romie Banerjee

September 2026

Abstract

In algorithmic statistics a string x is explained by a finite set containing it, and Kolmogorov's structure function records the smallest such model at each level of complexity. The strong models of Vereshchagin, those computable from the data by a total algorithm, are essentially the cells of simple partitions. So a partition of $\{0, 1\}^n$ can be read as a *hypothesis*, and the cell containing x as the model it assigns. We develop algorithmic statistics over *symmetric* partitions, the orbit partitions of groups acting on strings.

The Galois connection between subgroups and partitions assigns to each ambient group G a lattice Par_G of symmetric partitions. Each has a canonical certificate whose cost equals that of the partition, and hypotheses can be combined by joins and meets. The resulting structure function h_x^G and symmetric sophistication $\text{soph}^G(x)$ measure *which part of the regularity of x is symmetric*.

For $G = \text{Sym}(\{0, 1\}^n)$ every partition is symmetric. Cells then recover all Kolmogorov models, and cells of cheap partitions recover exactly the strong models, so normal and strange strings are characterised by symmetry. For $G = \text{GL}(n, 2)$ the cells are exactly the *linearly homogeneous* sets, which turns linear symmetry into a restricted model class. For every nonzero x , h_x^{GL} lies in a band between $C(x) - \alpha$ and $n - \alpha$, and both edges are attained. In particular, there are stochastic normal strings whose simple structure is entirely invisible to linear symmetry: $\text{soph}(x) \approx 0$ but $\text{soph}^{\text{GL}}(x) \approx C(x)$.

Finally, we coordinatise the space of permutation groups. Each group is an element of a Burnside ring (its *type*) together with a permutation (its *placement*), and restriction moves refine partitions cell by cell via the Mackey formula. In these coordinates the collapse for Sym is a statement about placement, a linear hypothesis is determined by its type up to n^2 bits, and the maximal gap theorem shows that any space of symmetry hypotheses small enough to search is small enough to miss simple structure.

1 Introduction

1.1 Models, strong models and partitions

Kolmogorov proposed to explain a string x by a finite set $S \ni x$, a *model*, and to measure the explanation by the two-part code $C(x) \lesssim C(S) + \log |S|$: first the model, then the position of x in it, which is treated as noise [3, 1, 11, 10]. The structure function $h_x(\alpha)$ records the smallest model affordable at complexity α , and a model is a sufficient statistic when the two parts together cost no more than $C(x)$.

Arbitrary finite sets are too generous as models. The universal models of Gács, Tromp and Vitányi [1] can be sufficient, yet they encode the number of strings of bounded complexity, so they cannot be found from the data. Vereshchagin [9] therefore singled out the *strong* models: those computable from x by a total algorithm. A total algorithm $x' \mapsto A(x')$ assigns a model to *every* string, and Milovanov [6] observed that this makes strong models essentially the *cells of simple partitions* of $\{0, 1\}^n$. So beneath algorithmic statistics lies a theory of partitions. A partition p is a *hypothesis*: a classification of all strings into classes of strings the hypothesis cannot tell apart. The model for x is its cell, and the description becomes

$$C(x) \lesssim \underbrace{C(p)}_{\text{hypothesis}} + \underbrace{C(B | p)}_{\text{which cell}} + \underbrace{\log |B|}_{\text{noise}}, \quad x \in B \in p. \quad (1.1)$$

1.2 Symmetric partitions

Which partitions are *structured* hypotheses? The most natural source of classifications in mathematics and physics is symmetry: two strings are indistinguishable if some transformation from a group maps one to the other. The hypothesis is then the *orbit partition* of a group U acting on strings, and the size of the cell of x is governed by its symmetry. By the orbit–stabiliser theorem,

$$\log |U \cdot x| = \log |U| - \log |\text{stab}_U(x)| = -\log \Pr_{u \in U} [ux = x],$$

so strings with large stabilisers lie in small cells and receive short descriptions in (1.1). This is the precise sense in which symmetric objects are simple: a symmetry hypothesis explains x well when it is cheap and x is fixed by a large part of it.

Different groups can have the same orbits and then make the same hypothesis, so the group is best seen as a *certificate* that a partition arises from symmetry. The classical Galois connection between subgroups and partitions [2] makes this precise. For an ambient group G acting on strings it singles out a lattice Par_G of *symmetric partitions*. Each symmetric partition has a canonical certificate, the largest subgroup with those orbits, which costs exactly as much as the partition. Hypotheses can be combined by joins and meets, and conjugation acts as relabelling. The theory of this paper is algorithmic statistics restricted to cells of symmetric partitions.

The choice of ambient group matters. For $G = \text{Sym}(\{0,1\}^n)$ every partition is symmetric, and symmetry adds nothing to the theory of strong models. Its certificates turn out to be structurally trivial symmetries placed in complicated ways (Section 5). At the other extreme, the orbit partition of the stabiliser of x in a large group has $\{x\}$ as a cell: a lossless description that costs $C(x)$. The informative cases lie in between and arise from structured ambient groups, above all the group $\text{GL}(n, 2)$ of linear symmetries of $\mathbb{F}_2^n$.

1.3 The question and the results

$C(x)$ can always be recovered from a symmetric partition at full price, so the question is not *whether* but *how*: *which part of the regularity of x is captured by cells of symmetric partitions, and at what cost?* We measure this by the structure function h_x^G over Par_G , and summarise it by the symmetric sophistication $\text{soph}^G(x)$, the least cost at which a cell of a symmetric partition is a sufficient statistic.

- (1) *Framework* (Sections 3–4). The Galois connection between subgroups and partitions: symmetric partitions, their canonical certificates and costs, an algebra of joins and meets, and relabelling. Partition structure functions come with a dictionary between the group and partition formulations.
- (2) *All permutations* (Section 5). Every partition is symmetric. Cells of symmetric partitions recover all Kolmogorov models, and cells of cheap ones recover exactly the strong models. The certificates realising arbitrary models have trivial type: all their information sits in the placement.
- (3) *Linear symmetry* (Section 6). The cells of Par_{GL} are the linearly homogeneous sets, but whole symmetric partitions are rigid. A linear hypothesis is determined by its type up to n^2 bits.
- (4) *A band and its extremes* (Sections 7–8). For nonzero x , $C(x) - \alpha \leq h_x^{\text{GL}}(\alpha) \leq n - \alpha$, both edges are attained, and there are stochastic normal strings with $\text{soph}(x) \approx 0$ but $\text{soph}^{\text{GL}}(x) \approx C(x)$: their regularity is simple and computable, yet no linear symmetry captures any of it. A figure places all the structure functions of the paper in one picture.
- (5) *Coordinates and search* (Section 9). Every permutation group is a Burnside ring element (its *type*) together with a permutation (its *placement*), which gives coordinates on the spaces of permutation groups and symmetric partitions, and moves for navigating them. The maximal gap theorem limits what any search in these coordinates can achieve: coordinates small enough to search are small enough to miss simple structure.

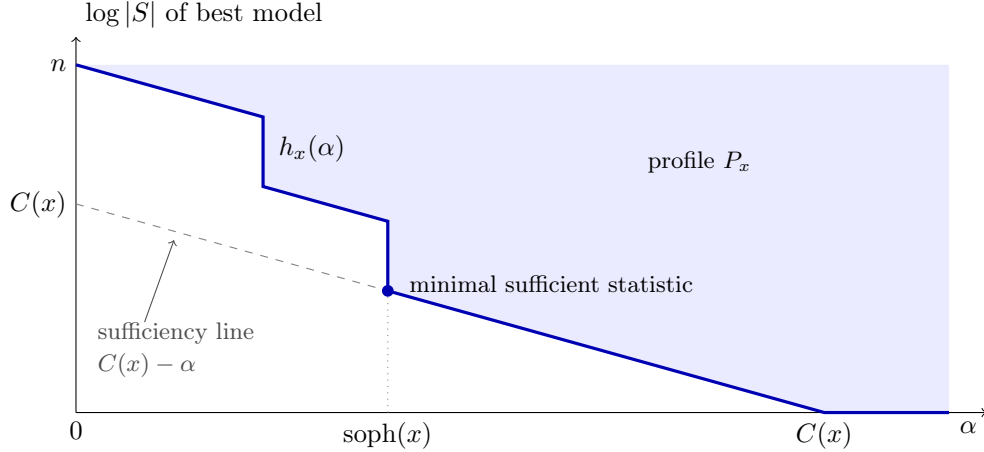


Figure 1: A typical structure function (schematic, after [11]). It lies above the sufficiency line, decreases with slope at most -1 , and touches the line first at the minimal sufficient statistic.

2 Preliminaries

2.1 Structure functions and sophistication

C is plain Kolmogorov complexity and $CT(y \mid x)$ the *total* conditional complexity: the length of a shortest program that maps x to y and halts on every input. Strings have length n . Statements said to hold “up to $O(\log n)$ ” hold up to additive $O(\log n)$ terms in all coordinates; c denotes a sufficiently large constant.

Definition 2.1. The *structure function* of $x \in \{0, 1\}^n$ is $h_x(\alpha) = \min\{\log |S| : x \in S \subseteq \{0, 1\}^n, C(S) \leq \alpha\}$ for $\alpha \geq c \log n$. A model is *sufficient* if $C(S) + \log |S| \leq C(x) + c \log n$, and the *sophistication* $\text{soph}(x) = \min\{\alpha : h_x(\alpha) + \alpha \leq C(x) + c \log n\}$ is the complexity of a minimal sufficient statistic [4, 1, 11]. The string is *stochastic* if $\text{soph}(x) = O(\log n)$, and *antistochastic* if $h_x(\alpha) \approx n - \alpha$ for all $\alpha < C(x)$ [5].

Lemma 2.2 ([11, 8]). For $x \in \{0, 1\}^n$ with $C(x) = \kappa$, up to $O(\log n)$: (i) $h_x(\alpha) \geq \kappa - \alpha$; (ii) $h_x(c \log n) \leq n$ and $h_x(\kappa) = 0$; (iii) (halving law) $h_x(\alpha) + \alpha$ is non-increasing. Conversely [11], every simple function with (i)–(iii) is the structure function of some x up to $O(\log n)$; such functions are called *admissible*.

Figure 1 shows a typical structure function. The curve starts near n , where the only affordable model is essentially $\{0, 1\}^n$, and reaches 0 at $\alpha \approx C(x)$, where $\{x\}$ is affordable. It never lies below the *sufficiency line* $C(x) - \alpha$ (Lemma 2.2(i)), and by the halving law it falls by at least one bit per additional bit of model complexity, possibly in sudden drops. It first meets the sufficiency line at the minimal sufficient statistic, of complexity $\text{soph}(x)$, and follows the line from there on. The region above the curve is the *profile* P_x of all pairs $(\alpha, \log |S|)$ achieved by models $S \ni x$. For a stochastic string the curve drops to the sufficiency line almost at once; for an antistochastic string it follows $n - \alpha$ until $\alpha \approx C(x)$.

2.2 Strong models and partitions

Definition 2.3 ([9, 6]). A model $A \ni x$ is ε -strong if $CT(A \mid x) \leq \varepsilon$, and $h_{x,\varepsilon}^{\text{str}}(\alpha) = \min\{\log |A| : x \in A, CT(A \mid x) \leq \varepsilon, C(A) \leq \alpha\}$. A string is (ε, δ) -normal if $h_{x,\varepsilon}^{\text{str}}(\alpha + \delta) \leq h_x(\alpha) + \delta$ for all α , and *strange* if it has no strong minimal sufficient statistic.

Strong models exclude the universal models of [1], which carry the information of the number of strings of bounded complexity [9]. Strange strings exist [9]; normal strings exist for every admissible shape [6]; antistochastic strings are normal [6]. The following lemma is the bridge to partitions.

Lemma 2.4 (Partition lemma [6]). If $A \ni x$ is ε -strong, there is a partition p of $\{0, 1\}^n$ with $C(p) \leq \varepsilon + O(\log n)$ and a cell $A_1 \ni x$ of p with $A_1 \subseteq A$ and $C(A_1) \leq C(A) + \varepsilon + O(\log n)$.

Proof. Let p be a total program with $p(x) = A$ and $|p| \leq \varepsilon$. Group the strings x' with $x' \in p(x')$ according to the value $p(x')$, and put every other string in a singleton cell. This partition is computable from p and n , and the cell of x is $A_1 = \{x'' \in A : p(x'') = A\}$, computable from A , p and n . $\square$

2.3 Restricted model classes

A family $\mathcal{A}$ of finite sets is *acceptable* [12] (see also [6, 8]) if it is enumerable, contains every $\{0, 1\}^n$, and satisfies the *covering condition*: for a fixed polynomial p , every $A \in \mathcal{A}$ and every $c < |A|$, the n -bit strings of A are covered by at most $p(n)|A|/c$ members of $\mathcal{A}$ of size at most c .

Theorem 2.5 ([12]). *If $\mathcal{A}$ is acceptable, h_x^A obeys the halving law, and for every admissible shape P there is x with h_x and h_x^A both $C(P) + O(\sqrt{n \log n})$ -close to P .*

3 The Galois connection

3.1 Symmetric partitions

Fix an *ambient* finite group G acting on a finite set X , with $G = G_n$ computable from n when $X = \{0, 1\}^n$. Let $L(G)$ be the subgroup lattice and $\text{Par}(X)$ the partition lattice, ordered by refinement ($p \leq q$ if every cell of p lies in a cell of q). Put

$$\text{orb}_G(U) = \{\text{orbits of } U\}, \quad \text{stab}_G(p) = \{g \in G : g(B) = B \text{ for every cell } B \in p\}.$$

Proposition 3.1 ([2]). $\text{orb}_G(U) \leq p \iff U \leq \text{stab}_G(p)$. Hence both maps are monotone, $U \leq \text{stab}_G \text{orb}_G(U)$, $\text{orb}_G \text{stab}_G(p) \leq p$, and the closed subgroups ($U = \text{stab}_G \text{orb}_G U$) correspond bijectively and order-preservingly to the symmetric partitions ($p = \text{orb}_G \text{stab}_G p$), which are exactly the orbit partitions of subgroups of G . Denote their lattice by Par_G .

Proof. If every U -orbit lies in a cell of p , each cell is a union of U -orbits and is U -invariant; conversely, if U preserves every cell its orbits lie in cells. The rest is the general theory of Galois connections. $\square$

The closed subgroup $\text{stab}_G(p)$ is the *largest* subgroup with orbit partition p : the *canonical certificate* of the hypothesis p .

Example 3.2.

- (a) $G = \text{Sym}(X)$: every partition is symmetric, and the canonical certificates are the *Young subgroups* $\prod_{B \in p} \text{Sym}(B)$.
- (b) $G = S_n$ permuting the coordinates of $\{0, 1\}^n$: $\text{orb}(S_n)$ is the partition into weight classes; subgroups give finer “combinatorial” partitions.
- (c) $G = \text{GL}(n, 2)$: symmetric partitions are rigid. If $W \neq \mathbb{F}_2^n$ is a subspace with $\dim W \geq 2$, the partition $\{W \setminus \{0\}\} \cup \{\text{singletons}\}$ is not symmetric: any g preserving it fixes every vector outside $W \setminus \{0\}$, these span $\mathbb{F}_2^n$, so $g = 1$. With a canonical complement D , the orbit partition of $\text{GL}(W) \times \{\text{id}_D\}$ is symmetric; its cells are $\{d\}$ and $d + (W \setminus \{0\})$, $d \in D$. A cell can appear only together with the other orbits of its certificate.

3.2 Canonical cost

$C(p)$ is the complexity of a partition p of $\{0, 1\}^n$ as a finite object, $C(p, B)$ that of p with one of its cells, and $C(U)$ the complexity of a generator list of U .

Proposition 3.3 (Canonical cost). *For $U \leq G$ with $p = \text{orb}_G(U)$ and a cell B of p :*

$$C(\text{stab}_G p) \leq C(p) + O(1) \leq C(U) + O(1), \quad C(\text{stab}_G p, B) = C(p, B) \pm O(1).$$

The canonical certificate is, up to $O(1)$, the cheapest group with the given orbits, and it costs exactly as much as the hypothesis.

Proof. Orbits are computable from generators; $\text{stab}_G(p)$ is computable from p by search in G_n , and n is recoverable from p . $\square$

3.3 The algebra of hypotheses

Since orb_G is a lower adjoint it preserves joins, and stab_G as an upper adjoint preserves meets:

$$\text{orb}_G(\langle U, V \rangle) = \text{orb}_G(U) \vee \text{orb}_G(V), \quad \text{stab}_G(p \wedge q) = \text{stab}_G(p) \cap \text{stab}_G(q).$$

So Par_G is closed under joins in $\text{Par}(X)$: the *join* of two hypotheses is the hypothesis of all their symmetries together, and is coarser. The meet in Par_G is $p \wedge_G q = \text{orb}_G(\text{stab}_G p \cap \text{stab}_G q) \leq p \wedge q$, the hypothesis of their *common* symmetries, which is finer. Both are computable, so $C(p \vee q), C(p \wedge_G q) \leq C(p) + C(q) + O(\log n)$. The cell of x in $p \wedge_G q$ lies inside the intersection of its cells in p and q . Meeting with a second cheap hypothesis is thus the natural way to sharpen an explanation (compare the halving discussion in Section 7).

3.4 Relabelling

G acts by conjugation on $L(G)$ and by translation on Par_G , and the two maps are equivariant: $\text{orb}_G(gUg^{-1}) = g \cdot \text{orb}_G(U)$ and $\text{stab}_G(g \cdot p) = g \text{stab}_G(p) g^{-1}$. The model $(g \cdot p, gB)$ for gx is the model (p, B) for x in new coordinates, and costs at most $C(g) + O(1)$ more.

4 Partition structure functions

Definition 4.1. Let $\mathcal{P}$ assign to each n a family of partitions of $\{0, 1\}^n$. A $\mathcal{P}$ -model for x is a pair (p, B) with $p \in \mathcal{P}$ and $x \in B \in p$. Define

$$\begin{aligned} h_x^{\mathcal{P}}(\alpha) &= \min\{\log |B| : (p, B) \text{ } \mathcal{P}\text{-model}, C(p, B) \leq \alpha\}, \\ h_{x,\varepsilon}^{\mathcal{P}}(\alpha) &= \min\{\log |B| : (p, B) \text{ } \mathcal{P}\text{-model}, C(p) \leq \varepsilon, C(p, B) \leq \alpha\}, \\ \text{soph}^{\mathcal{P}}(x) &= \min\{\alpha : h_x^{\mathcal{P}}(\alpha) + \alpha \leq C(x) + c \log n\}. \end{aligned}$$

$C(p)$ is the cost of the *hypothesis* and $C(B | p)$ that of the *cell*; $C(p, B) = C(p) + C(B | p)$ up to $O(\log n)$. For an ambient group G we write $h_x^G, h_{x,\varepsilon}^G$ and soph^G for $\mathcal{P} = \text{Par}_G$.

Proposition 4.2. Up to $O(1)$ shifts: (i) $h_x \leq h_x^{\mathcal{P}} \leq h_{x,\varepsilon}^{\mathcal{P}}$, and $\mathcal{P} \subseteq \mathcal{P}'$ implies $h_x^{\mathcal{P}'} \leq h_x^{\mathcal{P}}$; (ii) every cell of a partition of complexity at most ε is $(\varepsilon + O(1))$ -strong, so $h_{x,\varepsilon+c}^{\text{str}} \leq h_{x,\varepsilon}^{\mathcal{P}}$; (iii) $\text{soph}(x) \leq \text{soph}^{\mathcal{P}}(x)$ up to $O(\log n)$.

Proof. (i) A cell is a finite set containing x and $C(B) \leq C(p, B) + O(1)$. (ii) A total program knowing p maps x to its cell. $\square$

Groups and partitions. The group formulation defines $h_x^{L(G)}(\alpha)$ as the least $\log |U \cdot x|$ over subgroups $U \leq G$ with $C(U, U \cdot x) \leq \alpha$, and $h_{x,\varepsilon}^{L(G)}$ by adding $C(U) \leq \varepsilon$.

Theorem 4.3 (Dictionary). For every ambient G , up to $O(1)$ shifts in α and ε , $h_x^{L(G)} = h_x^G$ and $h_{x,\varepsilon}^{L(G)} = h_{x,\varepsilon}^G$.

Proof. A pair $(U, U \cdot x)$ gives the Par_G -model $(\text{orb}_G U, U \cdot x)$ at no greater cost. A Par_G -model (p, B) gives $(\text{stab}_G p, B)$, and $\text{stab}_G(p) \cdot x = B$ because p is symmetric. Costs are compared by Proposition 3.3. $\square$

Proposition 4.4 (Burnside form). If (p, B) is a Par_G -model for x with certificate $U = \text{stab}_G(p)$, then $|B| = |U|/|\text{stab}_U(x)| = 1/\Pr_{u \in U}[ux = x]$, so $C(x | p, B) \leq \log |B| + O(1)$. All but a 2^{-t} fraction of $x' \in B$ have $C(x' | p, B) \geq \log |B| - t$.

symmetry	algorithmic statistics
symmetric partition $p \in \text{Par}_G$	hypothesis (model class)
cell $B \ni x$	model
$\log U - \log \text{stab}_U(x) $	$\log B $
$C(p)$ small	strong model
x typical in its cell	small randomness deficiency
$\text{soph}^G(x)$	$\text{soph}(x)$

5 All permutations

Let Sym denote the ambient group $\text{Sym}(\{0, 1\}^n)$. By Example 3.2(a), Par_{Sym} is the lattice of *all* partitions.

Theorem 5.1. *Up to $O(1)$ shifts, $h_x^{\text{Sym}} = h_x$, so $\text{soph}^{\text{Sym}} = \text{soph}$. Up to $O(\varepsilon + \log n)$ shifts, $h_{x,\varepsilon}^{\text{Sym}} = h_{x,\varepsilon}^{\text{str}}$.*

Proof. For $S \ni x$, the partition $\{S\} \cup \{\text{singletons}\}$ costs $C(S) + O(1)$. For the second claim combine Proposition 4.2(ii) with the partition lemma (Lemma 2.4). $\square$

Corollary 5.2. *Up to $O(\varepsilon + \delta + \log n)$, a string is (ε, δ) -normal if and only if cells of partitions of complexity at most ε (equivalently, orbits of groups of complexity at most ε) are as good as arbitrary models at every level. It is strange if and only if no such cell is a minimal sufficient statistic. So there are strings with no sufficient cheap symmetry explanation [9], and every admissible shape is realised by a string whose cheap-symmetry structure function equals its Kolmogorov structure function [6].*

Theorem 5.1 says that unrestricted symmetry explains everything. As a statement about symmetry this is vacuous: in the coordinates of Section 9, the certificates realising arbitrary sets are structurally trivial symmetries whose information lies entirely in their placement (Proposition 9.4).

6 Linear symmetry

Identify $\{0, 1\}^n$ with $\mathbb{F}_2^n$ and take $G = \text{GL}(n, 2)$. Linear maps fix 0, so throughout $x \neq 0$.

6.1 Cells: linearly homogeneous sets

Which sets can be cells of linear-symmetric partitions? A linear map preserves all $\mathbb{F}_2$ -linear dependencies, so a set can be an orbit of a linear group only if its internal dependency structure looks the same from each of its points. We make this precise.

A *relation* in $S \subseteq \mathbb{F}_2^n \setminus \{0\}$ is a subset $T \subseteq S$ with $\sum T = 0$, an XOR dependency among elements of S . For example, in $S = \{e_1, e_2, e_1 + e_2\}$ the whole set is a relation, since $e_1 + e_2 + (e_1 + e_2) = 0$. The family of all relations is the *dependency structure* of S (the binary matroid it represents). A permutation π of S *preserves relations* if $\sum T = 0 \iff \sum \pi(T) = 0$ for all $T \subseteq S$, i.e. if it is an automorphism of the dependency structure.

Definition 6.1. A set $S \subseteq \mathbb{F}_2^n \setminus \{0\}$ is *linearly homogeneous* if the relation-preserving permutations of S act transitively on S : for any $s, s' \in S$ some automorphism of the dependency structure maps s to s' .

The analogy is with vertex-transitive graphs, where every vertex looks alike; here every point looks alike with respect to XOR dependencies instead of edges. Two extreme cases show the range of the notion. A linearly independent set has no relations, so every permutation preserves them and the set is homogeneous. A punctured subspace $W \setminus \{0\}$ has as many relations as possible, but $\text{GL}(W)$ moves any nonzero vector to any other while preserving them, so it too is homogeneous. Homogeneity fails when some points play special roles.

Example 6.2. Linearly homogeneous sets include: $W \setminus \{0\}$ for a subspace W ; affine subspaces $a + W$ with $a \notin W$; linearly independent sets; weight classes and other orbits of coordinate permutation groups; and cosets of multiplicative subgroups of $\mathbb{F}_2^\times$, identifying $\mathbb{F}_2^n$ with $\mathbb{F}_{2^n}$. The set $\{e_1, e_2, e_1 + e_2, e_3\}$ is not homogeneous: e_1 lies in the three-term relation $\{e_1, e_2, e_1 + e_2\}$ while e_3 lies in no relation, so no relation-preserving permutation maps e_1 to e_3 .

The key fact is a small piece of linear algebra: *a permutation of S extends to an invertible linear map if and only if it preserves relations*. This gives the characterisation of cells.

Theorem 6.3. $S \subseteq \mathbb{F}_2^n \setminus \{0\}$ is a cell of some partition in Par_{GL} if and only if it is linearly homogeneous. Moreover such a cell is a cell of a symmetric partition of complexity $C(S) + O(1)$, so h_x^{GL} is Kolmogorov's structure function restricted to linearly homogeneous sets.

Proof. If $U \cdot x = S$, every $u \in U$ restricts to a relation-preserving bijection of S , and these act transitively. Conversely, a relation-preserving π defines a linear bijection $L_\pi(\sum_{s \in T} s) = \sum_{s \in T} \pi(s)$ of $\text{span}(S)$. It is well defined, because $\sum T = \sum T'$ implies $\sum(T \triangle T') = 0$ and hence $\sum \pi(T) = \sum \pi(T')$, and it is injective, because π^{-1} also preserves relations. Extend it by the identity on a canonical complement. The resulting group is computable from S , its orbit through x is S , and its orbit partition costs $C(S) + O(1)$. $\square$

Why this matters: linear symmetry as a restricted model class. Theorem 6.3 places the linear theory inside the general theory of restricted model classes of Section 2.3.

- (1) *The groups disappear.* h_x^{GL} is defined by ranging over subgroups of $\text{GL}(n, 2)$ and their orbits. The theorem identifies it with h_x^A for the concrete, combinatorially defined family $\mathcal{A}_{\text{GL}}$ of linearly homogeneous sets. Questions about linear symmetry become questions about a family of sets, and the tools of Vereshchagin–Vitányi [12] apply directly.
- (2) *The restriction costs nothing extra.* A homogeneous set comes with a certificate computable from it, so the symmetry hypothesis costs no more than the set itself. Every difference between h_x^{GL} and h_x is therefore caused by *which* sets are admissible, not by any overhead of describing groups. This is a different kind of restriction from strong models, where the models are ordinary sets but must be computable from the data.
- (3) *It explains why linear symmetry does not collapse.* For $\text{Sym}(\{0, 1\}^n)$ every set is a cell, and symmetric models are all models (Theorem 5.1). For $\text{GL}(n, 2)$ only homogeneous sets qualify. A string whose good models all have inhomogeneous dependency structure cannot be explained well by linear symmetry. This is the gap measured by the band (Section 7) and realised at its maximum by Theorem 8.2.
- (4) *The class is small.* There are at most 2^{n^4+n} homogeneous subsets of $\mathbb{F}_2^n$, since each is an orbit of one of at most 2^{n^4} subgroups. Arbitrary sets number 2^{2^n} . This size difference is what the counting proof of Theorem 8.2 exploits.
- (5) *The class is not closed under halving.* Half of a homogeneous set is usually not homogeneous, since the two halves need not look alike internally. So the covering condition of acceptability, automatic for all sets and easy for affine subspaces, is a genuine question for $\mathcal{A}_{\text{GL}}$. This is why the halving law for h_x^{GL} is open (Question 7.2).
- (6) *Membership is a finite test.* Deciding whether a proposed set is a legitimate linear-symmetry model requires no search over groups: one checks that its dependency structure is point-transitive.

Since linear maps fix 0, homogeneous sets avoid 0, and $\mathcal{A}_{\text{GL}}$ contains $\mathbb{F}_2^n \setminus \{0\}$ rather than the whole space required in the definition of acceptability. This changes nothing up to $O(1)$.

Single cells are constrained only by homogeneity, but whole symmetric partitions are rigid (Example 3.2(c)). This matters for the strong version $h_{x,\varepsilon}^{\text{GL}}$, which requires a *cheap* symmetric partition containing the cell (Question 10.1). Linear hypotheses also have little room to hide information in placement: a subgroup of $\text{GL}(n, 2)$ is determined by its type, an abstract group with a faithful module, up to n^2 bits (Proposition 9.5).

6.2 Affine hypotheses

Every affine subspace is a cell of a cheap symmetric partition. For $W \setminus \{0\}$ this is Example 3.2(c). For $a + W$ with $a \notin W$, fix a complement D of $W \oplus \langle a \rangle$, and for $w \in W$ let τ_w fix $W \oplus D$ pointwise and send $a \mapsto a + w$. The group $\{\tau_w\} \cong W$ has cells $\{v\}$ for $v \in W \oplus D$ and $a + d + W$ for $d \in D$, at cost $C(W, a) + O(1)$. Writing h_x^{aff} for the structure function over affine subspaces, we get, up to $O(\log n)$,

$$h_x \leq h_x^{\text{GL}} \leq h_x^{\text{aff}}, \quad h_x \leq h_{x,\varepsilon}^{\text{str}} \leq h_{x,\varepsilon}^{\text{GL}}. \quad (6.1)$$

7 The band

Theorem 7.1 (Band). *Let $x \in \mathbb{F}_2^n \setminus \{0\}$ with $C(x) = \kappa$. Up to $O(\log n)$:*

- (i) $\kappa - \alpha \leq h_x(\alpha) \leq h_x^{\text{GL}}(\alpha) \leq h_x^{\text{aff}}(\alpha) \leq n - \alpha$ for $c \log n \leq \alpha \leq \kappa$;
- (ii) $h_x^{\text{GL}}(c \log n) \leq \log(2^n - 1)$ and $h_x^{\text{GL}}(\kappa) = 0$;
- (iii) $0 \leq h_x^{\text{GL}}(\alpha) - h_x(\alpha) \leq n - \kappa$, and $\text{soph}(x) \leq \text{soph}^{\text{GL}}(x) \leq \kappa$;
- (iv) (relabelling) for $g \in \text{GL}(n, 2)$, $h_{gx}^{\text{GL}}(\alpha + C(g) + c) \leq h_x^{\text{GL}}(\alpha)$.

Proof. (i) By Lemma 2.2(i) and (6.1); for the last inequality use the affine subspace of strings agreeing with x in the first j coordinates. (ii) Use the partitions $\{\{0\}, \mathbb{F}_2^n \setminus \{0\}\}$ and the discrete partition. (iii) Follows from (i) and (ii). (iv) By equivariance of the Galois connection. $\square$

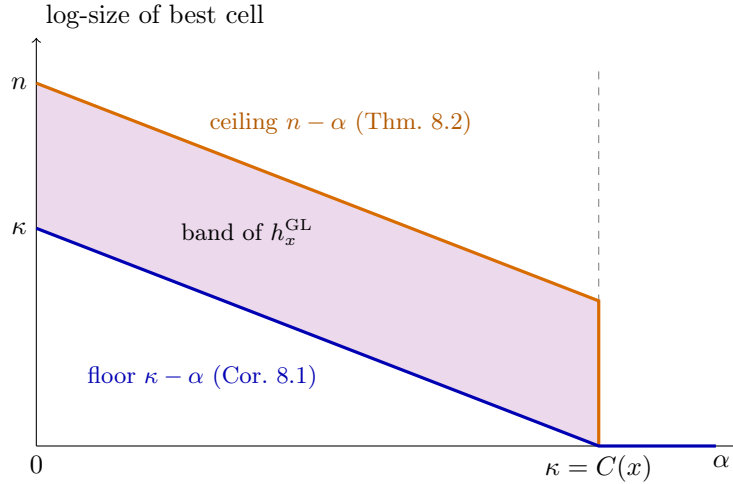


Figure 2: The band containing h_x^{GL} (Theorem 7.1); both edges are attained.

All the structure functions together. Figure 3 places the structure functions of this paper in one picture, for a typical *normal* string x of length n with $\kappa = C(x)$. Normal strings are the common case. A stochastic string has a simple, hence strong, sufficient model, so strangeness forces non-stochasticity, and non-stochastic strings have small a priori probability [10]. The figure shows six curves, of which three coincide for normal strings, together with two guide lines and three marked points. It should be read as follows.

- (1) *The guides.* The lower dashed line is the sufficiency line $\kappa - \alpha$, and the upper one is the ceiling $n - \alpha$. By Lemma 2.2(i) and Theorem 7.1(i), every curve in the figure lies between them for $c \log n \leq \alpha \leq \kappa$, up to $O(\log n)$. A model on the sufficiency line is sufficient; a curve on the ceiling means that the models of that kind are no better than those of a random string of length n .

- (2) *The order of the curves* (proved for every x). By Proposition 4.2, up to $O(1)$ shifts,

$$h_x \leq h_{x,\varepsilon}^{\text{str}} \leq h_{x,\varepsilon}^{\text{GL}}, \quad h_x \leq h_x^{\text{GL}} \leq h_{x,\varepsilon}^{\text{GL}}.$$

The cheap linear-symmetry curve pays for two restrictions at once: its models must be linear-symmetric *and* strong.

- (3) *The endpoints* (proved for every x). At the cheapest budget every curve is at most n : the one-cell partition and the partition $\{\{0\}, \mathbb{F}_2^n \setminus \{0\}\}$ cost $O(\log n)$. At $\alpha \approx \kappa$ every curve vanishes, since the discrete partition is cheap and its cell $\{x\}$ costs κ . All differences between the curves occur for $c \log n \leq \alpha \leq \kappa$.
- (4) *Kolmogorov's curve* h_x (solid blue) has the typical shape of Figure 1. It starts on the ceiling, decreases by at least one bit per bit of budget (the halving law), with two sudden drops, and first meets the sufficiency line at its minimal sufficient statistic (blue dot), of complexity $\text{soph}(x)$. From there it follows the sufficiency line down to $\alpha = \kappa$.
- (5) *The unrestricted symmetric curve* h_x^{Sym} (red dots) lies on h_x for *every* string, not only for normal ones. This is the collapse of Theorem 5.1: every finite set is a cell of a partition symmetric under $\text{Sym}(\{0,1\}^n)$, at no extra cost. In particular $\text{soph}^{\text{Sym}}(x) = \text{soph}(x)$, and the two tick labels at the blue dot coincide. The coincidence is uninformative about symmetry, since by Proposition 9.4 the certificates realising arbitrary models have trivial type.
- (6) *The strong curve* $h_{x,\varepsilon}^{\text{str}}$ (green dashes) also lies on h_x . This is the definition of normality: every point of the profile is attained by an ε -strong model, up to δ . By Theorem 5.1 the same curve is the *cheap* unrestricted symmetric curve $h_{x,\varepsilon}^{\text{Sym}}$. For a strange string it would separate from h_x and lie above it, while h_x^{Sym} would still lie on h_x .
- (7) *The linear-symmetry curve* h_x^{GL} (solid violet) uses cells of symmetric partitions of $\text{GL}(n, 2)$ of any cost, that is, linearly homogeneous sets (Theorem 6.3). In the figure it follows the ceiling up to $\alpha \approx \kappa/2$: at those budgets no linear symmetry sees any structure of x . It then drops, decreases along a segment of slope -1 , drops again, and first meets the sufficiency line at the violet dot, whose complexity is the symmetric sophistication $\text{soph}^{\text{GL}}(x)$. By Theorem 7.1(iii), $\text{soph}(x) \leq \text{soph}^{\text{GL}}(x) \leq \kappa$.
- (8) *The symmetry gap* (shaded) is the region between h_x and h_x^{GL} . At each budget it is the part of the regularity of x that linear symmetry does not see. Its horizontal extent at the sufficiency line, $\text{soph}^{\text{GL}}(x) - \text{soph}(x)$, is the extra budget symmetric explanations need before they become sufficient. By Corollary 8.1 the region can be empty, with $h_x^{\text{GL}} = h_x$. By Theorem 8.2 it can fill the whole band, with h_x^{GL} on the ceiling up to $\alpha \approx \kappa$, even for stochastic normal strings.
- (9) *The cheap linear-symmetry curve* $h_{x,\varepsilon}^{\text{GL}}$ (orange dashes) uses only cells of symmetric partitions of complexity at most ε . In the figure it agrees with h_x^{GL} on the ceiling, where the cylinder partitions fixing the first j coordinates are cheap, and lies above it afterwards. There the good linear-symmetric cells of x exist only inside expensive symmetric partitions: the rigidity effect of Question 10.1. It meets the sufficiency line only at the right end, just before $\alpha = \kappa$. Because x is normal, this separation from h_x^{GL} is caused by symmetry, not by strangeness.
- (10) *The marked points*. On the horizontal axis: $\text{soph}(x)$, equal to $\text{soph}^{\text{Sym}}(x)$, where h_x and h_x^{Sym} become sufficient; $\text{soph}^{\text{GL}}(x)$, where h_x^{GL} becomes sufficient; and κ , where all curves vanish. On the vertical axis: n , where all curves start, and κ , where the sufficiency line starts.

Items (1)–(6) and the inequality in (7) are theorems; item (10) only names the marked points. The particular shapes of h_x^{GL} and $h_{x,\varepsilon}^{\text{GL}}$ are *illustrative*. Which shape is typical is open, and so is whether the rigidity separation in (9) occurs at all. Whether the GL curves obey the halving law (Question 7.2) is also open. If they do not, they may have plateaus, which h_x never has.

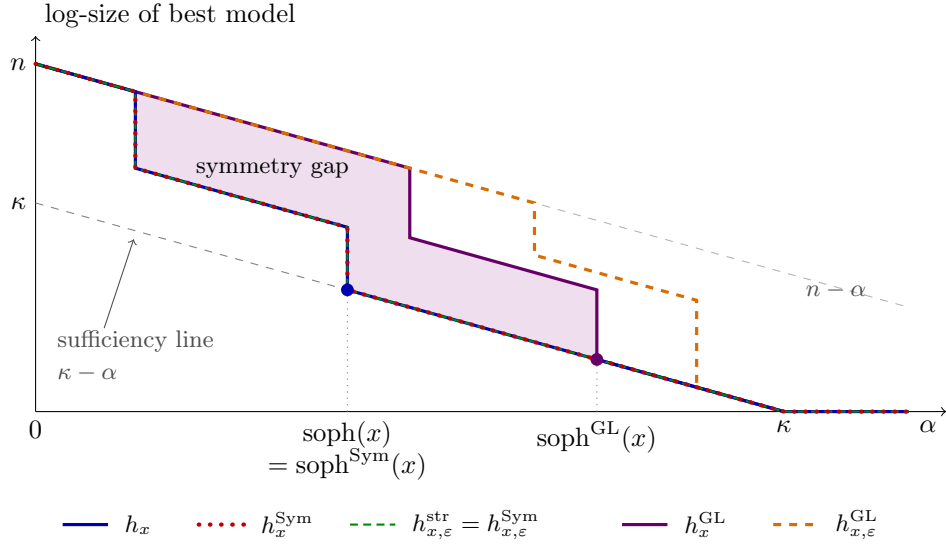


Figure 3: The structure functions for a typical normal string. The curves h_x , h_x^{Sym} and $h_{x,\varepsilon}^{\text{str}} = h_{x,\varepsilon}^{\text{Sym}}$ coincide; the first two coincide for every string. The shaded region is the symmetry gap. The ordering, the guides, the endpoints and the coincidences are proved; the shapes of the two GL curves are illustrative. See items (1)–(10) in the text.

Halving as refinement. Affine subspaces form an acceptable class: a coset of dimension d splits into at most $2|A|/c$ cosets of size at most c . By Theorem 2.5, h_x^{aff} obeys the halving law. For h_x^{GL} the question becomes one about the lattice Par_{GL} . Given a symmetric p with $x \in B \in p$, is there a cheap symmetric refinement whose cell containing x has size about $|B|/2$? The algebra of Section 3 suggests meeting p with a second cheap hypothesis q : the cell of x in $p \wedge_{\text{GL}} q$ lies in the intersection of its cells. Transitivity alone does not obstruct this. $\text{GL}(n, 2)$ is 2-transitive on nonzero vectors, so no subgroup between $\text{stab}(x)$ and $\text{GL}(n, 2)$ refines the orbit of x , but subgroups not containing $\text{stab}(x)$ can.

Question 7.2. *Is the family of cells of Par_{GL} acceptable, i.e. can every cell of size N be covered by $\text{poly}(n)N/c$ cells of size at most c ? Equivalently for the halving law: does every cell admit cheap symmetric refinements, for instance by meets? If not, does h_x^{GL} exhibit a plateau for some x ?*

8 The extremes of the band

8.1 Zero gap: symmetry explains everything

Corollary 8.1. *For every admissible shape P there is a string x with $h_x \approx h_x^{\text{GL}} \approx h_x^{\text{aff}} \approx P$, up to $C(P) + O(\sqrt{n \log n})$, so $\text{soph}^{\text{GL}}(x) = \text{soph}(x)$ to the same accuracy.*

Proof. Theorem 2.5 applied to affine subspaces, and (6.1). □

8.2 Maximal gap: structure invisible to symmetry

Theorem 8.2 (Maximal gap). *There is a constant c such that for all n , all $c \log n \leq m \leq n$ and all $t \geq 0$ there are strings $x \in \{0, 1\}^n$ with:*

- (a) $C(x) = m \pm O(t + \log n)$ and $h_x(\alpha) = m - \alpha \pm O(t + \log n)$ for $c \log n \leq \alpha \leq m$; x is stochastic and $(O(\log n), O(t + \log n))$ -normal;
- (b) $h_x^{\text{GL}}(\alpha) \geq n - \alpha - t - O(\log n)$ for all $\alpha \leq m - t - c \log n$.

In fact all but a 2^{-t} fraction of a set S with $C(S) = O(\log n)$ and $|S| \approx 2^m$ have these properties.

Proof. Counting cells. By Proposition 9.5, or simply because a group of order N has a generating set of size $\log N$, there are at most 2^{n^4} subgroups of $\text{GL}(n, 2)$, hence at most 2^{n^4+n} cells of symmetric partitions.

A set that no cell sees. Put each string into a random S independently with probability 2^{m-n} . For a cell B let $R_B = \max(2e 2^{m-n}|B|, n^5)$. The Chernoff bound $\Pr[X \geq R] \leq 2^{-R}$ for $R \geq 2e \mathbb{E}X$ [7] gives $\Pr[|S \cap B| \geq R_B] \leq 2^{-n^5}$. A union bound, and concentration of $|S|$, show that with positive probability

$$2^{m-1} \leq |S| \leq 2^{m+1} \quad \text{and} \quad |S \cap B| < \max(2e 2^{m-n}|B|, n^5) \text{ for every cell } B. \quad (8.1)$$

This property is decidable, so the lexicographically first such S has $C(S) = O(\log n)$.

Typical elements. All but a 2^{-t} fraction of $x \in S$ have $C(x) \geq m - t$; fix one.

(a) Cutting S into 2^j consecutive pieces gives models of complexity $j + O(\log n)$ and size at most 2^{m+1-j} , each computable from x by a total program of length $O(\log n)$. With Lemma 2.2(i) this gives h_x and normality.

(b) Fix α, s . Fewer than $2^{\alpha+1}$ models (p, B) have $C(p, B) \leq \alpha$, and the union U of those cells with $|B| \leq 2^s$ is enumerable from n, m, α, s . By (8.1), $|U \cap S| \leq 2^{\alpha+1} \max(2e 2^{m-n+s}, n^5)$. If $x \in U$, then

$$m - t \leq C(x) \leq \log |U \cap S| + O(\log n) \leq \alpha + \max(m - n + s, 5 \log n) + O(\log n).$$

For $\alpha \leq m - t - c \log n$ the second branch is impossible, so $s \geq n - \alpha - t - O(\log n)$. $\square$

Corollary 8.3 (Simple structure, invisible to symmetry). *If moreover $m \leq n - t - c' \log n$, these strings satisfy $\text{soph}(x) = O(\log n)$ but $\text{soph}^{\text{GL}}(x) \geq C(x) - O(t + \log n)$. They are stochastic and normal, yet no cell of a symmetric partition costing less than $C(x) - O(t + \log n)$ is sufficient: every sufficient symmetric model essentially names x itself. The gap between the two chains in (6.1) is a symmetry phenomenon, not a computability one.*

Proof. By (b), $h_x^{\text{GL}}(\alpha) + \alpha \geq n - t - O(\log n) > C(x) + c \log n$ for $\alpha \leq m - t - c \log n$. $\square$

Remark. The argument only uses that the partition family has $2^{\text{poly}(n)}$ cells and is decidable. So every such family, for instance Par_G for any ambient group of order $2^{\text{poly}(n)}$, has stochastic normal strings that it sees as antistochastic. The construction is not explicit.

9 Coordinates on the space of permutation groups

The Galois connection organises symmetry hypotheses into a lattice. To move through that lattice, and to search it, we need coordinates. This section shows that every permutation group is a Burnside ring element (its *type*) together with a permutation (its *placement*), compares the two ambient groups in these terms, and discusses what the maximal gap theorem implies for searching the coordinates for good models.

9.1 The coordinate theorem

We give coordinates on the set of all permutation groups on a finite set X , $|X| = N$; in our setting $X = \{0, 1\}^n$ and $N = 2^n$. Via the Galois connection of Section 3 these become coordinates on symmetric partitions, and they are used to navigate both spaces. To keep G for the ambient group, we write H for the abstract group.

A subgroup $U \leq S_X$ is the image of a faithful action $\rho: H \rightarrow S_X$ of an abstract group $H \cong U$. Up to isomorphism of H -sets, the action is determined by its class in the Burnside ring $B(H)$. Fix representatives $K_1, \dots, K_r$ of the conjugacy classes of subgroups of H . Then

$$[X]_\rho = \sum_{i=1}^r a_i [H/K_i], \quad a_i \geq 0, \quad \deg[X]_\rho := \sum_i a_i |H : K_i| = N.$$

The action is faithful iff $\bigcap_{a_i > 0} \text{core}_H(K_i) = 1$, where $\text{core}_H(K) = \bigcap_h hKh^{-1}$ is the kernel of H on H/K . By Burnside's theorem, $b \in B(H)$ is determined by its *marks* $\varphi_K(b) = |X_b^K|$, the number of points fixed by K .

Definition 9.1 (Canonical realisation). Let $B_N(H)$ be the set of faithful $b = \sum a_i [H/K_i]$ of degree N . Fix H concretely (with a total order on its elements) and the representatives K_i canonically. For $b \in B_N(H)$ let X_b be the disjoint union of a_i copies of H/K_i , ordered by i , then by copy, then by the least element of each coset. Identify X_b with X in lexicographic order. This gives a faithful action ρ_b and a *canonical subgroup* $U_b = \rho_b(H) \leq S_X$.

Theorem 9.2 (Coordinates). Let $\mathcal{S}_H(X) = \{U \leq S_X : U \cong H\}$ and

$$\Phi_H : B_N(H) \times S_X \rightarrow \mathcal{S}_H(X), \quad \Phi_H(b, \sigma) = \sigma U_b \sigma^{-1}.$$

- (i) Φ_H is surjective: every $U \cong H$ has coordinates (b, σ) .
- (ii) For fixed b , $\Phi_H(b, \sigma) = \Phi_H(b, \sigma')$ iff $\sigma^{-1}\sigma' \in N_{S_X}(U_b)$.
- (iii) $\Phi_H(b, \sigma)$ and $\Phi_H(b', \sigma')$ are conjugate in S_X iff $b' \in \text{Aut}(H) \cdot b$, where $\alpha \in \text{Aut}(H)$ acts by $[H/K] \mapsto [H/\alpha(K)]$.

Consequently

$$\mathcal{S}_H(X) = \bigsqcup_{[b] \in B_N(H)/\text{Aut}(H)} S_X/N_{S_X}(U_b) :$$

the type $[b]$ is a discrete coordinate labelling conjugacy classes, and the placement σ is a coordinate on the homogeneous space $S_X/N_{S_X}(U_b)$. The whole space of permutation groups on X is the disjoint union of these pieces over the isomorphism classes of H with a faithful action of degree N .

Proof. (i) Fix an isomorphism $\psi : H \rightarrow U$. The H -set X with action ψ is faithful of some class $b \in B_N(H)$, so there is an isomorphism of H -sets $\sigma : (X, \rho_b) \rightarrow (X, \psi)$, i.e. $\sigma \rho_b(h) \sigma^{-1} = \psi(h)$ for all h . Hence $U = \sigma U_b \sigma^{-1}$. (ii) This is the definition of the normaliser. (iii) If $\tau U_b \tau^{-1} = U_{b'}$, then $\alpha = \rho_{b'}^{-1} \circ c_\tau \circ \rho_b$ is an automorphism of H , and τ is an isomorphism of H -sets from X_b to $X_{b'}$ twisted by α , so $b' = \alpha \cdot b$. The converse reverses the argument. $\square$

Partitions and cells in coordinates. The orbits of U_b are the summands of X_b , which are consecutive intervals of X . Let $\lambda(b)$ be the integer partition of N with part $|H : K_i|$ repeated a_i times, and $p_{\lambda(b)}$ the corresponding partition of X into consecutive intervals. Then

$$\text{orb}(\Phi_H(b, \sigma)) = \sigma \cdot p_{\lambda(b)}.$$

So partitions have coordinates (λ, σ) , with σ defined modulo the stabiliser of p_λ in S_X . The projection $(b, \sigma) \mapsto (\lambda(b), \sigma)$ is orb written in coordinates: it forgets the structure of b beyond its cell sizes. If $\sigma^{-1}(x)$ lies in a summand of type H/K_i , then the cell of x has size $|H : K_i|$, its stabiliser is conjugate to K_i , and

$$\Pr_{u \in U}[ux = x] = \frac{|K_i|}{|H|}.$$

The marks $\varphi_K(b)$ give the whole distribution of cell sizes and fixed points over X ; Proposition 4.4 is the pointwise case.

Proposition 9.3 (Complexity in coordinates). Let $U = \Phi_H(b, \sigma)$, with b chosen canonically in its $\text{Aut}(H)$ -orbit. Then, up to $O(\log C(U))$,

$$C(U) = C(H, b) + \min\{C(\sigma' \mid H, b) : \Phi_H(b, \sigma') = U\}.$$

Proof. From U one computes H up to isomorphism, the canonical b , and the least σ in the coset $\sigma N_{S_X}(U_b)$. This least element is computable from any member of the coset together with (H, b) . Apply symmetry of information. $\square$

Navigation. Two basic moves act on coordinates.

- *Relabelling:* $(b, \sigma) \mapsto (b, \pi\sigma)$ conjugates the group by π and translates its partition, at cost at most $C(\pi)$.
- *Restriction* to a subgroup $L \leq H$: the subgroup $\sigma\rho_b(L)\sigma^{-1}$ has coordinates $(\text{Res}_L^H b, \sigma\tau)$ in $\mathcal{S}_L(X)$, where τ is computable from (H, L, b) . Its partition refines that of U cell by cell, according to the Mackey formula

$$\text{Res}_L^H [H/K] = \sum_{LhK \in L \backslash H/K} [L/(L \cap hKh^{-1})],$$

so a cell of size $|H : K|$ splits into cells of sizes $|L : L \cap hKh^{-1}|$.

Restriction moves down the subgroup lattice and refines hypotheses; relabelling moves within a conjugacy class. In coordinates, the halving question of Section 7 asks for a cheap subgroup $L \leq H$ whose Mackey decomposition splits the cell of x roughly in half.

Remark. For subgroups of $\text{GL}(n, 2)$ the same construction applies with $B(H)$ replaced by the isomorphism classes of faithful $\mathbb{F}_2 H$ -modules of dimension n (the Green ring), and S_X by $\text{GL}(n, 2)$. The fibres are then $\text{GL}(n, 2)/N_{\text{GL}(n, 2)}(U_b)$, and the placement costs at most n^2 bits (Proposition 9.5).

9.2 Types and placements: S_X versus $\text{GL}(n, 2)$

The coordinates separate what a symmetry hypothesis *is* (its type) from *where it sits* (its placement). The two ambient groups behave very differently.

Proposition 9.4 (Placement dominates). *Let $U_S = \text{Sym}(S) \times 1$ be the Young subgroup of a set $S \subseteq X$ with $|S| = k \geq 2$. In the coordinates of Theorem 9.2, its type is $H = S_k$, $b = [S_k/S_{k-1}] + (2^n - k)[S_k/S_k]$, and costs $O(n)$, while its placement costs*

$$\min\{C(\sigma \mid H, b) : \Phi_H(b, \sigma) = U_S\} = C(S) \pm O(n).$$

Proof. S is the unique cell of size greater than one, so $C(U_S) = C(S) \pm O(1)$. The type is described by k and n . Since $C(U_S) \leq n2^n + O(1)$, the error term in Proposition 9.3 is $O(n)$. $\square$

So the explanations that realise arbitrary Kolmogorov models are *structurally trivial symmetries placed in a complicated way*. A symmetry explanation is informative only when its information lies in the type. The placement for Sym can cost up to $\log(2^n!) \approx n2^n$ bits, and this is what makes the collapse possible.

Proposition 9.5 (Type determines the hypothesis up to n^2 bits). *Let $U \leq \text{GL}(n, 2)$ have coordinates (b, g) as in the remark at the end of Section 9.1: b is the class of a faithful $\mathbb{F}_2 H$ -module of dimension n and $g \in \text{GL}(n, 2)$ is a placement. Then, up to $O(\log n)$,*

$$C(H, b) \leq C(U) \leq C(H, b) + n^2.$$

Consequently the number of symmetric partitions in Par_{GL} is at most the number of types times $|\text{GL}(n, 2)| < 2^{n^2}$.

Proof. The type is computable from U . Conversely $U = gU_b g^{-1}$, where U_b is computable from the type and g is an $n \times n$ matrix. $\square$

Contrast Proposition 9.4: in S_X the placement can carry almost all of the information, whereas in $\text{GL}(n, 2)$ it carries at most n^2 bits. A linear symmetry hypothesis is essentially an abstract group with a module, and it cannot hide arbitrary information in its placement.

9.3 What the maximal gap theorem means for search

The coordinates make the search for symmetric models concrete. A hypothesis is a type (H, b) together with a placement, and one moves through the space by relabelling, by restriction with its Mackey splitting of cells, and by meets with other cheap hypotheses. It is natural to try to find good models of a given string x this way, by exhaustive, heuristic or learned search. Theorem 8.2 sets the limits of any such procedure.

Failure can be forced by the class, not the search. For the strings of Theorem 8.2, no cell of a linear-symmetric partition costing much less than $C(x)$ is a good model. No search strategy over Par_{GL} , however it navigates the coordinates, can find one, because none exists. Conversely, a model found by search only gives an *upper* bound on h_x^{GL} . Since structure functions are not computable, no search yields matching lower bounds.

Failure says little about x . The strings of Theorem 8.2 are stochastic and normal. Their sufficient statistics are simple, and even computable from x by a short total program (Corollary 8.3). So “no good symmetric model found” means “no linear-symmetric structure found”. It is no evidence that x is random, nor that x is strange.

Permutation-group coordinates are complete but uninformative. For the ambient group S_X , search is complete in principle: by Theorem 5.1, every Kolmogorov model is a cell of a symmetric partition. But by Proposition 9.4, the certificates that realise arbitrary models have trivial type, so the search would take place almost entirely among placements. The homogeneous spaces $S_X/N_{S_X}(U_b)$ can have size close to $(2^n)!$, and searching them is searching arbitrary finite sets under another name. The group structure helps only for models whose information lies mainly in the type.

Matrix-group coordinates are searchable but blind. For $\text{GL}(n, 2)$, the placement costs at most n^2 bits (Proposition 9.5). The effective search space is therefore the discrete set of types, small abstract groups with faithful modules, times at most 2^{n^2} placements: altogether $2^{\text{poly}(n)}$ hypotheses. This finiteness is what makes the search feasible, and it is exactly what the proof of Theorem 8.2 exploits: any family of $2^{\text{poly}(n)}$ decidable cells is blind to some stochastic normal strings. The same holds for $\text{AGL}(n, 2)$, for coordinate permutations, and for any finite combination of such classes (the remark after Corollary 8.3). In short:

A coordinate system in which the space of symmetry hypotheses is small enough to search is small enough to miss simple structure.

Completeness requires the full group S_X , where the hypotheses stop being meaningfully symmetric.

What search is good for. None of this makes symmetric search useless. Theorem 8.2 is non-constructive. It shows that blind strings exist, not how often they occur in data of interest, and if such data tend to have symmetric regularities, the search target exists (Corollary 8.1) and finding it is a genuine discovery. The theorem changes what a search should report. It should not report “the structure of x ”. It should report *how much of the structure of x is symmetric*: the best model found, with its deficiency $C(p, B) + \log |B| - C(x)$, where $C(x)$ is estimated by a compressor. A deficiency near the ceiling of Theorem 7.1 means that x looks random to linear symmetry. Four practical consequences follow.

- (a) *Pair symmetric search with strong-model search.* Blind strings can be normal, so a search over all simple partitions can succeed where the symmetric search fails. The difference between the two results estimates $\text{soph}^{\text{GL}}(x) - \text{soph}(x)$.
- (b) *Search types first.* For $\text{GL}(n, 2)$, enumerate small types (H, b) and only then optimise placements within each type.
- (c) *Refine by restriction and meets.* Mackey splitting and meets are the natural moves for halving the cell of x (Question 7.2).
- (d) *Treat stalls as possibly genuine.* If h_x^{GL} violates the halving law, a search that stalls at some cell size may have reached a real plateau of h_x^{GL} , not a local optimum of the algorithm.

Strangeness is rare, since strange strings are non-stochastic and non-stochastic strings have small a priori probability [10]. Blindness to symmetry is plausibly common among stochastic strings, because a symmetric family has only $2^{\text{poly}(n)}$ cells. So in practice the gap between symmetric and arbitrary models, not the gap between strong and arbitrary ones, is the dominant limitation.

10 Open problems

- (1) *Halving as refinement.* Resolve Question 7.2: does Par_{GL} admit cheap refinements of every cell, for instance by meets with cheap hypotheses, or by restriction moves whose Mackey decomposition halves the cell of x (Section 9)?
- (2) *Navigation.* Use the coordinates (b, σ) and the moves of restriction and relabelling to search the space of symmetric partitions for good models of a given string. Which moves are cheap, and how does the structure function change along them?
- (3) *Rigidity.*

Question 10.1. *Is there a string x for which h_x^{GL} and $h_{x,\varepsilon}^{\text{str}}$ are both small at some level, yet $h_{x,\varepsilon}^{\text{GL}}$ is large there? Such an x would have a good linearly homogeneous model and a good strong model, but no cheap symmetric partition, a separation caused by the rigidity of Par_{GL} rather than by computability.*

- (4) *Type-dominated explanations.* Charge the type and the placement separately, and define structure functions in which the placement cost is bounded. Which strings have sufficient symmetric explanations whose information lies mainly in the type? For Sym, Proposition 9.4 shows that arbitrary models need placement. For $\text{GL}(n, 2)$, how do the at most n^2 bits of placement trade against the type in optimal models?
- (5) *Counting symmetric partitions.* Bound the number of $p \in \text{Par}_{\text{GL}}$ with $C(p) \leq \varepsilon$ and given cell sizes, using types (Burnside and Green rings) and placements. This would sharpen Theorem 8.2, whose counting ranges over all subgroups.
- (6) *Realizability.* Which pairs (h_x, h_x^{GL}) occur? Conjecture: every pair (P, Q) with P admissible, Q non-increasing, $P \leq Q \leq n - \alpha$ and Q vanishing where P does, up to $o(n)$; in particular every value of $\text{soph}^{\text{GL}}(x)$ between $\text{soph}(x)$ and $C(x)$.
- (7) *Explicit extremal strings.* Find explicit stochastic strings with $\text{soph}^{\text{GL}}(x) \approx C(x)$. Candidates are $(y, f(y))$ for simple f whose graph has neither large Walsh coefficients nor large linear symmetry groups.
- (8) *Canonicity and heredity.* Is a sufficient cell of a cheap symmetric partition canonical, in the sense of Vereshchagin's theorem for strong minimal sufficient statistics [9]? Is the code of an optimal symmetric model of a string with $h_x^{\text{GL}} \approx h_x$ again of this kind, as for normal strings [6]?
- (9) *Other ambient groups.* Develop the lattices Par_G for the affine group $\text{AGL}(n, 2)$, for coordinate permutations S_n , and for groups of polynomial automorphisms, and determine which regularities each sees.

References

- [1] P. Gács, J. Tromp and P. M. B. Vitányi. Algorithmic statistics. *IEEE Transactions on Information Theory*, 47(6):2443–2463, 2001.
- [2] A. Kerber. *Applied Finite Group Actions*. Springer, 2nd edition, 1999.
- [3] A. N. Kolmogorov. Talk at the Information Theory Symposium, Tallinn, 1974.
- [4] M. Koppel. Complexity, depth, and sophistication. *Complex Systems*, 1, 1987.
- [5] A. Milovanov. Some properties of antistochastic strings. In *Computer Science – Theory and Applications (CSR 2015)*, LNCS 9139. Springer, 2015.

- [6] A. Milovanov. Algorithmic statistics: normal objects and universal models. In *Computer Science – Theory and Applications (CSR 2016)*, LNCS 9691. Springer, 2016. arXiv:1512.04510.
- [7] M. Mitzenmacher and E. Upfal. *Probability and Computing*. Cambridge University Press, 2005.
- [8] A. Shen, V. A. Uspensky and N. Vereshchagin. *Kolmogorov Complexity and Algorithmic Randomness*. American Mathematical Society, 2017.
- [9] N. Vereshchagin. Algorithmic minimal sufficient statistics: a new approach. *Theory of Computing Systems*, 56(2), 2015.
- [10] N. Vereshchagin and A. Shen. Algorithmic statistics: forty years later. arXiv:1607.08077, 2016.
- [11] N. K. Vereshchagin and P. M. B. Vitányi. Kolmogorov’s structure functions and model selection. *IEEE Transactions on Information Theory*, 50(12):3265–3290, 2004.
- [12] N. K. Vereshchagin and P. M. B. Vitányi. Rate distortion and denoising of individual data using Kolmogorov complexity. *IEEE Transactions on Information Theory*, 56(7):3438–3454, 2010.