Tropical Limits of Probability Spaces, Part I

TROPICAL LIMITS OF PROBABILITY SPACES, PART I

THE INTRINSIC KOLMOGOROV-SINAI DISTANCE AND THE ASYMPTOTIC EQUIPARTITION PROPERTY FOR CONFIGURATIONS

R. MATVEEV AND J. W. PORTEGIES

Abstract. The entropy of a finite probability space X measures the observable cardinality of large independent products $X^{\otimes n}$ of the probability space. If two probability spaces X and Y have the same entropy, there is an almost measure-preserving bijection between large parts of $X^{\otimes n}$ and $Y^{\otimes n}$. In this way, X and Y are asymptotically equivalent.

It turns out to be challenging to generalize this notion of asymptotic equivalence to configurations of probability spaces, which are collections of probability spaces with measure-preserving maps between some of them.

In this article we introduce the intrinsic Kolmogorov-Sinai distance on the space of configurations of probability spaces. Concentrating on the large-scale geometry we pass to the asymptotic Kolmogorov-Sinai distance. It induces an asymptotic equivalence relation on sequences of configurations of probability spaces. We will call the equivalence classes *tropical probability spaces*.

In this context we prove an Asymptotic Equipartition Property for configurations. It states that tropical configurations can always be approximated by homogeneous configurations. In addition, we show that the solutions to certain Information-Optimization problems are Lipschitz-continuous with respect to the asymptotic Kolmogorov-Sinai distance. It follows from these two statements that in order to solve an Information-Optimization problem, it suffices to consider homogeneous configurations.

Finally, we show that spaces of trajectories of length n of certain stochastic processes, in particular stationary Markov chains, have a tropical limit.

0. Introduction

The aim of the present article is to develop a theory of tropical probability spaces, which are asymptotic classes of finite probability spaces. Together with the accompanying techniques, we expect them to be relevant to problems arising in information theory, causal inference, artificial intelligence and neuroscience.

As a matter of introduction and motivation of the research presented in the article, we start by considering a few simple examples.

0.1. **Single probability spaces.** We consider a finite probability space $X = (S, p)$, where S is a finite set, and p is a probability measure on S. For simplicity, assume for now that the measure p has full support. Next, we consider the, so-called, Bernoulli sequence of probability spaces

$$X^{\otimes n} = (S^n, p^{\otimes n})$$

where S^n denotes the n-fold Cartesian product of S, and $p^{\otimes n}$ is the n-fold product measure.

This situation arises in several contexts. For example, in physics, $X^{\otimes n}$ would encode the state of the system comprised of many identical non-interacting (weakly interacting) subsystems with state space X. In information theory, $X^{\otimes n}$ would describe the output of an i.i.d. random source. In dynamical systems or stochastic processes the setting corresponds to Bernoulli shifts and Bernoulli processes.

The *entropy* of X is the exponential growth rate of the *observable* cardinality of tensor powers of X. The observable cardinality, loosely speaking, is the cardinality of the set $X^{\otimes n}$ after (the biggest possible) set of small measure of elements, each with negligible measure, has been removed. It turns out that the observable cardinality of $X^{\otimes n}$ might be much smaller than $|S|^n$, the cardinality of the whole of $X^{\otimes n}$, in the following sense.

The *Asymptotic Equipartition Property* states that for every $\varepsilon > 0$ and sufficiently large n one can find a, so-called, *typical subset* $A_\varepsilon^{(n)} \subset S^n$, such that it takes up almost all of the mass of $X^{\otimes n}$ and the probability distribution on $A_\varepsilon^{(n)}$ is almost uniform on the normalized logarithmic scale,

(i) $p^{\otimes n}(A_\varepsilon^{(n)}) \geq 1 - \varepsilon$
(ii) For any $a, a' \in A_\varepsilon^{(n)}$ holds $\left|\frac{1}{n}\ln p(a) - \frac{1}{n}\ln p(a')\right| \leq \varepsilon$

The cardinality $|A_\varepsilon^{(n)}|$ may be much smaller than $|S|^n$, but it will still grow exponentially with n. Even though there are many choices for such a set $A_\varepsilon^{(n)}$, the exponential growth rate with respect to n is well-defined upto 2ε. In fact, there exists a number h_X such that for any choice of the typical subset $A_\varepsilon^{(n)}$ holds

$$\mathbf{e}^{n\cdot h_X - \varepsilon} \leq |A_\varepsilon^{(n)}| \leq \mathbf{e}^{n\cdot h_X + \varepsilon}$$

The limit of the growth rate as $\varepsilon \to 0+$ is called the entropy of X, as explained in more detail in Section 1.3

$$\mathcal{E}nt(X) := \lim_{\varepsilon\downarrow 0}\lim_{n\to\infty}\frac{1}{n}\ln|A_\varepsilon^{(n)}|$$

By the law of large numbers

$$\mathcal{E}nt(X) = h_X = -\sum_{x\in S} p(x)\ln p(x)$$

which is the formula by which the Shannon entropy is usually introduced.

Entropy is especially easy to evaluate if the space is uniform, since for any finite probability space with the uniform distribution holds

$$\mathcal{E}nt(X) = \ln|X| \tag{0.1}$$

This point of view on entropy goes back to the original idea of Boltzmann, according to which entropy is the logarithm of the number of equiprobable states, that a system, comprised of many identical weakly interacting subsystems, may take on.

0.1.1. *Asymptotic equivalence.* The Asymptotic Equipartion Property implies that the sequence $X^{\otimes n}$ is asymptotically equivalent to a sequence of uniform spaces in the following sense. Let us denote by p_U the probability distribution that is supported on $A_\varepsilon^{(n)}$, and is uniform on its support. Then a sequence from independent samples according to p_U is very hard to discriminate from a sequence of independent samples from $p_X^{\otimes n}$.

Similarly, when X and Y are probability spaces with the same entropy, the sets X and Y are asymptotically equivalent in the sense that there is a bijection between the typical sets, which can be seen as a change of code. This is essentially the content of Shannon's source coding theorem.

In [Gro12], Gromov proposed this existence of an "almost-bijection" as a basis of an asymptotic equivalence relation on sequences of probability spaces. Even though we were greatly influenced by ideas in [Gro12], we found that Gromov's definition does not extend easily to **configurations** of probability spaces.

By a configuration of probability spaces we mean a collection of probability spaces with measure-preserving maps between some of them. We will give a precise definition in Section 1.2, but will consider some particular examples below. Formalizing and studying a notion of asymptotic equivalence for configurations of probability spaces is the main topic of the present article.

0.2. **Configurations of probability spaces.** Suppose that now instead of a single probability space, we consider a pair of probability spaces $X = (\underline{X}, p_X)$ and $Y = (\underline{Y}, p_Y)$ with a joint distribution, that is a probability measure on $\underline{X} \times \underline{Y}$ that pushes forward to p_X and p_Y under coordinate projections. In other words, we consider a triple of probability spaces X, Y and U with a pair of measure-preserving maps $U \to X$ and $U \to Y$. This is what we later call a minimal two-fan of probability spaces

$$\begin{tikzcd} & U \arrow[dl] \arrow[dr] & \\ X & & Y \end{tikzcd}$$

and is a particular instance of a configuration of probability spaces.

0.2.1. *Three examples.* Three examples of such an object are shown on Figure 1, which is to be interpreted in the following way. Each of the spaces X_i and Y_i, $i = 1, 2, 3$, have cardinality six and a uniform distribution, where the weight of each atom is $\frac{1}{6}$. The spaces U_i, $i = 1, 2, 3$, have cardinality 12 and the distribution is also uniform with all weights being $\frac{1}{12}$. The support of the measure on U_i's is colored grey on the pictures. The maps from U_i to X_i and Y_i are coordinate projections.

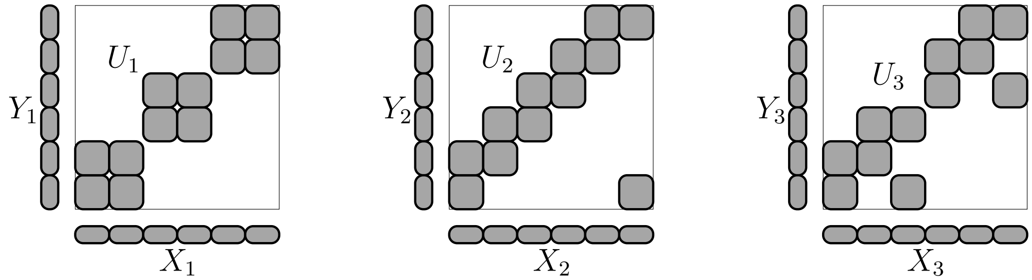

FIGURE 1. Examples of pairs of probability spaces together with a joint distribution.

In view of equation (0.1) we have for each $i = 1, 2, 3$,

$$\begin{aligned} \mathcal{E}nt(X_i) &= \ln 6 \\ \mathcal{E}nt(Y_i) &= \ln 6 \\ \mathcal{E}nt(U_i) &= \ln 12 \end{aligned}$$

Now we would like to ask the following.

Question.

> *Is it possible to find an almost-bijection between sufficiently high powers of $(X_i^{\otimes n} \leftarrow Z_i^{\otimes n} \rightarrow Y_i^{\otimes n})$ by $(X_j^{\otimes n} \leftarrow Z_j^{\otimes n} \rightarrow Y_j^{\otimes n})$ for $i \neq j$ with an arbitrary given precision as in Shannon's coding theorem as described at the end of the previous subsection 0.1? More generally, what is the proper generalization of an asymptotic equivalence relation as discussed in the previous subsection to sequences of tensor powers of two-fans?*

We would like to argue that even though the entropies of the constituent spaces are all (pairwise) the same, all three examples above should be pairwise asymptotically different.

To establish that the examples in Figure 1 are different, that is, not isomorphic (see also Section 1.2) is relatively easy, since they have non-isomorphic symmetry groups. However, we present a different argument, that lends itself for generalization to prove that the examples at hand are not *asymptotically* equivalent and that also gives a quantitative difference between them.

To distinguish Example 1 from both 2 and 3, one could argue along the following lines. We could try to add a third space $Z = (\underline{Z}, p_Z)$ to the pair X and Y and provide a *joint distribution* p_Q on

$$Q = (\underline{X} \times \underline{Y} \times \underline{Z}, p_Q)$$

such that the projection of p_Q on the first two factors is p_U and on the third factor is p_Z.

Once we do that, we could evaluate entropies of various push-forwards of p_Q. Denote by $V = (\underline{X} \times \underline{Z}, p_V)$ and $W = (\underline{Y} \times \underline{Z}, p_W)$, where p_V and p_W are push-forwards of p_Q under corresponding coordinate projections. All the probability spaces now fit into a commutative diagram

$$\begin{tikzcd}
 & Q \arrow[dl] \arrow[d] \arrow[dr] & \\
U \arrow[d] \arrow[dr] & V \arrow[dl] \arrow[dr] & W \arrow[dl] \arrow[d] \\
X & Y & Z
\end{tikzcd}$$

where each arrow is a reduction, which is simply a measure-preserving map between probability spaces.

We consider the set of all possible extensions of the above form and denote it by $\mathrm{Ext}(X, Y, U)$. For any extension $\mathbf{E} = (X, Y, Z, U, V, W, Q)$ in $\mathrm{Ext}(X, Y, U)$ we have four "new" entropies

$$\mathcal{E}nt(Q), \quad \mathcal{E}nt(V), \quad \mathcal{E}nt(W), \quad \mathcal{E}nt(Z) \tag{0.2}$$

in addition to the "known" entropies of X, Y and U. The vector

$$\mathcal{E}nt_*(\mathbf{E}) := \big(\mathcal{E}nt(X), \mathcal{E}nt(Y), \mathcal{E}nt(Z), \mathcal{E}nt(U), \mathcal{E}nt(V), \mathcal{E}nt(W), \mathcal{E}nt(Q)\big)$$

is the entropy vector of the extension $\mathbf{E}$.

The set of all possible values of the entropy vector for all extensions of (X, Y, U)

$$\begin{aligned}&\mathbf{\Gamma}^\circ(X, Y, U) := \\ &\Big\{\mathcal{E}nt_*(\mathbf{E}) \in \mathbb{R}^7 \,\Big\|\, \mathbf{E} \text{ is an extension of } (X, Y, U)\Big\} \subset \mathbb{R}^7\end{aligned}$$

is what we call the unstabilized relative entropic set of the two-fan $(X \leftarrow U \rightarrow Y)$.

0.2.2. *The unstabilized relative entropic sets for the examples.* It turns out that these *unstabilized* relative entropic sets of $(X_1 \leftarrow U_1 \rightarrow Y_1)$ and $(X_2 \leftarrow U_2 \rightarrow Y_2)$ are different

$$\mathbf{\Gamma}^\circ(X_1, Y_1, U_1) \neq \mathbf{\Gamma}^\circ(X_2, Y_2, U_2)$$

To see this, let us calculate some particular points in the unstabilized relative entropic sets of the Examples 1–3. We consider the constrained Information-Optimization problem, of finding an extension $\mathbf{E} = (X, Y, Z, U, V, W, Q)$ of (X, Y, U) such that

(i) the space Z is a reduction of U, that is

$$\mathcal{E}nt(Q) = \mathcal{E}nt(U)$$

(ii) the spaces X and Y are independent conditioned on Z,

$$\mathcal{E}nt(Q) + \mathcal{E}nt(Z) = \mathcal{E}nt(V) + \mathcal{E}nt(W)$$

(iii) the sum

$$\mathcal{E}nt(X \lfloor Z) + \mathcal{E}nt(Y \lfloor Z)$$

is maximal, subject to conditions (i) and (ii).

It is very easy to read the solutions $\hat{\mathbf{E}}_1$, $\hat{\mathbf{E}}_2$ and $\hat{\mathbf{E}}_3$ of this optimization problem for Examples 1, 2 and 3 right from the pictures in Figure 1. Indeed, condition (i) says that Z_i must be a partition of U_i. Condition (ii) says that each set in the partition must be "rectangular", that is it must be a Cartesian product of a subset of $\underline{X}_i$ and a subset of $\underline{Y}_i$. The quantity to be maximized is the average log-area of the sets in the partition.

Optima are very easy to find "by hand". For Example 1 it is a partition of U_1 into three 2×2 squares. In Examples 2 and 3, one of the solutions is the partition of U_2, (resp. U_3) into 1×2 rectangles. Thus, the optimal values are $(2 \ln 2)$ for Example 1, and $(\ln 2)$ for Examples 2 and 3.

0.2.3. *The stabilized relative entropic set.* We have just seen that Examples 1 and 2 can be told apart by determining the unstabilized relative entropic set. However, this is not really what we are interested in. Rather, we wonder whether high tensor powers can be distinguished this way.

This relates to why we used the adjective "unstabilized": because the relative entropic set usually grows (is not stable) under taking tensor powers. That is, for every $n, k \in \mathbb{N}$ it holds that

$$k \cdot \mathbf{\Gamma}^{\circ}(X^{\otimes n}, Y^{\otimes n}, U^{\otimes n}) \subset \mathbf{\Gamma}^{\circ}(X^{\otimes (k \cdot n)}, Y^{\otimes (k \cdot n)}, U^{\otimes (k \cdot n)}) \tag{0.3}$$

but in general the set on the right-hand side can be strictly larger than the set on the left.

In view of the inclusion (0.3) we may define the *stabilized* relative entropic set

$$\overline{\mathbf{\Gamma}}(X, Y, U) = \mathrm{Closure}\left(\lim_{n \to \infty} \frac{1}{n} \mathbf{\Gamma}^{\circ}(X^{\otimes n}, Y^{\otimes n}, U^{\otimes n})\right)$$

This set turns out to be convex.

0.2.4. *The stabilized relative entropic set for the examples.* In fact, the *stabilized* relative entropic set also differentiates between Examples 1 and 2

$$\overline{\mathbf{\Gamma}}(X_1, Y_1, U_1) \neq \overline{\mathbf{\Gamma}}(X_2, Y_2, U_2)$$

The proof of this fact follows the same lines as in Section 0.2.2, but the stabilization makes the argument much more technical.

We expect that the stabilized relative entropic set cannot differentiate between Examples 2 and 3. However, there are other types of relative entropic sets, and other Information-Optimization problems that *can* differentiate between Examples 2 and 3.

The relative entropic sets are discussed in Section 7.

0.3. **Information-Optimization problems and relative entropic sets.** In Section 0.2.2 we used an Information-Optimization problem to find particular points in the (unstabilized) relative entropic set. This is no coincidence, and the link between stabilized Information-Optimization problems and the stable relative entropic set can be made very explicit. Because the stable relative entropic set is convex, it can be completely characterized by Information-Optimization problems and vice versa.

Such Information-Optimization problems play a very important role in information theory [Yeu12], causal inference [SA15], artificial intelligence [VDP13], information decomposition [BRO+14], robotics [ABD+08], and neuroscience [Fri09]. The techniques developed in the article allow one to address this type of problems easily and efficiently.

0.4. **The intrinsic Kolmogorov-Sinai distance.** As we mentioned at the end of Section 0.1, one is tempted to define asymptotically equivalent configurations along the lines of Shannon's source coding theorem following [Gro12]. Two configurations would be asymptotically equivalent if there is an almost measure-preserving bijection between subspaces of almost full measure in their high tensor powers.

However, we found this approach inconvenient. Instead of finding an almost measure-preserving bijection between large parts of the two spaces, we consider a stochastic coupling (transportation plan, joint distribution) between a pair of spaces and measure its deviation from being an isomorphism of probability spaces, that is a measure-preserving bijection. Such a measure of deviation from being an isomorphism then leads to the notion of intrinsic Kolmogorov-Sinai distance, and its stable version – the asymptotic Kolmogorov-Sinai distance, as explained in Section 4.

In the case of single probability spaces we define the *intrinsic Kolmogorov-Sinai distance* between two probability spaces $X = (\underline{X}, p_X)$ and $Y = (\underline{Y}, p_Y)$ by

$$\mathbf{k}(X,Y) := \inf\left\{\left[\,\mathcal{E}nt(Z) - \mathcal{E}nt(X)\right] + \left[\,\mathcal{E}nt(Z) - \mathcal{E}nt(Y)\right]\right\}$$

where the infimum is taken over all choices of the joint distribution $Z = (\underline{X} \times \underline{Y}, p_Z)$. Note that each of the summands is nonnegative and vanishes if and only if the corresponding marginalization $Z \to X$ or $Z \to Y$ is an isomorphism of probability spaces. In this sense the distance measures the deviation from the existence of a measure-preserving bijection between X and Y.

Furthermore, we define the *asymptotic Kolmogorov-Sinai distance* between two probability spaces X and Y by

$$\boldsymbol{\kappa}(X,Y) = \lim_{n\to\infty} \frac{1}{n}\mathbf{k}(X^{\otimes n}, Y^{\otimes n}).$$

This definition could be generalized to configurations of probability spaces and we will say that two configurations are asymptotically equivalent if the asymptotic Kolmogorov-Sinai distance between them vanishes.

0.5. **Asymptotic Equipartition Property.** Examples 1, 2, and 3 above have the property that the symmetry group acts transitively on the support of the measure on U_i and they are particular instances of what we call homogeneous configurations.

In Section 6, we show an *Asymptotic Equipartion Property for configurations*: Theorem 6.1 states that every sequence of tensor powers of a configuration can be approximated in the asymptotic Kolmogorov-Sinai distance by a sequence of homogeneous configurations.

This Asymptotic Equipartition Property allows one to substitute configurations of probability spaces by homogeneous approximations. Homogeneous probability spaces are just uniform probability spaces, and as a first simple consequence of the Asymptotic Equipartition Property, the asymptotic Kolmogorov-Sinai distance between probability spaces X and Y can be computed and equals

$$\boldsymbol{\kappa}(X,Y) = |\mathcal{E}nt(X) - \mathcal{E}nt(Y)|.$$

Homogeneous *configurations* are, unlike homogeneous probability spaces, rather complex objects. Nonetheless, they seem to be simpler than arbitrary configurations of probability spaces for the types of problems that we would like to address.

More specifically, we show in Section 7 that the optimal values in (stabilized) Information-Optimization problems only depend on the asymptotic class of a configuration and that they are continuous with respect to the asymptotic Kolmogorov-Sinai distance; in many cases, the optimizers are continuous as well. The Asymptotic Equipartition Property implies that for the purposes of calculating optimal values and approximate optimizers, one only needs to consider homogeneous configurations and this can greatly simplify computations.

Summarizing, the Asymptotic Equipartition Property and the continuity of Information-Optimization problems are important justifications for the choice of asymptotic equivalence relation and the introduction of the intrinsic and asymptotic Kolmogorov-Sinai distances.

0.6. **The article.** The article has the following structure. Section 1 is devoted to the basic setup used throughout the text. In Section 2 we explain what we mean by configurations of probability spaces, give examples, describe simple properties and operations. Further, in Section 3 we generalize the notion of probability distribution to that on configurations and discuss the theory of types for configurations. In Section 4 the intrinsic Kolmogorov-Sinai distance and the asymptotic Kolmogorov-Sinai distance are introduced and some technical tools for the estimation of Kolmogorov distance are developed. Section 5 contains estimates on the distances between types. We use these estimates in the proof of the Asymptotic Equipartition Property for configurations in Section 6. Section 7 deals with extensions of configurations. We prove there the Extension Lemma, which is used to show continuity of extensions and implies, in particular, that solutions of the constrained optimization problem for the entropies of extensions are Lipschitz-continuous with respect to the asymptotic Kolmogorov-Sinai distance, thus they only depend on the asymptotic classes of configurations. In Section 8 we briefly discuss a special type of configurations called mixtures, which will play an important role in the construction of tropical probability spaces. Finally, in Section 9 we introduce the notion of tropical probability spaces and configurations thereof, and list some of their properties. We will continue our study of tropical probability spaces and configurations in subsequent articles.

Some technical and not very illuminating proofs are deferred to Section T Technical Proofs. In the electronic version one can move between the proof in the technical section and the statement in the main text by following the link (arrow up or down).

Acknowledgments. We would like to thank Tobias Fritz, František Matúš, Misha Movshev and Johannes Rauh for inspiring discussions. We are grateful to the participants of the Wednesday Morning Session at the CASA group at the Eindhoven University of Technology for valuable feedback on the introduction of the article. Finally, we thank the Max Planck Institute for Mathematics in the Sciences, Leipzig, for its hospitality.

1. Category of probability spaces and configurations

This section is devoted to the basic setup used throughout the present article. We introduce a category of probability spaces and reductions, similar to categories introduced in [BFL11] and [Gro12], and define configurations of probability spaces and the corresponding category. The last subsection recalls the notion of entropy and its elementary properties.

1.1. **Probability spaces and reductions.** Below we will consider probability spaces such that the support of the probability measure is finite. Any such space contains a full-measure subspace isomorphic to a finite space, thus we call such objects *finite probability spaces*. For a probability space $X = (S, p_X)$ denote by $\underline{X} = \operatorname{supp} p_X$ the support of the measure and by $|X|$ its cardinality. Slightly abusing the language, we call this quantity the *cardinality* of X.

For a pair of probability spaces a *reduction* $X \to Y$ is a class of measure-preserving maps, with two maps being equivalent if they coincide on a set of full measure. The composition of two reductions is itself a reduction. Two probability spaces are isomorphic if there is a measure-preserving bijection between the supports of the probability measures. Such a bijection defines an invertible reduction from one space into another. Clearly the cardinality $|X|$ is an isomorphism invariant. The automorphism group $\operatorname{Aut}(X)$ is the group of all self-isomorphisms of X.

A probability space X is called *homogeneous* if the automorphism group $\operatorname{Aut}(X)$ acts transitively on the support $\underline{X}$ of the measure. The property of being homogeneous is an isomorphism invariant. In the isomorphism class of a homogeneous space there is a representative with uniform measure.

The finite probability spaces and reductions form a category, that we denote by **Prob**. The subcategory of homogeneous spaces will be denoted by $\mathbf{Prob_h}$. The isomorphism in the category coincides with the notion of isomorphism above.

The category **Prob** is not a small category. However it has a small full subcategory, that contains an object for every isomorphism class in **Prob** and for every pair of objects in it, it contains all the available morphisms between them. From now on we imagine that such a subcategory was chosen and fixed and replaces **Prob** in all considerations below.

There is a product in **Prob** given by the Cartesian product of probability spaces, that we will denote by $X \otimes Y := (\underline{X} \times \underline{Y}, p_X \otimes p_Y)$. There are canonical reductions $X \otimes Y \to X$ and $X \otimes Y \to Y$ given by projections to factors. For a pair of reductions $f_i : X_i \to Y_i$, $i = 1, 2$ their tensor product is the reduction $f_1 \otimes f_2 : X_1 \otimes X_2 \to Y_1 \otimes Y_2$, which is equal to the class of the Cartesian product of maps representing f_i's. The tensor product is not however a categorical product. The product leaves the subcategory of homogeneous spaces invariant.

The probability measure on X will be usually denoted by p_X or simply p, when the risk of confusion is low.

1.2. **Configurations of probability spaces.** Essentially, a configuration $\mathcal{X} = \{X_i; f_{ij}\}$ is a commutative diagram consisting of a finite number of probability spaces and reductions between some of them, that is transitively closed, while a morphism $\rho : \mathcal{X} \to \mathcal{Y}$ between two configurations $\mathcal{X} = \{X_i; f_{ij}\}$ and $\mathcal{Y} = \{Y_i; g_{ij}\}$ of the same combinatorial type is a collection of reductions between corresponding individual objects $\rho_i : X_i \to Y_i$, that commute with the reductions within each configuration, $\rho_j \circ f_{ij} = g_{ij} \circ \rho_i$.

We need to keep track of the combinatorial structure of the collection of reductions within a configuration. There are several possibilities for doing so:

- the reductions form a directed transitively closed graph without loops;
- the spaces in the configuration form a poset;
- the underlying combinatorial structure could be recorded as a finite category.

The last option seems to be most convenient since it has many operations necessary for our analysis already built-in.

A *diagram category* $\mathbf{G}$ is a finite category such that for each pair of objects O_1, O_2 in $\mathbf{G}$ the morphism space between them

$$\mathrm{Hom}_{\mathbf{G}}(O_1, O_2) \cup \mathrm{Hom}_{\mathbf{G}}(O_2, O_1)$$

contains at most one element.

For a diagram category $\mathbf{G}$ a *configuration of probability spaces* modeled on $\mathbf{G}$ is a functor $\mathcal{X} : \mathbf{G} \to \mathbf{Prob}$. The collection of all configurations of probability spaces modeled on a fixed diagram category $\mathbf{G}$ forms the category of functors $\mathbf{Prob}\langle\mathbf{G}\rangle := [\mathbf{G}, \mathbf{Prob}]$. The objects of $\mathbf{Prob}\langle\mathbf{G}\rangle$ are configurations, that is functors from $\mathbf{G}$ to $\mathbf{Prob}$, while morphisms in $\mathbf{Prob}\langle\mathbf{G}\rangle$ are natural transformations between them. For a configuration $\mathcal{X} \in \mathbf{Prob}\langle\mathbf{G}\rangle$, the diagram category $\mathbf{G}$ will be called the *combinatorial type* of $\mathcal{X}$.

For a diagram category $\mathbf{G}$ or a configuration $\mathcal{X} \in \mathbf{Prob}\langle\mathbf{G}\rangle$ we denote by $[\![\mathbf{G}]\!] = [\![\mathcal{X}]\!]$ the number of objects in the category $\mathbf{G}$.

An object O in a diagram category $\mathbf{G}$ will be called *initial*, if it is not a target of any morphism except for the identity. Likewise a *terminal* object is not a source of any morphism, except for the identity morphism. Note that this terminology is somewhat unconventional from the point of view of category theory.

A diagram category is called *complete* if it has a unique initial object. Thus a configuration modeled on a complete category includes a space that reduces to all other spaces in the configuration.

The above terminology transfers to configurations modeled on $\mathbf{G}$: An initial space in $\mathcal{X} \in \mathbf{Prob}\langle\mathbf{G}\rangle$ is one that is not a target space of any reduction within the configuration, a terminal space is not a source of any non-trivial reduction and $\mathcal{X}$ is complete if $\mathbf{G}$ is, that is there is a unique initial space.

The tensor product of probability spaces extends to a tensor product of configurations. For $\mathcal{X}, \mathcal{Y} \in \mathbf{Prob}\langle\mathbf{G}\rangle$, such that $\mathcal{X} = \{X_i; f_{ij}\}$ and $\mathcal{Y} = \{Y_i; g_{ij}\}$ define

$$\mathcal{X} \otimes \mathcal{Y} := \{X_i \otimes Y_i; f_{ij} \otimes g_{ij}\}$$

Occasionally we will also talk about configuration of sets. Denote by $\mathbf{Set}$ the category of finite sets and surjective maps. Then all of the above constructions could be repeated for sets instead of probability spaces. Thus we could talk about the category of configurations of sets $\mathbf{Set}\langle\mathbf{G}\rangle$.

Given a reduction $f : X \to Y$ between two probability spaces, the restriction $\underline{f} : \underline{X} \to \underline{Y}$ is a well-defined surjective map. Given a configuration $\mathcal{X} = \{X_i; f_{ij}\}$ of probability spaces, there is an underlying configuration of sets, obtained by taking the supports of measures on each level and restricting reductions on these supports. We will denote it by $\underline{\mathcal{X}} = \{\underline{X}_i; f_{ij}\}$, where $\underline{X}_i := \operatorname{supp} p_{X_i}$. Thus we have a forgetful functor

$$\underline{\cdot} : \mathbf{Prob}\langle\mathbf{G}\rangle \to \mathbf{Set}\langle\mathbf{G}\rangle$$

For now we will consider two important examples of diagram categories and configurations modeled on them. We give further examples in Section 2.1.

1.2.1. *Two-fans:* A two-fan is a configuration modeled on the category $\mathbf{\Lambda}$ with three objects, one initial and two terminal.

$$\mathbf{\Lambda} = (O_1 \leftarrow O_{12} \to O_2)$$

There is a special significance to two-fans, since these are the simplest non-trivial configurations, as we will see later.

Essentially, a two-fan $X \leftarrow Z \to Y$ is a triple of probability spaces and a pair of reductions between them.

A reduction of a two-fan $X \leftarrow Z \to Y$ to another two-fan $X' \leftarrow Z' \to Y'$ is a triple of reductions $Z \to Z'$, $Y \to Y'$ and $X \to X'$ that commute with the reductions within each fan, that is, the following diagram is commutative

$$\begin{array}{ccccc} X & \longleftarrow & Z & \longrightarrow & Y \\ \downarrow & & \downarrow & & \downarrow \\ X' & \longleftarrow & Z' & \longrightarrow & Y' \end{array}$$

Isomorphisms and the automorphism group $\mathrm{Aut}(\cdot)$ are defined accordingly. Note that terminal spaces in a two-fan are labeled and reductions preserve the labeling.

A two-fan $X \leftarrow Z \rightarrow Y$ is called *minimal* if for a.e. $x \in X$ and $y \in Y$ there is a unique $z \in Z$, that reduces to y and to x. Given a two-fan $X \leftarrow Z \rightarrow Y$, there is always a reduction to a minimal two-fan $X \leftarrow Z' \rightarrow Y$. Such minimal reduction is unique up to isomorphism. Explicitly, take $Z' := \underline{X} \times \underline{Y}$ as a set and consider a probability distribution on Z' induced by a map $Z \rightarrow Z'$ which is the Cartesian product of the reductions $X \leftarrow Z \rightarrow Y$ in the original two-fan.

The notion of being minimal is in fact a categorical notion. It could be equivalently defined by saying that a two-fan $\mathcal{X} = (X_1 \leftarrow X_{12} \rightarrow X_2)$ is minimal if for any reduction $\lambda : \mathcal{X} \rightarrow \mathcal{X}'$ holds: if both λ_1 and λ_2 are isomorphisms, then λ_{12} is also an isomorphism. Consequently, if one specifies reductions from the terminal spaces of a minimal two-fan to another two-fan, then there exists at most one extension to the reduction of the whole fan.

The inclusion of a pair of probability spaces X and Y as terminal vertices in a minimal two-fan is equivalent to specifying a joint distribution on $\underline{X} \times \underline{Y}$.

An arbitrary configuration $\mathcal{X}$ will be called *minimal* if with every two-fan, it also contains a minimal two-fan with the same terminal spaces. We will denote the space of minimal configurations modelled on a diagram category $\mathbf{G}$ by $\mathbf{Prob}\langle\mathbf{G}\rangle_{\mathbf{m}}$.

Given a two-fan

$$\mathcal{F} = (X \leftarrow Z \rightarrow Y)$$

with terminal spaces X and Y, and a point $x \in X$ with $p_X(x) > 0$, one may construct a *conditional probability distribution* $p_Y(\,\cdot\,\lfloor x)$ on $\underline{Y}$. We denote the corresponding space $Y\lfloor x := \big(\underline{Y}, p_Y(\,\cdot\,\lfloor x)\big)$. The usual bar "|", that is normally used for conditioning, interferes with our notations for cardinality of spaces. We will give more details in Section 2.7.

1.2.2. *A diamond configuration.* A "diamond" configuration is modeled on a diamond category $\diamond$, that consists of a two-fan and a "co-fan":

$$\diamond = \left(\begin{tikzcd} & O_{12} \arrow[dl] \arrow[dr] & \\ O_1 \arrow[dr] & & O_2 \arrow[dl] \\ & O_\bullet & \end{tikzcd}\right)$$

Of course, there is also a morphism $O_{12} \rightarrow O_\bullet$, which lies in the transitive closure of the given four morphisms. As a rule, we will skip writing morphisms, that are implied by the transitive closure.

A diamond configuration is minimal if the top two-fan in it is minimal.

1.3. **Entropy.** Our working definition of *entropy* will be based on the following version of the asymptotic equipartition theorem for Bernoulli process, see [CT91].

Theorem 1.1. *Suppose X is a finite probability space, then for any $\varepsilon > 0$ and any $n \gg 0$ there exists a subset $A_\varepsilon^{(n)} \subset X^{\otimes n}$ such that*

(i) $p(A_\varepsilon^{(n)}) \geq 1 - \varepsilon$

(ii) *For any $a, a' \in A_\varepsilon^{(n)}$ holds*

$$\left| \frac{\ln p(a)}{n} - \frac{\ln p(a')}{n} \right| \leq \varepsilon$$

Moreover, if $A_\varepsilon^{(n)}$ and $B_\varepsilon^{(n)}$ are two subsets of $X^{\otimes n}$ satisfying two conditions above, then their cardinalities satisfy

$$\left| \frac{\ln |A_\varepsilon^{(n)}|}{n} - \frac{\ln |B_\varepsilon^{(n)}|}{n} \right| \leq 2\varepsilon \tag{1.1}$$

⊠

Then we define

$$\mathcal{E}nt(X) \coloneqq \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{n} \ln |A_\varepsilon^{(n)}|$$

Clearly, in view of the property (1.1) in the Theorem 1.1 the limit above is well-defined and is independent of the choice of the typical subsets $A_\varepsilon^{(n)}$.

Entropy satisfies the so-called Shannon inequality, see for example [CT91], namely for any minimal diamond configuration

$$\begin{tikzcd} & X_{12} \arrow[dl] \arrow[dr] \arrow[dd] & \\ X_1 \arrow[dr] & & X_2 \arrow[dl] \\ & X_\bullet & \end{tikzcd}$$

the following inequality holds,

$$\mathcal{E}nt(X_1) + \mathcal{E}nt(X_2) \geq \mathcal{E}nt(X_{12}) + \mathcal{E}nt(X_\bullet) \tag{1.2}$$

Furthermore, entropy is additive with respect to the tensor product, that is, for a pair of probability spaces $X, Y \in \mathbf{Prob}$ holds

$$\mathcal{E}nt(X \otimes Y) = \mathcal{E}nt(X) + \mathcal{E}nt(Y) \tag{1.3}$$

Further, for a pair X, Y of probability spaces included in a minimal two-fan $(X \leftarrow Z \rightarrow Y)$ we define the conditional entropy

$$\mathcal{E}nt(X \lfloor Y) \coloneqq \mathcal{E}nt(Z) - \mathcal{E}nt(X)$$

The above quantity is always non-negative in view of Shannon inequality (1.2). Moreover, the following identity holds, see [CT91]

$$\mathcal{E}nt(X \lfloor Y) = \int_Y \mathcal{E}nt(X \lfloor y) \, \mathrm{d}\, p_Y(y) \tag{1.4}$$

2. CONFIGURATIONS

In this section we will look at configurations in more detail. We start by considering some important examples.

2.1. **Examples of configurations.**

2.1.1. *Singleton.* We denote by $\bullet$ a diagram category with a single object. Clearly configurations modeled on $\bullet$ are just probability spaces and we have $\mathbf{Prob} \equiv \mathbf{Prob}\langle\bullet\rangle$.

2.1.2. *Chains.* The chain $\mathbf{C}_n$ of length $n \in \mathbb{N}$ is a category with n objects $\{O_i\}_{i=1}^n$ and morphisms from O_i to O_j whenever $i \le j$. A configuration $\mathcal{X} \in \mathbf{Prob}\langle\mathbf{C}_n\rangle$ is a chain of reductions

$$\mathcal{X} = (X_1 \to X_2 \to \cdots \to X_n)$$

2.1.3. *Two-fan.* The two-fan $\mathbf{\Lambda}$ is a category with three objects $\{O_1, O_{12}, O_2\}$ and two non-identity morphisms $O_{12} \to O_1$ and $O_{12} \to O_2$. See also Section 1.2.1. Two-fans are the simplest configurations for which asymptotic equivalence classes contain more information than just entropies of the entries.

Recall that a fan $(X \leftarrow Z \to Y)$ is called minimal, if for any pair of points $x \in X$ and $y \in Y$ with positive weights there exists at most one $z \in Z$, that reduces to x and to y. Equivalently, for any super-configuration

$$\begin{tikzcd}
 & Z \arrow[d] \arrow[ddl, two heads] \arrow[ddr, two heads] & \\
 & Z' \arrow[dl, two heads] \arrow[dr, two heads] & \\
X & & Y
\end{tikzcd}$$

the reduction $Z \to Z'$ must be an isomorphism.

2.1.4. *Full configuration.* The full category $\mathbf{\Lambda}_n$ on n objects is a category with objects $\{O_I\}_{I \in 2^{\{1,\ldots,n\}} \setminus \{\varnothing\}}$ indexed by all non-empty subsets $I \in 2^{\{1,\ldots,n\}}$ and a morphism from O_I to O_J, whenever $J \subseteq I$.

For a collection of random variables $X_1, \ldots, X_n$ one may construct a minimal full configuration $\mathcal{X} \in \mathbf{Prob}\langle\mathbf{\Lambda}_n\rangle$ by considering all joint distributions and "marginalization" reductions. We denote such a configuration by $\langle X_1, \ldots, X_n\rangle$. On the other hand, the terminal vertices of a full configuration can be viewed as random variables on the domain of definition given by the (unique) initial space.

Suppose $\mathcal{X} \in \mathbf{Prob}\langle\mathbf{\Lambda}_n\rangle$ is a minimal full configuration with terminal vertices $X_1, \ldots, X_n$. It is convenient to view $\mathcal{X}$ as a distribution on the Cartesian product of the underlying sets of the terminal vertices:

$$p_{\mathcal{X}} \in \Delta(\underline{X}_1 \times \cdots \times \underline{X}_n)$$

Once the underlying sets of the terminal spaces are fixed, there is a one-to-one correspondence between the full minimal configurations and distributions as above.

2.1.5. *"Two-tents" configuration.* The "two-tents" category $\mathbf{M}_2$ consists of five objects, of which two are initial and three are terminal, and morphisms are as follows

$$\mathbf{M}_2 = \left(\begin{tikzcd} O_{12} \arrow[d] \arrow[dr] & & O_{23} \arrow[dl] \arrow[d] \\ O_1 & O_2 & O_3 \end{tikzcd}\right)$$

Thus, a typical two-tents configuration consists of five probability spaces and reduction as in

$$\mathcal{X} = (X \leftarrow U \rightarrow Y \leftarrow V \rightarrow Z)$$

The probability spaces U and V are initial and X, Y and Z are terminal.

2.1.6. *"Many-tents" configuration.* The previous example could be generalized to a "many-tents" category

$$\mathbf{M}_n = (O_1 \leftarrow O_{12} \rightarrow O_2 \leftarrow \cdots \rightarrow O_{n-1} \leftarrow O_{n-1,n} \rightarrow O_n)$$

2.1.7. *"Fence" configuration.* The "fence" category $\mathbf{W}_3$ consists of six objects and the morphisms are

$$\mathbf{W}_3 = \left(\begin{tikzcd} O_{12} \arrow[d] \arrow[dr] & O_{13} \arrow[dl] \arrow[dr] & O_{23} \arrow[dl] \arrow[d] \\ O_1 & O_2 & O_3 \end{tikzcd}\right)$$

2.1.8. *Co-fan.* A co-fan $\mathbf{V}$ is a category with three objects and morphisms as in the diagram

$$\mathbf{V} = \left(\begin{tikzcd} O_1 \arrow[dr] & & O_2 \arrow[dl] \\ & O_\bullet & \end{tikzcd}\right)$$

2.1.9. *"Diamond" configurations.* A "diamond" configuration $\diamond$ is modeled on a diamond category that consists of a fan and a co-fan

$$\diamond = \left(\begin{tikzcd} & O_{12} \arrow[dl] \arrow[dr] & \\ O_1 \arrow[dr] & & O_2 \arrow[dl] \\ & O_\bullet & \end{tikzcd}\right)$$

See also Section 1.2.2.

Examples 2.1.1, 2.1.2, 2.1.3, 2.1.4 and 2.1.9 are complete. Examples 2.1.1, 2.1.2 and 2.1.8 do not contain a two-fan. Tropical limits of such configurations are very simple. Essentially, such tropical limits correspond to the tuple of numbers corresponding to the entropies of the constituent spaces. Therefore, we call configurations not containing a two-fan *simple*.

2.2. **Constant configurations.** Suppose X is a probability space and $\mathbf{G}$ is a diagram category. One may form a *constant* $\mathbf{G}$*-configuration* by considering a functor that maps all objects in $\mathbf{G}$ to X and all the morphisms to the identity morphism $X \to X$. We denote such a constant configuration by $X^{\mathbf{G}}$ or simply by X, when $\mathbf{G}$ is clear from the context. Any constant configuration is automatically minimal.

If $\mathcal{Y} = \{Y_i; f_{ij}\}$ is another $\mathbf{G}$-configuration, then a reduction $\rho : \mathcal{Y} \to X^{\mathbf{G}}$ (which we write sometimes simply as $\rho : \mathcal{Y} \to X$) is a collection of reductions $\rho_i : Y_i \to X$, such that

$$f_{ij} \circ \rho_i = \rho_j$$

2.3. **Configurations of configurations.** Of course, the operation of "configuration" could be iterated, so given a pair $\mathbf{G}_1$, $\mathbf{G}_2$ of diagram categories we could form a $\mathbf{G}_2$-configuration of $\mathbf{G}_1$-configurations, so we could speak, for example, about a two-fan of configurations of the same type.

$$\mathbf{Prob}\langle\mathbf{G}_1, \mathbf{G}_2\rangle \coloneqq \mathbf{Prob}\langle\mathbf{G}_1\rangle\langle\mathbf{G}_2\rangle = \mathbf{Prob}\langle\mathbf{G}_1 \square \mathbf{G}_2\rangle$$

where $\mathbf{G}_1 \square \mathbf{G}_2$ is the "Cartesian product of graphs" (as every diagram category could be considered as a transitively closed directed graph). This operation is commutative, thus, for example, a two-fan of $\mathbf{G}$-configurations is a $\mathbf{G}$-configuration of two-fans.

We will rarely need anything beyond a two-fan of configurations.

A two-fan $\mathcal{F} = (\mathcal{X} \leftarrow \mathcal{Z} \to \mathcal{Y})$ of $\mathbf{G}$-configurations is called minimal if in any extension of $\mathcal{F}$ of the form

$$\begin{tikzcd} & \mathcal{Z} \arrow[dl] \arrow[d, "f"] \arrow[dr] & \\ & \mathcal{Z}' \arrow[l] \arrow[r] & \\ \mathcal{X} & & \mathcal{Y} \end{tikzcd}$$

the reduction $f : \mathcal{Z} \to \mathcal{Z}'$ must be an isomorphism of $\mathbf{G}$-configurations.

Recall that a two-fan of $\mathbf{G}$-configurations could also be viewed as a $\mathbf{G}$-configuration of two-fans of probability spaces. In the following lemma we show that in order to verify the minimality of a two-fan of configurations it is sufficient to check the minimality of all the constituent two-fans.

Lemma 2.1. ↓ *Let* $\mathbf{G}$ *be a diagram category. Then*

(i) *A two-fan* $\mathcal{F} = (\mathcal{X} \leftarrow \mathcal{Z} \to \mathcal{Y})$ *of* $\mathbf{G}$*-configurations is minimal, if and only if the constituent two-fans of probability spaces* $\mathcal{F}_i = (X_i \leftarrow Z_i \to Y_i)$ *are all minimal.*

(ii) *For any two-fan* $\mathcal{F} = (\mathcal{X} \leftarrow \mathcal{Z} \to \mathcal{Y})$ *of* $\mathbf{G}$*-configurations its minimal reduction exists, that is, there exists a minimal two-fan* $\mathcal{F}' = (\mathcal{X} \leftarrow$

$\mathcal{Z}' \to \mathcal{Y}$) *included in the following diagram*

$$\begin{tikzcd}[column sep=small]
 & \mathcal{Z} \arrow[d] \arrow[ddl] \arrow[ddr] & \\
 & \mathcal{Z}' \arrow[dl] \arrow[dr] & \\
\mathcal{X} & & \mathcal{Y}
\end{tikzcd}$$

⊠

Even though this lemma is rather elementary, there are many similar statements that are not true. Thus we are compelled to provide a proof, which can be found in Section T on page 59.

Similarly, a full configuration $\mathcal{F}$ of $\mathbf{G}$-configurations is called minimal if for every two-fan of $\mathbf{G}$-configurations in $\mathcal{F}$ there is a minimal two-fan in $\mathcal{F}$ of $\mathbf{G}$-configurations with the same terminal configurations.

Lemma 2.1 has the following corollary and counterpart for full configurations of $\mathbf{G}$-configurations.

Corollary 2.2. *Let* $\mathbf{G}$ *be a diagram category. Then*

(i) *A full configuration* $\mathcal{F}$ *of* $\mathbf{G}$*-configurations is minimal, if and only if the constituent full configurations of probability spaces* $\mathcal{F}_i$ *are all minimal.*

(ii) *For any full configuration* $\mathcal{F}$ *of* $\mathbf{G}$*-configurations its minimal reduction exists.*

⊠

2.4. **Restrictions and extensions.** Suppose $R : \mathbf{G}_1 \to \mathbf{G}_2$ is a functor between two diagram categories. For a configuration $\mathcal{X} : \mathbf{G}_2 \to \mathbf{Prob}$, the pullback configuration $\mathcal{Y} = R^*\mathcal{X} \in \mathbf{Prob}\langle\mathbf{G}_1\rangle$ defined as the composition

$$\mathcal{Y} := \mathcal{X} \circ R$$

is called an *R-restriction* of $\mathcal{X}$ to $\mathbf{G}_1$ and $\mathcal{X}$ is the *extension* of $\mathcal{Y}$. If the functor R is injective then we call $\mathcal{Y} = R^*\mathcal{X}$ a *sub-configuration* of $\mathcal{X}$ and write $\mathcal{Y} \subset \mathcal{X}$, likewise $\mathcal{X}$ will be called a *super-configuration* of $\mathcal{Y}$.

The restriction operation is functorial in the sense that given two configurations $\mathcal{X}, \mathcal{X}' \in \mathbf{Prob}\langle\mathbf{G}_1\rangle$ and a reduction $f : \mathcal{X} \to \mathcal{X}'$, there is a canonical reduction $R^*f : R^*\mathcal{X} \to R^*\mathcal{X}'$. Thus R^* can be considered as a functor

$$R^* : \mathbf{Prob}\langle\mathbf{G}_2\rangle \to \mathbf{Prob}\langle\mathbf{G}_1\rangle$$

Some important examples of restrictions and extensions are below.

2.4.1. *Restriction of a full configuration to a smaller full configuration.* Recall that, as explained in Section 2.1.4, the terminal vertices of a full configuration could be considered as random variables and any collection of random variables "generates" a full configuration.

For a full configuration $\mathcal{X} = \langle X_i\rangle_{i=1}^n$ and a subset $I \subset \{1, \dots, n\}$ we denote by $R_I^*\mathcal{X} = \langle X_i\rangle_{i\in I}$ the restriction of $\mathcal{X}$ to a full configuration generated by X_i,

$i \in I$. We will also make use of the notation $R^*_{k,l}$ for the restriction operator $R^*_{\{1,\dots,k\}} : \mathbf{Prob}\langle \mathbf{\Lambda}_l \rangle \to \mathbf{Prob}\langle \mathbf{\Lambda}_k \rangle$.

2.4.2. *Restriction of a $\mathbf{\Lambda}_3$-configuration to an $\mathbf{M}_2$-configuration.* Given a full configuration $\mathcal{X} \in \mathbf{Prob}\langle \mathbf{\Lambda}_3 \rangle$ we may "forget" part of the data. If we, for example, forget the top space and the relation between a pair out of three terminal spaces we end up with the two-tents configuration. This operation corresponds to the inclusion functor

$$M : \left(\begin{tikzcd} O_{12} \arrow[d] \arrow[dr] & & O_{23} \arrow[dl] \arrow[d] \\ O_1 & O_2 & O_3 \end{tikzcd} \right) \longrightarrow \left(\begin{tikzcd} & Q_{123} \arrow[dl] \arrow[d] \arrow[dr] & \\ Q_{12} \arrow[d] \arrow[dr] & Q_{31} \arrow[dl] \arrow[dr] & Q_{23} \arrow[dl] \arrow[d] \\ Q_1 & Q_2 & Q_3 \end{tikzcd} \right)$$

that preserves the sub-indices.

We show in Section 2.5 below that the corresponding restriction operator

$$M^* : \mathbf{Prob}\langle \mathbf{\Lambda}_3 \rangle \to \mathbf{Prob}\langle \mathbf{M}_3 \rangle$$

is surjective, both on objects and all morphisms. Thus, as a map of collections of objects it has a right inverse. However, no natural right inverse exists.

2.4.3. *Restriction of a $\mathbf{\Lambda}_3$-configuration to a $\mathbf{W}_3$-configuration.* Starting with a full configuration we might choose to forget the initial space (and reductions, for which it was the domain). The remaining configuration has the combinatorial type of a fence. This operation corresponds to the functor

$$W : \left(\begin{tikzcd} O_{12} \arrow[d] \arrow[dr] & O_{31} \arrow[dl] \arrow[dr] & O_{23} \arrow[dl] \arrow[d] \\ O_1 & O_2 & O_3 \end{tikzcd} \right) \longrightarrow \left(\begin{tikzcd} & Q_{123} \arrow[dl] \arrow[d] \arrow[dr] & \\ Q_{12} \arrow[d] \arrow[dr] & Q_{31} \arrow[dl] \arrow[dr] & Q_{23} \arrow[dl] \arrow[d] \\ Q_1 & Q_2 & Q_3 \end{tikzcd} \right)$$

The corresponding operator

$$W^* : \mathbf{Prob}\langle \mathbf{\Lambda}_3 \rangle \to \mathbf{Prob}\langle \mathbf{W}_3 \rangle$$

is not surjective. To find out when a $\mathbf{W}_3$-configuration is extendable to a $\mathbf{\Lambda}_3$-configuration is an interesting problem, see for example, [ABK+15] and references therein. It is our hope that the methods developed in this article might be useful to address these questions.

2.4.4. *Doubling.* This will be the first example of an interesting functor between diagram categories, which is not injective. In this situation the term "restriction" does not really reflect the operation of pull-back well, however we did not come up with a better terminology.

The doubling operation is the restriction of a two-fan to a two-tents configuration. Consider the two-fan category $\mathbf{\Lambda} = (O_1 \leftarrow O_{12} \rightarrow O_2)$ and a two-tents category $\mathbf{M}_2 = (Q_1 \leftarrow Q_{12} \rightarrow Q_2 \leftarrow Q_{23} \rightarrow Q_3)$. Define the functor $D : \mathbf{M}_2 \rightarrow \mathbf{\Lambda}$ by setting

$$D : \begin{cases} Q_1 \mapsto O_1 \\ Q_2 \mapsto O_2 \\ Q_3 \mapsto O_1 \\ Q_{12} \mapsto O_{12} \\ Q_{23} \mapsto O_{12} \end{cases}$$

Note that D extends uniquely to the spaces of morphisms, since each morphism space is either empty or a one-point set.

Thus $D^*(X \leftarrow Z \rightarrow Y) = (X \leftarrow Z \rightarrow Y \leftarrow Z \rightarrow X)$, where the "left" and "right" two-fans are isomorphic.

This operation along with a particular Information-Optimization problem is related to the so-called copy operation, that was used to find many non-Shannon information inequalities, as described, for example, in [DFZ11].

2.5. **Adhesion.** Given a minimal two-tents configuration $\mathcal{X} = (X \leftarrow U \rightarrow Y \leftarrow V \rightarrow Z) \in \mathbf{Prob}\langle\mathbf{M}_2\rangle_{\mathbf{m}}$ one could always construct an extension of $\mathcal{X}$ to a full configuration $\mathbf{ad}(\mathcal{X}) \in \mathbf{Prob}\langle\mathbf{\Lambda}_3\rangle_{\mathbf{m}}$ in the following way: As explained in Section 2.1.4, to construct a minimal full configuration with terminal vertices X, Y and Z it is sufficient to provide a distribution on $\underline{X} \times \underline{Y} \times \underline{Z}$ with the correct marginals. We do this by setting

$$p(x,y,z) := \frac{p_U(x,y) \cdot p_V(y,z)}{p_Y(y)}$$

It is straightforward to check that the appropriate restriction of the full configuration defined in the above manner is indeed the original two-tents configuration. Essentially, to extend we need to provide a relationship (coupling) between spaces X and Z and we do it by declaring X and Z independent relative to Y. This is an instance of operation called *adhesion*, see [Mat07].

If we call the top vertex in the full configuration W, the entropies achieve equality in the Shannon inequality, that is

$$\mathcal{E}nt(U) + \mathcal{E}nt(V) - \mathcal{E}nt(W) - \mathcal{E}nt(Y) = 0.$$

Adhesion provides a right inverse $\mathbf{ad}$ to the restriction functor M^* described in Section 2.4.2

$$\begin{tikzcd} \mathbf{Prob}\langle\mathbf{\Lambda}_3\rangle_{\mathbf{m}} \arrow[r, "M^*"] & \mathbf{Prob}\langle\mathbf{M}_2\rangle_{\mathbf{m}} \arrow[l, bend left, dashed, "\mathbf{ad}"] \end{tikzcd}$$

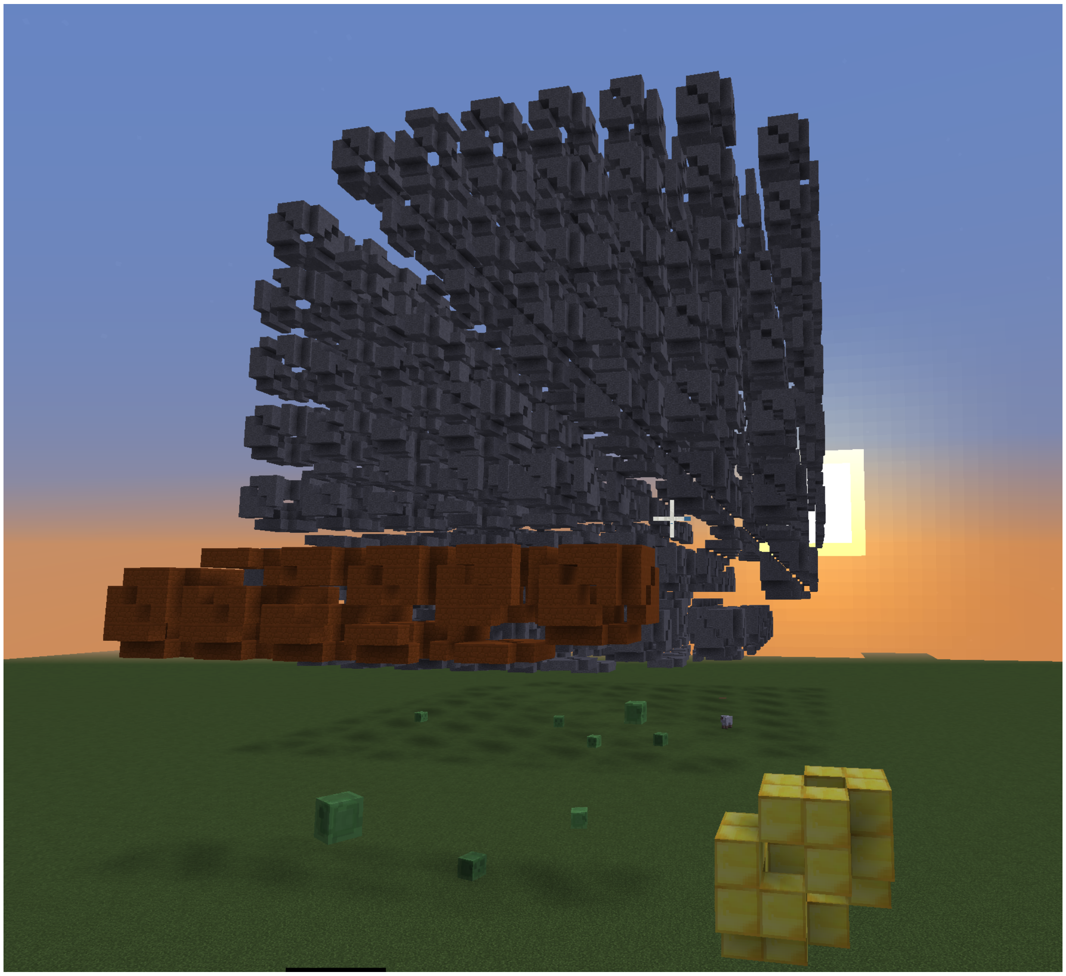

FIGURE 2. Examples of homogeneous configurations

It is important to note though, that the map **ad** is not functorial and, in fact, no functorial inverse of M^* exists.

2.6. **Homogeneous configurations.** A configuration $\mathcal{X} \in \mathbf{Prob}\langle\mathbf{G}\rangle$ modeled on some diagram category $\mathbf{G}$ is called *homogeneous* if its automorphism group $\operatorname{Aut}(\mathcal{X})$ acts transitively on every probability space in $\mathcal{X}$. Three examples of homogeneous configurations were given in the introduction. Other examples of a homogeneous configurations (of combinatorial type $\mathbf{\Lambda}_3$) are shown in Figure 2. The subcategory of all homogeneous configurations modeled on $\mathbf{G}$ will be denoted $\mathbf{Prob}\langle\mathbf{G}\rangle_{\mathbf{h}}$.

In fact, for $\mathcal{X}$ to be homogeneous it is sufficient that the $Aut(\mathcal{X})$ acts transitively on every initial space in $\mathcal{X}$. Thus, if $\mathcal{X}$ is complete with initial space X_0, to check homogeneity it is sufficient to check the transitivity of the action of the symmetries of $\mathcal{X}$ on X_0.

By functoriality of the restriction operator, any restriction of a homogeneous configuration is also homogeneous. In other words, if $R : \mathbf{G} \to \mathbf{G}'$ is a functor and

$$R^* : \mathbf{Prob}\langle\mathbf{G}'\rangle \to \mathbf{Prob}\langle\mathbf{G}\rangle$$

is the associated restriction operator, then

$$R^*\big(\mathbf{Prob}\langle\mathbf{G}'\rangle_{\mathbf{h}}\big) \subset \mathbf{Prob}\langle\mathbf{G}\rangle_{\mathbf{h}}$$

In particular, all the individual spaces of a homogeneous configuration are homogeneous

$$\mathbf{Prob}\langle\mathbf{G}\rangle_{\mathbf{h}} \subset \mathbf{Prob}_{\mathbf{h}}\langle\mathbf{G}\rangle$$

However homogeneity of the whole of the configuration is a stronger property than homogeneity of the individual spaces in the configuration, thus in general

$$\mathbf{Prob}\langle\mathbf{G}\rangle_{\mathbf{h}} \subsetneq \mathbf{Prob}_{\mathbf{h}}\langle\mathbf{G}\rangle$$

A single probability space is homogeneous if and only if there is a representative in its isomorphism class with uniform measure and the same holds true for chain configurations, for the co-fan or any other configuration that does not contain a two-fan. However, for more complex configurations, for example for two-fans, no such simple description is available.

2.6.1. *Universal construction of homogeneous configurations.* Examples of homogeneous configurations could be constructed in the following manner. Suppose Γ is a finite group and $\{H_i\}$ is a collection of subgroups. Consider a collection of sets $\underline{X}_i := \Gamma/H_i$ and consider a natural surjection $f_{ij} : \underline{X}_i \to \underline{X}_j$ whenever H_i is a subgroup of H_j. Equipping each $\underline{X}_i$ with the uniform distribution one can turn the configuration of sets $\{\underline{X}_i; f_{ij}\}$ into a homogeneous configuration. It will be complete if there is a smallest subgroup (under inclusion) among H_i's.

Such a configuration will be complete and minimal, if together with any pair of groups H_i and H_j in the collection, their intersection $H_i \cap H_j$ also belongs to the collection $\{H_i\}$.

In fact, any homogeneous configuration arises this way. Suppose configuration $\mathcal{X} = \{X_i; f_{ij}\}$ is homogeneous, then we set $\Gamma = \mathrm{Aut}(\mathcal{X})$ and choose a collection of points $x_i \in X_i$ such that $f_{ij}(x_i) = x_j$ and denote by $H_i := \mathrm{Stab}(x_i) \subset \Gamma$. Then, if one applies the construction of the previous paragraph to Γ, with the collection of subgroups $\{H_i\}$, one recovers the original configuration $\mathcal{X}$.

2.7. **Conditioning.** Suppose a configuration $\mathcal{X}$ contains a fan

$$\mathcal{F} = \Big(X \xleftarrow{f} Z \xrightarrow{g} Y\Big)$$

Given a point $x \in X$ with a non-zero weight one may consider *conditional probability distributions* $p_Z(\,\cdot\,\lfloor x)$ on $\underline{Z}$, and $p_Y(\,\cdot\,\lfloor x)$ on $\underline{Y}$. The distribution $p_Z(\,\cdot\,\lfloor x)$ is supported on $f^{-1}(x)$ and is given by

$$p_Z(z\lfloor x) = \frac{p_Z(z)}{p_X(x)}$$

The distribution $p_Y(\,\cdot\,\lfloor x)$ is the pushforward of $p_Z(\,\cdot\,\lfloor x)$ under g

$$p_Y(\,\cdot\,\lfloor x) = g_* p_Z(\,\cdot\,\lfloor x)$$

Recall that if $\mathcal{F}$ is minimal, the underlying set of Z can be assumed to be the product $\underline{X} \times \underline{Y}$. In that case

$$p_Y(y\lfloor x) = \frac{p_Z(x,y)}{p_X(x)}$$

We denote the corresponding space $Y\lfloor x := \big(\underline{Y}, p_Y(\,\cdot\,\lfloor x)\big)$, as discussed at the end of Section 1.2.1.

Under some assumptions it is possible to condition a whole sub-configuration of $\mathcal{X}$. More specifically, if a configuration $\mathcal{X}$ contains a sub-configuration $\mathcal{Y}$ and a probability space X satisfying the condition that

> for every Y in $\mathcal{Y}$ there is a fan in $\mathcal{X}$ with terminal vertices X and Y,

then we may condition the whole of $\mathcal{Y}$ on $x \in X$ given that $p_X(x) > 0$.

For $x \in X$ with positive weight we denote by $\mathcal{Y}\lfloor x$ the configuration of spaces in $\mathcal{Y}$ conditioned on $x \in X$. The configuration $\mathcal{Y}\lfloor x$ has the same combinatorial type as $\mathcal{Y}$ and will be called the *slice* of $\mathcal{Y}$ over $x \in X$. Note that the space X itself may or may not belong to $\mathcal{Y}$. The conditioning $\mathcal{Y}\lfloor x$ may depend on the choice of a fan between $\mathcal{Y}$ and X, however when $\mathcal{X}$ is complete the conditioning $\mathcal{Y}\lfloor x$ is well-defined and is independent of the choice of fans.

Suppose now that there are two subconfiguration $\mathcal{Y}$ and $\mathcal{Z}$ in $\mathcal{X}$ and in addition $\mathcal{Z}$ is a constant configuration, $\mathcal{Z} = Z^{\mathbf{G}'}$ for some diagram category $\mathbf{G}'$. Let $z \in \underline{Z}$, then $\mathcal{Y}\lfloor z$ is well defined and is independent of the choice of the space in $\mathcal{Z}$, the element of which z is to be considered.

If $\mathcal{X}$ is homogeneous, then $\mathcal{Y}\lfloor x$ is also homogeneous and its isomorphism class does not depend on the choice of $x \in \underline{X}$.

2.8. **Entropy.** For a $\mathbf{G}$-configuration $\mathcal{X} = \{X_i, f_{ij}\}$ define the entropy function

$$\mathcal{E}nt_* : \mathbf{Prob}\langle\mathbf{G}\rangle \to \mathbb{R}^{[\![G]\!]}, \quad \mathcal{E}nt_* : \mathcal{X} = \{X_i, f_{ij}\} \mapsto \big(\mathcal{E}nt(X_i)\big) \in \mathbb{R}^{[\![G]\!]}$$

It will be convenient for us to equip the target $\mathbb{R}^{[\![\mathbf{G}]\!]}$ with the ℓ^1-norm. Thus

$$|\,\mathcal{E}nt_*(\mathcal{X})|_1 = \sum_{i=1}^{[\![\mathbf{G}]\!]} \mathcal{E}nt(X_i)$$

If $\mathcal{X}$ is a complete $\mathbf{G}$-configuration with initial space X_0, then by Shannon inequality (1.2) there is an obvious estimate

$$\mathcal{E}nt(X_0) \leq |\,\mathcal{E}nt_*(\mathcal{X})|_1 \leq [\![\mathcal{X}]\!] \cdot \mathcal{E}nt(X_0)$$

3. DISTRIBUTIONS AND TYPES

In this section we recall some elementary inequalities for (relative) entropies and the total variation distance for distributions on finite sets. Furthermore, we generalize the notion of a probability distribution on a set to a distribution on a configuration of sets. Finally, we give a perspective on the theory of types, and also introduce types in the context of complete configurations.

3.1. **Distributions.**

3.1.1. *Single probability spaces.* For a finite set S we denote by ΔS the collection of all probability distributions on S. It is a unit simplex in the real vector space $\mathbb{R}^S$. We often use the fact that it is a compact, convex set, whose interior points correspond to fully supported probability measures on S.

For $\pi_1, \pi_2 \in \Delta S$ denote by $|\pi_1 - \pi_2|_1$ the total variation of the signed measure $(\pi_1 - \pi_2)$ and define the entropy of the distribution π_1 by

$$h(\pi_1) := -\sum_{x \in X} \pi_1(x) \ln \pi_1(x) \tag{3.1}$$

If, in addition, π_2 lies in the interior of ΔS define the relative entropy by

$$D(\pi_1 \,\|\, \pi_2) := \sum_{x \in X} \pi_1(x) \ln \frac{\pi_1(x)}{\pi_2(x)} \tag{3.2}$$

The entropy of a probability space is often defined through formula (3.1). It is a standard fact, and can be verified with the help of Lemma 3.2 below, that for $\pi \in \Delta S$ holds

$$h(\pi) = \mathcal{E}nt(S, \pi) \tag{3.3}$$

which justifies the name "entropy" for the function $h : \Delta S \to \mathbb{R}$.

Define a *divergence ball* of radius $\varepsilon > 0$ centered at $\pi \in \operatorname{Interior} \Delta S$ as

$$B_\varepsilon(\pi) := \{\pi' \in \Delta S \,\|\, D(\pi' \,\|\, \pi) \le \varepsilon\} \tag{3.4}$$

For a fixed π and $\varepsilon \ll 1$ the ball $B_\varepsilon(\pi)$ also lies in the interior of ΔS.

Lemma 3.1. *Let S be a finite set, then*

(i) *For any $\pi_1, \pi_2 \in \Delta S$, Pinsker's inequality holds*
$$|\pi_1 - \pi_2|_1 \le \sqrt{2D(\pi_1 \,\|\, \pi_2)}$$

(ii) *For any $\pi_2 \in \operatorname{Interior} \Delta S$ there exists a positive constant $C = C_{\pi_2}$ such that for any $\pi_1 \in \Delta S$, holds*
$$|\pi_1 - \pi_2|_1 \ge C\sqrt{D(\pi_1 \,\|\, \pi_2)}$$

(iii) *Suppose π is a point in the interior of ΔS and $r > 0$ is such that $B_r(\pi)$ also lies in the interior of ΔS. There exist a constant $C = C_{\pi,r}$ such that for any $\varepsilon \le r$ holds*
$$\max\{|h(\pi_1) - h(\pi_2)| \,\|\, \pi_1, \pi_2 \in B_\varepsilon(\pi)\} \le C\sqrt{\varepsilon}$$

⊠

The first claim of the Lemma, Pinsker's inequality, is a well-known inequality in for instance information theory, and a proof can be found in [CT91].

The second claim follows from the fact that for the fixed $\pi_2 \in \operatorname{Interior} \Delta S$ the relative entropy as the function of the first argument is bounded, smooth on the interior of the simplex and has a minimum at π_2.

To prove the last claim, note that the entropy function h is smooth in the interior of the simplex. Then this last claim follows from the first claim.

3.1.2. *Distributions on configurations.* A map $f : S \to S'$ between two finite sets induces an affine map $f_* : \Delta S \to \Delta S'$.

For a configuration of sets $\mathcal{S} = \{S_i; f_{ij}\}$ we define the *space of distributions on the configuration* $\mathcal{S}$ by

$$\Delta\mathcal{S} := \left\{ (\pi_i) \in \prod_i \Delta S_i \,\middle\|\, (f_{ij})_* \pi_i = \pi_j \right\}$$

Essentially, an element of $\Delta\mathcal{S}$ is a collection of distributions on the sets S_i in $\mathcal{S}$ that is consistent with respect to the maps f_{ij}. The consistency conditions $(f_{ij})_*\pi_i = \pi_j$ form a collection of linear equations with integer coefficients with respect to the standard convex coordinates in $\prod \Delta S_i$. Thus, $\Delta\mathcal{S}$ is a rational affine subspace in the product of simplices. In particular, $\Delta\mathcal{S}$ has a convex structure.

If $\mathcal{S}$ is complete with initial set S_0, then specifying a distribution $\pi_0 \in \Delta S_0$ uniquely determines distributions on all of the S_i's by setting $\pi_i := (f_{0i})_*\pi_0$. In such a situation we have

$$\Delta\mathcal{S} \cong \Delta S_0 \tag{3.5}$$

If $\mathcal{S}$ is not complete and $S_0, \ldots, S_k$ is a collection of its initial sets, then $\Delta\mathcal{S}$ is isomorphic to an affine subspace of the product $\Delta S_0 \times \cdots \times \Delta S_k$ cut out by linear equations with integer coefficients corresponding to co-fans in $\mathcal{S}$ with initial sets among $S_0, \ldots, S_k$.

To simplify notation, for a probability space X or a configuration $\mathcal{X}$ we will write

$$\Delta X := \Delta \underline{X}$$
$$\Delta \mathcal{X} := \Delta \underline{\mathcal{X}}$$

We now discuss briefly the theory of types. Types are special subspaces of tensor powers that consist of seqences with the same "empirical distribution" as explained in details below. For a more detailed discussion the reader is referred to [CT91] and [Csi98]. We generalize the theory of types to complete configurations of sets and complete configurations of probability spaces.

The theory of types for configurations, that are not complete, is more complex and will be addressed in a subsequent article.

3.2. **Types for single probability spaces.** Let S be a finite set. For $n \in \mathbb{N}$ denote also

$$\Delta^{(n)} S := \Delta S \cap \frac{1}{n}\mathbb{Z}^S$$

a collection of rational points in ΔS with denominator n. (We say that a rational number $r \in \mathbb{Q}$ has denominator $n \in \mathbb{N}$ if $r \cdot n \in \mathbb{Z}$)

Define the *empirical distribution map* $\mathbf{q} : S^n \to \Delta S$, that sends $(s_i)_{i=1}^n = \mathbf{s} \in S^n$ to the empirical distribution $\mathbf{q}(\mathbf{s}) \in \Delta S$ given by

$$\mathbf{q}(\mathbf{s})(a) = \frac{1}{n} \cdot \big| \{ i \,\|\, s_i = a \} \big| \quad \text{for any } a \in S$$

Clearly the image of $\mathbf{q}$ lies in $\Delta^{(n)} S$.

For $\pi \in \Delta^{(n)} S$, the space $T_\pi^{(n)} S := \mathbf{q}^{-1}(\pi)$ equipped with the uniform measure is called a *type* over π. The symmetric group $\mathbb{S}_n$ acts on $S^{\otimes n}$ by permuting the coordinates. This action leaves the empirical distribution invariant and therefore could be restricted to each type, where it acts transitively. Thus, for $\pi \in \Delta^{(n)} S$ the probability space $(T_\pi^{(n)} S, u)$ with u being a uniform ($\mathbb{S}_n$-invariant) distribution, is a homogeneous space.

Suppose $X = (\underline{X}, p)$ is a probability space. Let τ_n be the pushforward of $p^{\otimes n}$ under the empirical distribution map $\mathbf{q} : \underline{X}^n \to \Delta X$. Clearly $\operatorname{supp} \tau_n \subset \Delta^{(n)} X$, thus $(\Delta X, \tau_n)$ is a finite probability space. Therefore we have a reduction

$$\mathbf{q} : X^{\otimes n} \to (\Delta X, \tau_n)$$

which we call the *empirical reduction*. If $\pi \in \Delta^{(n)} X$ is such that $\tau_n(\pi) > 0$, then

$$T_\pi^{(n)} \underline{X} = X^{\otimes n} \lfloor \pi \tag{3.6}$$

In particular, it follows that the right-hand side does not depend on the probability p on X as long as π is "compatible" to it.

The following lemma records some standard facts about types, which can be checked by elementary combinatorics and found in [CT91].

Lemma 3.2. *Let X be a probability space and $\mathbf{x} \in X^{\otimes n}$, then*

(i)
$$|\Delta^{(n)} X| = \binom{n + |X|}{|X|} \leq \mathbf{e}^{|X| \cdot \ln(n+1)}$$

(ii)
$$p^{\otimes n}(\mathbf{x}) = \mathbf{e}^{-n\left[h\left(\mathbf{q}(\mathbf{x})\right) + D\left(\mathbf{q}(\mathbf{x}) \,\|\, p\right)\right]}$$

(iii)
$$\mathbf{e}^{n \cdot h(\pi) - |X| \cdot \ln(n+1)} \leq |T^{(n)}_\pi \underline{X}| \leq \mathbf{e}^{n \cdot h(\pi)}$$

(iv)
$$\mathbf{e}^{-n \cdot D(\pi \,\|\, p) - |X| \cdot \ln(n+1)} \leq \tau_n(\pi) = p^{\otimes n}(T^{(n)}_\pi \underline{X}) \leq \mathbf{e}^{-n \cdot D(\pi \,\|\, p)}$$

⊠

If $X = (\underline{X}, p_X)$ is a probability space with rational probability distribution with denominator n, then the type over p_X will be called the true type of X

$$T^{(n)} X := T^{(n)}_{p_X} \underline{X}$$

As a corollary to Lemma 3.2 and equation (3.3) we obtain the following.

Corollary 3.3. *For a finite set S and $\pi \in \Delta^{(n)} S$ holds*

$$n \cdot h(\pi) - |S| \cdot \ln(n+1) \leq \mathcal{E}nt(T^{(n)}_\pi S) \leq n \cdot h(\pi)$$

In particular, for a finite probability space $X = (S, p)$ with a rational distribution p with denominator n holds

$$n \cdot \mathcal{E}nt(X) - |S| \cdot \ln(n+1) \leq \mathcal{E}nt(T^{(n)} X) \leq n \cdot \mathcal{E}nt(X)$$

⊠

The following important theorem is known as Sanov's theorem. It can be derived from Lemma 3.2 and found in [CT91].

Theorem 3.4. (Sanov's Theorem) *Let X be a finite probability space and let $\mathbf{q} : X^{\otimes n} \to (\Delta X, \tau_n)$ be the empirical reduction. Then for every $r > 0$,*

$$\tau_n(\Delta X \backslash B_r(p)) \leq \mathbf{e}^{-n \cdot r + |X| \cdot \ln(n+1)}$$

where $B_r(p)$ is the divergence ball (relative entropy ball) defined in (3.4). ⊠

3.3. **Types for complete configurations.** In this subsection we generalize the theory of types for configurations modeled on a complete category. The theory for a non-complete configurations is more complex and will be addressed in our future work. We will give three equivalent definitions of a type for a complete configuration, each of which will be useful in its own way. Before we describe the three approaches we need some preparatory material.

Lemma 3.5. *Given a diamond configuration of probability spaces*

$$\mathcal{D} = \begin{pmatrix} X & \xrightarrow{f} & Y \\ \downarrow\rho_1 & & \downarrow\rho_2 \\ A & \xrightarrow{g} & B \end{pmatrix}$$

the following two conditions are equivalent

(i) *The minimal reduction of the diamond $\mathcal{D}$ is isomorphic to the adhesion of its co-fan or equivalently the following independence condition holds*

$$A \perp\!\!\!\perp Y \lfloor B$$

(ii) *For any $a, a' \in \underline{A}$ such that $f(a) = f(a') = b \in B$ holds*

$$Y\lfloor a = Y\lfloor a' = Y\lfloor b$$

⊠

Suppose $\mathcal{D}$ is the diamond as in the Lemma 3.5. The top row $X \to Y$ is a two-chain subconfiguration of $\mathcal{D}$ and we can consider a conditioning by an element $a \in \underline{A}$

$$X\lfloor a \to Y\lfloor a$$

If $\mathcal{D}$ satisfies any of two conditions in Lemma 3.5, then $Y\lfloor a = Y\lfloor b$ for $b = g(a)$. Thus, we constructed a reduction

$$X\lfloor a \to Y\lfloor b$$

Suppose we have a reduction $f : X \to Y$ between a pair of probability spaces. Then for any $n \in \mathbb{N}$ there is an induced reduction $f_* : (\Delta X, \tau_n) \to (\Delta Y, \tau_n)$ that can be included in the following diamond configuration

$$\begin{array}{ccc} X^{\otimes n} & \xrightarrow{f^{\otimes n}} & Y^{\otimes n} \\ \downarrow \mathbf{q} & & \downarrow \mathbf{q} \\ (\Delta X, \tau_n) & \xrightarrow{f_*} & (\Delta Y, \tau_n) \end{array}$$

that satisfies conditions in Lemma 3.5. It means that there is a reduction

$$Tf : T^{(n)}_{\pi} X \to T^{(n)}_{\pi'} Y$$

for $\pi \in \Delta^{(n)} X$ and $\pi' = f_* \pi \in \Delta^{(n)} Y$.

Now we are ready to give the definitions of types. Let $\mathcal{X} \in \mathbf{Prob}\langle\mathbf{G}\rangle$ be a complete configuration, $\mathcal{X} = \{X_i; f_{ij}\}$ with initial space X_0 and let $\pi \in \Delta^{(n)}\mathcal{X}$.

3.3.1. *Type of a configuration as the configuration of types.* Define the type $T^{(n)}_{\pi} \underline{\mathcal{X}}$ as the $\mathbf{G}$-configuration, whose individual spaces are types of the individual spaces of $\mathcal{X}$ over the corresponding push-forwards of π

$$T^{(n)}_{\pi} \underline{\mathcal{X}} := \left\{ T^{(n)}_{\pi_i} \underline{X}_i ; T f_{ij} \right\}$$

3.3.2. *Types as $\mathbb{S}_n$-orbits in the tensor power.* By a *section* in $\mathcal{X}$ we mean a consistent collection of points

$$\mathbf{x} = (x_0, \ldots, x_{[\![\mathcal{X}]\!]-1}) \in \prod_{i=0}^{[\![\mathcal{X}]\!]-1} \underline{X}_i$$

such that $f_{ij} x_i = x_j$, whenever f_{ij} is defined. For any j define the projection $\rho_j : \prod \underline{X}_i \to \underline{X}_j$, so that $\rho_j(\mathbf{x}) = x_j$.

The symmetric group $\mathbb{S}_n$ acts on the collection of sections in the tensor power $\mathcal{X}^{\otimes n}$, by permuting the coordinates. Let $\mathcal{T} \subset \prod \underline{X}_i$ be an orbit of the action such that $\rho_0(\mathcal{T}) = T_\pi^{(n)} \underline{X}_0$. Suppose that the pair (i,j) is such, that f_{ij} is defined. Since $f_{ij}^{\otimes n} : X_i^{\otimes n} \to X_j^{\otimes n}$ is $\mathbb{S}_n$-equivariant, we have a map

$$Tf : \rho_i(\mathcal{T}) \to \rho_j(\mathcal{T})$$

We can turn $\mathcal{T}$ into a $\mathbf{G}$-configuration, which we will call a type of $\mathcal{X}$

$$\hat{T}_\pi^{(n)} \underline{\mathcal{X}} := \{\rho_i(\mathcal{T}), Tf_{ij}\}$$

where π is the value of the impirical distribution on $\rho_0(\mathcal{T})$.

Since the initial space in $\hat{T}_\pi^{(n)} \underline{\mathcal{X}}$ coincides with the initial space in $T_\pi^{(n)} \underline{\mathcal{X}}$ and all the reductions coincide, we conclude that

$$\hat{T}_\pi^{(n)} \underline{\mathcal{X}} = T_\pi^{(n)} \underline{\mathcal{X}}$$

3.3.3. *Type as conditionining of the tensor power.* We can extend $\mathcal{X}^{\otimes n}$ to a configuration $\hat{\mathcal{X}}$ by adding $(\Delta\mathcal{X}, \tau_n)$ and the empirical reduction $X_0^{\otimes n} \to (\Delta\mathcal{X}, \tau_n)$.

Let $\pi \in \Delta\mathcal{X}$ with $\tau_n(\pi) > 0$ and recall $\Delta X_0 \cong \Delta\mathcal{X}$. We may now define $\mathcal{X}^{\otimes n} \lfloor \pi$ as in Section 2.7. Define a type of $\mathcal{X}$ over $\pi \in \overset{(n)}{\Delta} \mathcal{X}$ by

$$\check{T}_\pi^{(n)} \underline{\mathcal{X}} := \mathcal{X}^{\otimes n} \lfloor \pi$$

By definition, it holds that

$$X_0^{\otimes n} \lfloor \pi = T_\pi^{(n)} X_0$$

Let $\pi_i = (f_{0i})_* \pi$. Using Lemma 3.5 and discussion thereafter we conclude that

$$X_i^{\otimes n} \lfloor \pi = X_i^{\otimes n} \lfloor \pi_i$$

and therefore

$$\check{T}_\pi^{(n)} \underline{\mathcal{X}} = T_\pi^{(n)} \underline{\mathcal{X}}$$

3.3.4. *The empirical two-fan.* We construct a two-fan of $\mathbf{G}$-configurations with terminal vertices $\mathcal{X}^{\otimes n}$ and $(\Delta\mathcal{X}, \tau_n)^{\mathbf{G}}$ (the constant $\mathbf{G}$-configuration in which every probability space is $(\Delta\mathcal{X}, \tau_n)$, and every reduction is an identity)

$$\mathcal{R}_n(\mathcal{X}) = \left(\begin{tikzcd} & \tilde{\mathcal{X}}^{(n)} \arrow[dl] \arrow[dr] & \\ \mathcal{X}^{\otimes n} & & (\overset{(n)}{\Delta} \mathcal{X}, \tau_n)^{\mathbf{G}} \end{tikzcd} \right) \tag{3.7}$$

With the help of Lemma 2.1, we construct $\mathcal{R}_n(\mathcal{X})$ as the minimal reduction of the two-fan of $\mathbf{G}$-configurations

$$\begin{tikzcd} & (X_0^{\otimes n})^{\mathbf{G}} \arrow[dl, "f_{0*}^n"'] \arrow[dr, "\mathbf{q}^{\mathbf{G}}"] & \\ \mathcal{X}^{\otimes n} & & (\Delta\mathcal{X}, \tau_n)^{\mathbf{G}} \end{tikzcd}$$

Let $\pi \in (\Delta\mathcal{X}, \tau_n)$ with $\tau_n(\pi) > 0$. Then within $\mathcal{R}_n(\mathcal{X})$ holds

$$\mathcal{X}^{\otimes n} \lfloor \pi = T_\pi^{(n)} \mathcal{X}$$

For every $n \in \mathbb{N}$ and $\pi \in \Delta^{(n)} \underline{X}_0$ the type $T_\pi^{(n)} \mathcal{X}$ is a homogeneous configuration. Suppose that a complete configuration $\mathcal{X}$ is such that the probability distribution p_0 on the initial set is rational with the denominator n, then we call $T_p^{(n)} \underline{\mathcal{X}}$ the *true type* of $\mathcal{X}$ and denote

$$T^{(n)} \mathcal{X} := T_{p_0}^n \underline{\mathcal{X}}$$

4. THE KOLMOGOROV-SINAI DISTANCE

We turn the space of configurations into a pseudo-metric space by introducing the intrinsic Kolmogorov-Sinai distance and asymptotic Kolmogorov-Sinai distance. For brevity, we will usually call it the Kolmogorov distance and asymptotic Kolmogorov distance. The intrinsic Kolmogorov-Sinai distance is obtained by taking an infimum of the shared information distance over all possible joint distributions on two probability spaces. The name is justified by the fact that the shared information distance (not under this name) appears in the proof of the theorem about generating partitions for ergodic systems by Kolmogorov and Sinai, see for example [Sin76]. Note that the Kolmogorov distance in statistics refers to a different notion.

4.1. **Kolmogorov distance and asymptotic Kolmogorov distance.**

4.1.1. *Kolmogorov distance in the case of single probability spaces.* For a two-fan $\mathcal{F} = (X \leftarrow Z \rightarrow Y)$ define a "distance" $\mathrm{kd}(\mathcal{F})$ between probability spaces X and Y with respect to $\mathcal{F}$ by

$$\begin{aligned} \mathrm{kd}(\mathcal{F}) &:= \mathcal{E}nt(Z \lfloor Y) + \mathcal{E}nt(Z \lfloor X) \\ &= 2\,\mathcal{E}nt(Z) - \mathcal{E}nt(X) - \mathcal{E}nt(Y) \end{aligned}$$

Essentially $\mathrm{kd}(\mathcal{F})$ measures the deviation of the statistical map defined by $\mathcal{F}$ from being a deterministic bijection between X and Y.

The minimal reduction $\mathcal{F}'$ of $\mathcal{F}$ satisfies

$$\mathrm{kd}(\mathcal{F}') \leq \mathrm{kd}(\mathcal{F}) \tag{4.1}$$

If the two-fan $\mathcal{F}$ is minimal the "distance" $\mathrm{kd}(\mathcal{F})$ can also be calculated by

$$\mathrm{kd}(\mathcal{F}) = h(p_X) + h(p_Y) - 2D(p_Z \,\|\, p_X \otimes p_Y),$$

where h and D are respectively the entropy and relative entropy functions defined in (3.1) and (3.2).

For a pair of probability spaces X, Y define the *intrinsic Kolmogorov-Sinai distance* as

$$\mathbf{k}(X,Y) := \inf\left\{\mathrm{kd}(\mathcal{F}) \,\|\, \mathcal{F} = (X \leftarrow Z \to Y) \text{ is a two-fan}\right\}$$

The optimization takes place over all two-fans with terminal spaces X and Y. In view of inequality (4.1) one could as well optimize over the space of *minimal* two-fans, which we will also refer to as *couplings* between X and Y. The tensor product of X and Y trivially provides a coupling and the set of couplings is compact, therefore an optimum is always achieved and it is finite.

The bivariate function $\mathbf{k} : \mathbf{Prob} \times \mathbf{Prob} \to \mathbb{R}_{\geq 0}$ defines a notion of pseudo-distance and it vanishes exactly on pairs of isomorphic probability spaces. This follows directly from the Shannon inequality (1.2), and a more general statement will be proven in Proposition 4.1 below.

4.1.2. *Kolmogorov distance for complete configurations.* The definition of Kolmogorov distance for complete configurations repeats almost literally the definition for single spaces. We fix a complete diagram category $\mathbf{G}$ and will be considering configurations from $\mathbf{Prob}\langle\mathbf{G}\rangle$.

Consider three configurations $\mathcal{X} = \{X_i, f_{ij}\}$, $\mathcal{Y} = \{Y_i, g_{ij}\}$ and $\mathcal{Z} = \{Z_i, h_{ij}\}$ from $\mathbf{Prob}\langle\mathbf{G}\rangle$. Recall that a two-fan $\mathcal{F} = (\mathcal{X} \leftarrow \mathcal{Z} \to \mathcal{Y})$ is a $\mathbf{G}$-configuration of two-fans

$$\mathcal{F}_i = (X_i \leftarrow Z_i \to Y_i)$$

Define

$$\begin{aligned}\mathrm{kd}(\mathcal{F}) &:= \sum_i \mathrm{kd}(\mathcal{F}_i)\\ &= \sum_i \big(2\,\mathcal{E}nt(Z_i) - \mathcal{E}nt(X_i) - \mathcal{E}nt(Y_i)\big)\end{aligned}$$

The quantity $\mathrm{kd}(\mathcal{F})$ vanishes if and only if the fan $\mathcal{F}$ provides isomorphisms between all individual spaces in $\mathcal{X}$ and $\mathcal{Y}$ that commute with the inner structure of the configurations, that is, it provides an isomorphism between $\mathcal{X}$ and $\mathcal{Y}$ in $\mathbf{Prob}\langle\mathbf{G}\rangle$.

The *intrinsic Kolmogorov-Sinai distance between configurations* is defined in analogy with the case of single probability spaces

$$\mathbf{k}(\mathcal{X},\mathcal{Y}) := \inf\left\{\mathrm{kd}(\mathcal{F}) \,\|\, \mathcal{F} = (\mathcal{X} \leftarrow \mathcal{Z} \to \mathcal{Y})\right\}$$

where the infimum is over all two-fans of $\mathbf{G}$-configurations with terminal vertices $\mathcal{X}$ and $\mathcal{Y}$.

The following proposition records that the intrinsic Kolmogorov distance is in fact a pseudo-distance on $\mathbf{Prob}\langle\mathbf{G}\rangle$, provided $\mathbf{G}$ is a complete diagram category (that is when $\mathbf{G}$ has a unique initial space).

Proposition 4.1. ↓ *Let* $\mathbf{G}$ *be a complete diagram category. Then the bivariate function*

$$\mathbf{k} : \mathbf{Prob}\langle\mathbf{G}\rangle \times \mathbf{Prob}\langle\mathbf{G}\rangle \to \mathbb{R}$$

is a pseudo-distance on $\mathbf{Prob}\langle\mathbf{G}\rangle$.
Moreover, two configurations $\mathcal{X},\mathcal{Y}\in\mathbf{Prob}\langle\mathbf{G}\rangle$ *satisfy* $\mathbf{k}(\mathcal{X},\mathcal{Y})=0$ *if and only if* $\mathcal{X}$ *is isomorphic to* $\mathcal{Y}$ *in* $\mathbf{Prob}\langle\mathbf{G}\rangle$. ⊠

The idea of the proof is very simple. In the case of single probability spaces X,Y,Z a coupling between X and Z can be constructed from a coupling between X and Y and a coupling between Y and Z by adhesion on Y, see Section 2.5. The triangle inequality then follows from a Shannon inequality. However, since we are dealing with configurations the combinatorial structure requires careful treatment. Therefore, we provide a detailed proof on page 61.

It is important to note, that the proof uses the fact that $\mathbf{G}$ is complete. In fact, even though the definition of $\mathbf{k}$ could be easily extended to some bivariate function on the space of configurations of any fixed combinatorial type, it fails to satisfy the triangle inequality in general, because the composition of couplings requires completeness of $\mathbf{G}$.

4.1.3. *The asymptotic Kolmogorov-Sinai distance.* Let $\mathbf{G}$ be a complete diagram category. We define the *asymptotic Kolmogorov-Sinai distance* between two configurations $\mathcal{X},\mathcal{Y}\in\mathbf{Prob}\langle\mathbf{G}\rangle$ by

$$\boldsymbol{\kappa}(\mathcal{X},\mathcal{Y})=\lim_{n\to\infty}\frac{1}{n}\mathbf{k}(\mathcal{X}^{\otimes n},\mathcal{Y}^{\otimes n}). \tag{4.2}$$

We will show in Corollary 4.5, that the sequence

$$n\mapsto\mathbf{k}(\mathcal{X}^{\otimes n},\mathcal{Y}^{\otimes n})$$

is subadditive, and therefore the limit in the definition (4.2) of $\boldsymbol{\kappa}(\mathcal{X},\mathcal{Y})$ always exists and for all $n\in\mathbb{N}$ holds

$$\boldsymbol{\kappa}(\mathcal{X},\mathcal{Y})\leq\frac{1}{n}\cdot\mathbf{k}(\mathcal{X}^{\otimes n},\mathcal{Y}^{\otimes n}). \tag{4.3}$$

As a corollary of Proposition 4.1 and definition (4.2) we immediately obtain that also the asymptotic Kolmogorov-Sinai distance is a pseudo-distance on $\mathbf{Prob}\langle\mathbf{G}\rangle$.

Corollary 4.2. *Let* $\mathbf{G}$ *be a complete diagram category. Then the bivariate function*

$$\boldsymbol{\kappa}:\mathbf{Prob}\langle\mathbf{G}\rangle\times\mathbf{Prob}\langle\mathbf{G}\rangle\to\mathbb{R}$$

is a pseudo-distance on $\mathbf{Prob}\langle\mathbf{G}\rangle$ *satisfying the following homogeneity property. For any pair of configurations* $\mathcal{X},\mathcal{Y}\in\mathbf{Prob}\langle\mathbf{G}\rangle$ *and any* $n\in\mathbb{N}_0$ *holds*

$$\boldsymbol{\kappa}(\mathcal{X}^{\otimes n},\mathcal{Y}^{\otimes n})=n\cdot\boldsymbol{\kappa}(\mathcal{X},\mathcal{Y})$$

⊠

We will show in a later section, however, that there are probability spaces X and Y for which $\boldsymbol{\kappa}(X,Y)=0$ that are not isomorphic.

In the rest of this section we derive some elementary properties of the intrinsic Kolmogorov distance and the asymptotic Kolmogorov distance.

4.2. **Lipschitz property for operations.** In this section we show that certain natural operations on configurations, namely the tensor product, entropy function and restriction operator, are Lipschitz continuous. In Section 7 we will show Lipschitz continuity of certain extension operations.

4.2.1. *Tensor product.* We show that the tensor product on the space of configurations is 1-Lipschitz. Later this will allow us to give a simple description of tropical configurations, that is of points in the asymptotic cone of $\mathbf{Prob}\langle\mathbf{G}\rangle$, as limits of certain sequences of "classical" configurations.

Proposition 4.3. ↓ *Let* $\mathbf{G}$ *be a complete diagram category. Then with respect to the Kolmogorov distance on* $\mathbf{Prob}\langle\mathbf{G}\rangle$ *the tensor product*

$$\otimes : (\mathbf{Prob}\langle\mathbf{G}\rangle, \mathbf{k})^2 \to (\mathbf{Prob}\langle\mathbf{G}\rangle, \mathbf{k})$$

is 1-Lipschitz in each variable, that is, for every triple $\mathcal{X}, \mathcal{Y}, \mathcal{Y}' \in \mathbf{Prob}\langle\mathbf{G}\rangle$ *the following bound holds*

$$\mathbf{k}(\mathcal{X}\otimes\mathcal{Y}, \mathcal{X}\otimes\mathcal{Y}') \le \mathbf{k}(\mathcal{Y}, \mathcal{Y}')$$

⊠

This statement is a direct consequence of additivity of entropy with respect to the tensor product. Details can be found on page 63.

It follows directly from definition (4.2) and Proposition 4.3, that the asymptotic Kolmogorov distance enjoys a similar property.

Corollary 4.4. *Let* $\mathbf{G}$ *be a complete diagram category. Then with respect to the Kolmogorov distance on* $\mathbf{Prob}\langle\mathbf{G}\rangle$ *the tensor product*

$$\otimes : (\mathbf{Prob}\langle\mathbf{G}\rangle, \boldsymbol{\kappa})^2 \to (\mathbf{Prob}\langle\mathbf{G}\rangle, \boldsymbol{\kappa})$$

is 1-Lipschitz in each variable. ⊠

As another corollary we obtain the subadditivity properties of the intrinsic Kolmogorov distance and asymptotic Kolmogorov distance.

Corollary 4.5. *Let* $\mathbf{G}$ *be a complete diagram category and let* $\mathcal{X}, \mathcal{Y}, \mathcal{U}, \mathcal{V} \in \mathbf{Prob}\langle\mathbf{G}\rangle$*, then*

$$\mathbf{k}(\mathcal{X}\otimes\mathcal{U}, \mathcal{Y}\otimes\mathcal{V}) \le \mathbf{k}(\mathcal{X}, \mathcal{Y}) + \mathbf{k}(\mathcal{U}, \mathcal{V}).$$

and

$$\boldsymbol{\kappa}(\mathcal{X}\otimes\mathcal{U}, \mathcal{Y}\otimes\mathcal{V}) \le \boldsymbol{\kappa}(\mathcal{X}, \mathcal{Y}) + \boldsymbol{\kappa}(\mathcal{U}, \mathcal{V}).$$

⊠

It implies in particular that shifts are non-expanding maps in $(\mathbf{Prob}\langle\mathbf{G}\rangle, \mathbf{k})$ or $(\mathbf{Prob}\langle\mathbf{G}\rangle, \boldsymbol{\kappa})$.

Corollary 4.6. *Let* $\mathbf{G}$ *be a complete diagram category and* $\boldsymbol{\delta} = \mathbf{k}, \boldsymbol{\kappa}$ *be either Kolmogorov distance or asymptotic Kolmogorov distance on* $\mathbf{Prob}\langle\mathbf{G}\rangle$*. Let* $\mathcal{U} \in \mathbf{Prob}\langle\mathbf{G}\rangle$*. Then the shift map*

$$\mathcal{U}\otimes\cdot : (\mathbf{Prob}\langle\mathbf{G}\rangle, \boldsymbol{\delta}) \to (\mathbf{Prob}\langle\mathbf{G}\rangle, \boldsymbol{\delta}), \quad \mathcal{X} \mapsto \mathcal{U}\otimes\mathcal{X}$$

is a non-expanding map with respect to either Kolmogorov distance or asymptotic Kolmogorov distance. ⊠

4.2.2. *Entropy.* Recall that we defined the entropy function

$$\mathcal{E}nt_* : \mathbf{Prob}\langle\mathbf{G}\rangle \to \mathbb{R}^{[\![\mathbf{G}]\!]}$$

by evaluating the entropy of all individual spaces in a $\mathbf{G}$-configuration. The target space $\mathbb{R}^{[\![\mathbf{G}]\!]}$ will be endowed with the ℓ^1-norm with respect to the natural coordinate system. With such a choice, the entropy function is 1-Lipschitz with respect to the Kolmogorov distance on $\mathbf{Prob}\langle\mathbf{G}\rangle$.

Proposition 4.7. ↓ *Suppose* $\mathbf{G}$ *is a complete diagram category and* $\boldsymbol{\delta} = \mathbf{k}, \boldsymbol{\kappa}$ *is either Kolmogorov distance or asymptotic Kolmogorov distance on* $\mathbf{Prob}\langle\mathbf{G}\rangle$*. Then the entropy function*

$$\mathcal{E}nt_* : (\mathbf{Prob}\langle\mathbf{G}\rangle, \boldsymbol{\delta}) \to (\mathbb{R}^{[\![\mathbf{G}]\!]}, |\cdot|_1), \quad \mathcal{X} = \{X_i, f_{ij}\} \mapsto (\mathcal{E}nt X_i)_i \in \mathbb{R}^{[\![\mathbf{G}]\!]}$$

is 1-Lipschitz. ⊠

Again, the proof of the proposition above is an application of Shannon's inequality, see page 64 for details.

4.2.3. *Restrictions.* The restriction operators are also Lipschitz, as shown in the next proposition.

Proposition 4.8. ↓ *Suppose* $R : \mathbf{G}' \to \mathbf{G}$ *is a functor between two complete diagram categories and* $\boldsymbol{\delta}$ *stands for either Kolmogorov or asymptotic Kolmogorov distance. Then the restriction operator*

$$R^* : (\mathbf{Prob}\langle\mathbf{G}\rangle, \boldsymbol{\delta}) \to (\mathbf{Prob}\langle\mathbf{G}'\rangle, \boldsymbol{\delta}), \quad \mathcal{X} \mapsto \mathcal{X} \circ R$$

is Lipschitz. ⊠

As can be seen from the proof on page 64, the Lipschitz constant in the proposition above can be bounded by $[\![\mathbf{G}']\!]$. In fact, a more careful analysis provides a better bound by the maximal number of objects in $\mathbf{G}'$ that are mapped by R to a single object in $\mathbf{G}$.

4.3. **The Slicing Lemma.** The Slicing Lemma, Proposition 4.9 below, allows to estimate the Kolmogorov distance between two configurations with the integrated Kolmogorov distance between "slices", which are configurations obtained by conditioning on another probability space.

The Slicing Lemma, along with the local estimate in Section 4.4, turned out to be a very powerful tool for estimation of the Kolmogorov distance and will be used below on many occasions.

As described in Section 2.2, by a reduction of a configuration $\mathcal{X} = \{X_i, f_{ij}\}$ to a single space U we mean a collection of reductions $\{\rho_i : X_i \to U\}$ from the individual spaces in $\mathcal{X}$ to U, that commute with the reductions within $\mathcal{X}$

$$\rho_j \circ f_{ij} = \rho_i$$

Alternatively, whenever a single probability space appears together with a **G**-configuration in a commutative diagram, it should be replaced by a constant **G**-configuration.

Proposition 4.9. ↓ (Slicing Lemma) *Suppose* $\mathbf{G}$ *is a complete diagram category and we are given* $\mathcal{X}, \hat{\mathcal{X}}, \mathcal{Y}, \hat{\mathcal{Y}} \in \mathbf{Prob}\langle\mathbf{G}\rangle$ *– four* $\mathbf{G}$*-configurations and* $U, V, W \in \mathbf{Prob}$ *– probability spaces, that are included into the following three-tents configuration*

$$\begin{tikzcd}[column sep=small] & \hat{\mathcal{X}} \arrow[dl] \arrow[dr] & & W \arrow[dl] \arrow[dr] & & \hat{\mathcal{Y}} \arrow[dl] \arrow[dr] & \\ \mathcal{X} & & U & & V & & \mathcal{Y} \end{tikzcd}$$

such that the two-fan $(U \leftarrow W \rightarrow V)$ *is minimal. Then the following estimate holds*

$$\begin{aligned} \mathbf{k}(\mathcal{X}, \mathcal{Y}) \leq & \int_W \mathbf{k}(\mathcal{X}\lfloor u, \mathcal{Y}\lfloor v) \, \mathrm{d}\, p_W(u,v) \\ & + [\![\mathbf{G}]\!] \cdot \mathrm{kd}(U \leftarrow W \rightarrow V) \\ & + \sum_i \big[\mathcal{E}nt(U\lfloor X_i) + \mathcal{E}nt(V\lfloor Y_i) \big] \end{aligned}$$

⊠

The idea of the proof of the Slicing Lemma (page 65) is as follows. For every pair $(u,v) \in \underline{W}$ we consider an optimal two-fan $\mathcal{G}_{uv}$ coupling $\mathcal{X}\lfloor u$ and $\mathcal{Y}\lfloor v$. These fans have the same underlying configuration of sets. Then we construct a coupling between $\mathcal{X}$ and $\mathcal{Y}$ as a convex combination of distributions of $\mathcal{G}_{uv}$'s weighted by $p_W(u,v)$. The estimates on the resulting two-fan then imply the proposition.

Various implications of the Slicing Lemma are summarized in the next corollary.

Corollary 4.10. *Let* $\mathbf{G}$ *be a complete diagram category,* $\mathcal{X}, \mathcal{Y} \in \mathbf{Prob}\langle\mathbf{G}\rangle$ *and* $U \in \mathbf{Prob}$.

(i) *Given a "two-tents" configuration*

$$\mathcal{X} \leftarrow \hat{\mathcal{X}} \rightarrow U \leftarrow \hat{\mathcal{Y}} \rightarrow \mathcal{Y}$$

the following inequality holds

$$\mathbf{k}(\mathcal{X}, \mathcal{Y}) \leq \int_U \mathbf{k}(\mathcal{X}\lfloor u, \mathcal{Y}\lfloor u) \, \mathrm{d}\, p_U(u) + 2 \cdot [\![\mathbf{G}]\!] \cdot \mathcal{E}nt(U)$$

(ii) *Given a fan*

$$\mathcal{X} \leftarrow \hat{\mathcal{X}} \rightarrow U$$

the following inequality holds

$$\mathbf{k}(\mathcal{X}, \mathcal{Y}) \leq \int_{UV} \mathbf{k}(\mathcal{X}\lfloor u, \mathcal{Y}) \, \mathrm{d}\, p_U(u) + 2 \cdot [\![\mathbf{G}]\!] \cdot \mathcal{E}nt(U)$$

(iii) *Let* $\mathcal{X} \to U$ *be a reduction, then*

$$\mathbf{k}(\mathcal{X},\mathcal{Y}) \le \int_U \mathbf{k}(\mathcal{X}\lfloor u,\mathcal{Y})\,\mathrm{d}\,p_U(u) + [\![\mathbf{G}]\!]\cdot \mathcal{E}nt(U)$$

(iv) *For a co-fan* $\mathcal{X} \to U \leftarrow \mathcal{Y}$ *holds*

$$\mathbf{k}(\mathcal{X},\mathcal{Y}) \le \int_U \mathbf{k}(\mathcal{X}\lfloor u,\mathcal{Y}\lfloor u)\,\mathrm{d}\,p_U(u)$$

⊠

4.4. **Local estimate.** Fix a complete diagram category $\mathbf{G}$ and consider a $\mathbf{G}$-configuration of sets $\mathcal{S} \in \mathbf{Set}\langle\mathbf{G}\rangle$ with S_0 being an initial set in $\mathcal{S}$. As discussed in Section 3.1.2, the space of distributions on $\mathcal{S}$ could be identified with the space of distributions on the initial set

$$\Delta\mathcal{S} \cong \Delta S_0$$

Therefore, all $\mathbf{G}$-configurations of probability spaces with the underlying configuration of sets equal to $\mathcal{S}$ are in one-to-one correspondence with the interior points of ΔS_0. The set $\operatorname{Interior}\Delta S_0$ consists of fully supported measures on the set S_0 and carries a total variation distance, which is just an ℓ^1-distance with respect to the convex coordinates on the simplex ΔS_0. Our task presently is to compare the total variation distance with the Kolmogorov distance on the space of configurations with the fixed underlying configuration of sets.

The upper bound on Kolmogorov distance, that we derive below, has two summands. One is linear in the total variation distance with the slope proportional to the log-cardinality of S_0. The second one is super-linear in the total variation distance, but it does not depend on $\mathcal{S}$. So we have the following interesting observation: of course, the super-linear summand always dominates the linear one locally. However as the cardinality of $\mathcal{S}$ becomes large it is the linear summand that starts playing the main role.

4.4.1. *The estimate.* Suppose we are given a configuration of sets $\mathcal{S} = \{S_i, f_{ij}\} \in \mathbf{Set}\langle\mathbf{G}\rangle$ modeled on a complete diagram category $\mathbf{G}$ with the initial set S_0. We use once again the isomorphism

$$\Delta\mathcal{S} \xrightarrow{\cong} \Delta S_0$$

that sends $p \in \Delta\mathcal{S}$ to its component in the initial space $p_0 \in \Delta S_0$, while its inverse is given by $p = \{(f_{0i})_* p_0\}$. For a pair of distributions $p_0, q_0 \in \Delta S_0$ denote by $|p_0 - q_0|$ the total variation of the difference.

For $\alpha \in [0,1]$ consider a binary probability space with the weight of one of the atoms equal to α

$$\Lambda_\alpha := \Big(\{\square,\blacksquare\}\,;\, p_{\Lambda_\alpha}(\square) = 1-\alpha,\, p_{\Lambda_\alpha}(\blacksquare) = \alpha\Big)$$

Proposition 4.11. ↓ *Let $\mathcal{S} = \{S_i, f_{ij}\} \in \mathbf{Set}\langle\mathbf{G}\rangle$ be a configuration of sets modeled on a complete diagram category $\mathbf{G}$ with the initial set S_0. Let $p, q \in \Delta\mathcal{S}$ be two probability distributions. Denote $\mathcal{X} := (\mathcal{S}, p)$, $\mathcal{Y} := (\mathcal{S}, q)$ and $\alpha = \frac{1}{2}|p_0 - q_0|_1$. Then*

$$\mathbf{k}(\mathcal{X}, \mathcal{Y}) \leq 2 \cdot [\![\mathbf{G}]\!] \cdot \big(\alpha \cdot \ln|S_0| + \mathcal{E}nt(\Lambda_\alpha)\big)$$

⊠

To prove the local estimate we decompose both p and q into a convex combination of a common part $\hat{p}$ and rests p^+ and q^+. The coupling between the common parts gives no contribution to the distance, and the worst possible estimate on the other parts is still enough to get the bound in the lemma, by using Corollary 4.10 part (i). Details of the proof can be found on page 67.

In fact, the lower bound also holds, more specifically given a complete $\mathbf{G}$-configuration of sets $\mathcal{S}$ and $p \in \Delta\mathcal{S}$ there is a constant $C > 0$ such that for any $q \in \Delta\mathcal{S}$ with $|p - q|_1 \ll 1$ holds

$$C \cdot \mathcal{E}nt(\Lambda_\alpha) \leq \mathbf{k}(\mathcal{X}, \mathcal{Y})$$

where $\mathcal{X} = (\mathcal{S}, p)$, $\mathcal{Y} = (\mathcal{S}, q)$ and $\alpha = \frac{1}{2}|p - q|_1$. We will not use this fact and therefore do not include a proof.

Once the $\mathbf{G}$-configuration of sets $\mathcal{S}$ is fixed, there is a map from $\Delta\mathcal{S}$ to $\mathbf{Prob}\langle\mathbf{G}\rangle$. As can be seen from the discussion above, even though the map is continuous it is not Lipschitz.

5. Distance between types

As explained in Section 3.3, given a complete $\mathbf{G}$-configuration $\mathcal{S}$ of sets and a rational distribution $\pi \in \Delta\mathcal{S}$ we construct a homogeneous configuration $T_\pi^{(n)}\mathcal{S}$, which is called the type of $\mathcal{S}$ over π. Our goal in this section is to estimate the Kolmogorov distance between two types over two different distributions $\pi_1, \pi_2 \in \Delta^{(n)}\mathcal{S}$ in terms of the total variation distance $|\pi_1 - \pi_2|_1$.

For this purpose we use a "lagging" technique which is explained below.

5.1. The lagging trick.

Let Λ_α be a binary probability space,

$$\Lambda_\alpha = \big(\{\square, \blacksquare\}; p_{\Lambda_\alpha}(\blacksquare) = \alpha\big)$$

and let $\mathcal{X} = \{(\underline{X}_i, p_i); f_{ij}\}$, $\mathcal{Z} = \{(\underline{Z}_i, q_i); g_{ij}\}$ be two configurations modeled on a complete diagram category $\mathbf{G}$ and included in a minimal two-fan

$$\Lambda_\alpha \xleftarrow{\lambda} \mathcal{Z} \xrightarrow{\rho} \mathcal{X}$$

Recall that the left terminal vertex in this two-fan should be interpreted as a constant $\mathbf{G}$-configuration $\Lambda_\alpha^{\mathbf{G}}$.

Assume further that the distribution q on $\mathcal{Z}$ is rational with denominator $n \in \mathbb{N}$, that is $q \in \Delta^{(n)}\underline{\mathcal{Z}}$. It follows that p and p_{Λ_α} are also rational with the same denominator n.

We construct a *lagging two-fan*

$$\mathcal{L} := \left(T^{((1-\alpha)n)}(\mathcal{X}\lfloor\Box) \xleftarrow{l} T^{(n)}\mathcal{Z} \xrightarrow{T\rho} T^{(n)}\mathcal{X}\right) \tag{5.1}$$

as follows. The right leg $T\rho$ of $\mathcal{L}$ is induced by the right leg ρ of the original two-fan. The left leg

$$l : T^{(n)}\mathcal{Z} \to T^{((1-\alpha)n)}(\mathcal{X}\lfloor\Box)$$

is obtained by erasing symbols that reduce to $\blacksquare$ and applying ρ to the remaining symbols. The target space for the reduction l is the true type of $\mathcal{X}\lfloor\Box$ which is "lagging" behind $T^{(n)}\mathcal{Z}$ by a factor of $(1-\alpha)$. More specifically, the reduction l is constructed as follows. Let $\lambda_j : Z_j \to \Lambda_\alpha$ be the components of the reduction $\lambda : \mathcal{Z} \to \Lambda_\alpha$.

Given $\overline{z} = (z_i)_{i=1}^n \in T^{(n)}Z_j$ define the subset of indexes

$$I_{\overline{z}} := \{i \parallel \lambda_j(z_i) = \Box\}$$

and define the j^{th} component of l by

$$l_j\big((z_i)_{i=1}^n\big) := (\rho(z_i))_{i\in I_{\overline{z}}}$$

By equivariance each l_j is a reduction of homogeneous spaces, since the inverse image of any point has the same cardinality. Moreover the reductions l_j commute with the reductions in $T^{(n)}\mathcal{Z}$ as explained in Section 3.3 and therefore l is a reduction of configurations.

The next lemma uses the lagging two-fan to estimate the Kolmogorov distance between its terminal configurations.

Lemma 5.1. *Let $\mathcal{X}, \mathcal{Z} \in$ **Prob**$\langle\mathbf{G}\rangle$ be two configurations modeled on a complete diagram category $\mathbf{G}$ and included in a minimal two-fan*

$$\Lambda_\alpha \xleftarrow{\lambda} \mathcal{Z} \xrightarrow{\rho} \mathcal{X}$$

where distribution on $\mathcal{Z}$ is rational with denominator $n \in \mathbb{N}$. Then

$$\begin{aligned}\mathbf{k}&\left(T^{((1-\alpha)n)}(\mathcal{X}\lfloor\Box)\,,\,T^{(n)}\mathcal{X}\right)\\ &\le n\cdot[\![G]\!]\cdot[2\,\mathcal{E}nt(\Lambda_\alpha)+\alpha\cdot\ln|X_0|]+2\cdot[\![\mathbf{G}]\!]\cdot|X_0|\cdot\ln(n+1)\\ &= n\cdot[\![G]\!]\cdot\big[2\,\mathcal{E}nt(\Lambda_\alpha)+\alpha\cdot\ln|X_0|\big]+\mathcal{O}\left(|X_0|\cdot\ln n\right)\end{aligned}$$

⊠

The Lemma (and Proposition 5.2 below) are closely related to the local estimate, Proposition 4.11. It is an immediate consequence of the Slicing Lemma, in particular Corollary 4.10 part (ii) that

$$\mathbf{k}\left(\mathcal{X}\lfloor\Box\,,\,\mathcal{X}\right)\le[\![G]\!]\cdot[2\,\mathcal{E}nt(\Lambda_\alpha)+\alpha\cdot\ln|X_0|]$$

This is a tacit ingredient in the proof of the local estimate. By the subadditivity of the Kolmogorov distance,

$$\mathbf{k}\left((\mathcal{X}\lfloor\Box)^{\otimes n}\,,\,\mathcal{X}^{\otimes n}\right)\le n\cdot[\![G]\!]\cdot[2\,\mathcal{E}nt(\Lambda_\alpha)+\alpha\cdot\ln|X_0|]$$

This bound is almost the estimate in Lemma 5.1, except Lemma 5.1 estimates the distance between types rather than tensor powers. We will soon see that tensor powers and types are very close in the Kolmogorov distance. However, for the purpose of the proof of Lemma 5.1, it suffices to know that their entropies are close, an estimate that is provided by Corollary 3.3.

Proof (of Lemma 5.1): We will use the lagging two-fan constructed in Equation (5.1), namely

$$\mathcal{L} := \left(T^{((1-\alpha)n)}(\mathcal{X}\lfloor\square) \xleftarrow{l} T^{(n)}\mathcal{Z} \xrightarrow{T\rho} T^{(n)}\mathcal{X}\right)$$

as a coupling to estimate the Kolmogorov distance

$$\mathbf{k}\left(T^{((1-\alpha)n)}(\mathcal{X}\lfloor\square)\,,\,T^{(n)}\mathcal{X}\right) \leq \mathrm{kd}(\mathcal{L})$$

Recall that by Corollary 3.3 for a probability space X with a rational distribution we have

$$n\cdot\mathcal{E}nt(X) - |X|\cdot\ln(n+1) \leq \mathcal{E}nt(T^{(n)}X) \leq n\cdot\mathcal{E}nt(X)$$

Thus we can estimate $\mathrm{kd}(\mathcal{L})$ as follows

$$\begin{aligned}
\mathrm{kd}(\mathcal{L}) = \sum_i \Big[&\big(\mathcal{E}nt(T^{(n)}Z_i) - \mathcal{E}nt(T^{(n)}X_i)\big)\\
&+ \big(\mathcal{E}nt(T^{(n)}Z_i) - \mathcal{E}nt(T^{((1-\alpha)n)}(X_i\lfloor\square))\big)\Big]\\
\leq n\cdot\sum_i &\Big[\big(\mathcal{E}nt(Z_i) - \mathcal{E}nt(X_i)\big) + \big(\mathcal{E}nt(Z_i) - (1-\alpha)\,\mathcal{E}nt(X_i\lfloor\square)\big)\Big]\\
&+ 2\cdot[\![\mathbf{G}]\!]\cdot|X_0|\cdot\ln(n+1)
\end{aligned}$$

By minimality of the original two-fan and Shannon inequality (1.2) we have a bound

$$\mathcal{E}nt(Z_i) - \mathcal{E}nt(X_i) \leq \mathcal{E}nt(\Lambda_\alpha)$$

The second part in the sum can be estimated using relation (1.4) as follows

$$\begin{aligned}
\mathcal{E}nt(Z_i) - (1-\alpha)\,\mathcal{E}nt(X_i\lfloor\square) &= \mathcal{E}nt(\Lambda_\alpha) + \mathcal{E}nt(X_i\lfloor\Lambda_\alpha) - (1-\alpha)\,\mathcal{E}nt(X_i\lfloor\square)\\
&= \mathcal{E}nt(\Lambda_\alpha) + (1-\alpha)\,\mathcal{E}nt(X_i\lfloor\square) + \alpha\,\mathcal{E}nt(X_i\lfloor\blacksquare) -\\
&\quad - (1-\alpha)\,\mathcal{E}nt(X_i\lfloor\square)\\
&\leq \mathcal{E}nt(\Lambda_\alpha) + \alpha\cdot\ln|X_i|
\end{aligned}$$

Combining all of the above we obtain the estimate in the conclusion of the lemma. ⊠

5.2. **Distance between types.** In this section we use the lagging trick as described above to estimate the distance between types over two different distributions in $\Delta\mathcal{S}$ where $\mathcal{S}$ is a complete configuration of sets.

Proposition 5.2. *Suppose $\mathcal{S}$ is a complete* **G***-configuration of sets with initial set S_0. Suppose $p, q \in \Delta^{(n)}\mathcal{S}$ and let $\alpha = \frac{1}{2}|p_0 - q_0|_1$. Then*

$$\begin{aligned}\mathbf{k}(T_p^{(n)}\mathcal{S}, T_q^{(n)}\mathcal{S}) &\leq 2n \cdot [\![\mathbf{G}]\!] \cdot [\alpha \cdot \ln|S_0| + 2\,\mathcal{E}nt(\Lambda_\alpha)] + 4[\![\mathbf{G}]\!] \cdot |X_0| \cdot \ln(n+1) \\ &= 2n \cdot [\![\mathbf{G}]\!] \cdot [\alpha \cdot \ln|S_0| + 2\,\mathcal{E}nt(\Lambda_\alpha)] + \mathcal{O}(|X_0| \cdot \ln n)\end{aligned}$$

⊠

As in the local estimate, the idea of the proof is to write p and q as a convex combination of a common distribution $\hat{p}$ and "small amounts" of p^+ and q^+, respectively. Then we use the lagging trick to estimate distances between types over p and $\hat{p}$, as well as between types over q and $\hat{p}$. We now present details of the proof.

Proof (of Proposition 5.2): Recall that for a complete configuration $\mathcal{S}$ with initial set S_0 we have

$$\Delta\mathcal{S} \cong \Delta S_0 \tag{5.2}$$

Our goal now is to write p and q as the convex combination of three other distributions $\hat{p}$, p^+ and q^+ as in

$$\begin{aligned}p &= (1-\alpha)\cdot\hat{p} + \alpha\cdot p^+ \\ q &= (1-\alpha)\cdot\hat{p} + \alpha\cdot q^+\end{aligned}$$

We could do it the following way. Let $\alpha := \frac{1}{2}|p_0 - q_0|_1$. If $\alpha = 1$ then the proposition follows trivially by constructing a tensor-product fan, so from now on we assume that $\alpha < 1$. Define three probability distributions $\hat{p}_0$, p_0^+ and q_0^+ on S_0 by setting for every $x \in S_0$

$$\begin{aligned}\hat{p}_0(x) &:= \frac{1}{1-\alpha}\min\{p_0(x), q_0(x)\} \\ p_0^+ &:= \frac{1}{\alpha}\big(p_0 - (1-\alpha)\hat{p}_0\big) \\ q_0^+ &:= \frac{1}{\alpha}\big(q_0 - (1-\alpha)\hat{p}_0\big)\end{aligned}$$

Denote by $\hat{p}, p^+, q^+ \in \Delta\mathcal{S}$ the distributions corresponding to $\hat{p}_0, p_0^+, q_0^+ \in \Delta S_0$ under the affine isomorphism (5.2). Thus we have

$$\begin{aligned}p &= (1-\alpha)\hat{p} + \alpha\cdot p^+ \\ q &= (1-\alpha)\hat{p} + \alpha\cdot q^+\end{aligned}$$

Now we construct a pair of two-fans of **G**-configurations

$$\begin{aligned}\Lambda_\alpha \leftarrow \tilde{\mathcal{X}} \rightarrow \mathcal{X} \\ \Lambda_\alpha \leftarrow \tilde{\mathcal{Y}} \rightarrow \mathcal{Y}\end{aligned} \tag{5.3}$$

by setting

$$\begin{aligned}
\mathcal{X} &:= (\mathcal{S}, p)\\
\mathcal{Y} &:= (\mathcal{S}, q)\\
\tilde{X}_i &:= \Big(S_i \times \underline{\Lambda}_\alpha;\ \tilde{p}_i(s,\square) = (1-\alpha)\hat{p}_i(s),\ \tilde{p}_i(s,\blacksquare) = \alpha \cdot p_i^+(s)\Big)\\
\tilde{Y}_i &:= \Big(S_i \times \underline{\Lambda}_\alpha;\ \tilde{q}_i(s,\square) = (1-\alpha)\hat{p}_i(s),\ \tilde{q}_i(s,\blacksquare) = \alpha \cdot q_i^+(s)\Big)
\end{aligned}$$

and

$$\begin{aligned}
\tilde{\mathcal{X}} &:= \big\{\tilde{X}_i;\ f_{ij} \times \mathrm{Id}\big\}\\
\tilde{\mathcal{Y}} &:= \big\{\tilde{Y}_i;\ f_{ij} \times \mathrm{Id}\big\}
\end{aligned}$$

The reductions in (5.3) are given by coordinate projections. We have the following isomorphisms

$$\mathcal{X}\lfloor\square \cong \mathcal{Y}\lfloor\square \cong (\mathcal{S}, \hat{p})$$

To estimate the distance between types we now apply Lemma 5.1 to the fans in (5.3)

$$\begin{aligned}
\mathbf{k}(T_p^{(n)}\mathcal{S}, T_q^{(n)}\mathcal{S}) &= \mathbf{k}(T^{(n)}\mathcal{X}, T^{(n)}\mathcal{Y})\\
&\leq \mathbf{k}\left(T^{(n)}\mathcal{X}, T^{((1-\alpha)n)}(\mathcal{X}\lfloor\square)\right) + \mathbf{k}\left(T^{((1-\alpha)n)}(\mathcal{Y}\lfloor\square), T^{(n)}\mathcal{Y}\right)\\
&\leq 2n \cdot [\![\mathbf{G}]\!] \cdot [\alpha \cdot \ln|S_0| + 2\,\mathcal{E}nt(\Lambda_\alpha)] + 4[\![\mathbf{G}]\!] \cdot |X_0| \cdot \ln(n+1)
\end{aligned}$$

⊠

The reason for the similarity between the local estimate and the distance estimate between types will become clear in the next section, when we establish the asymptotic equivalence between the Bernoulli sequence of probability spaces and sequence of types over rational distributions approximating the true distribution.

6. Asymptotic equipartition property for configurations

Below we prove that any Bernoulli sequence can be approximated by a sequence of homogeneous configurations. This is essentially the *Asymptotic Equipartition Theorem for configurations.*

Theorem 6.1. *Suppose $\mathcal{X} \in \mathbf{Prob}\langle\mathbf{G}\rangle$ is a complete configuration of probab ility spaces. Then there exists a sequence $\overline{\mathcal{H}} = (\mathcal{H}_n)_{n=0}^{\infty}$ of homogeneous configurations of the same c ombinatorial type as $\mathcal{X}$ such that*

$$\frac{1}{n}\mathbf{k}(\mathcal{X}^{\otimes n}, \mathcal{H}_n) = \mathcal{O}\left(\sqrt{\frac{\ln^3 n}{n}}\right)$$

More precisely, the sequence $\overline{\mathcal{H}}$ may be chosen such that for all $n \geq |X_0|$

$$\frac{1}{n}\mathbf{k}(\mathcal{X}^{\otimes n}, \mathcal{H}_n) \leq C(|X_0|, [\![\mathbf{G}]\!]) \cdot \sqrt{\frac{\ln^3 n}{n}} \tag{6.1}$$

where $C(|X_0|, [\![\mathbf{G}]\!])$ is a constant only depending on $|X_0|$ and $[\![\mathbf{G}]\!]$. ⊠

Proof: Denote by $\mathcal{S} = \underline{\mathcal{X}}$ the underlying configuration of sets and by $p_{\mathcal{X}}$ the true distribution on $\mathcal{S}$, such that

$$\mathcal{X} = (\mathcal{S}, p_X)$$

We will construct the approximating homogeneous sequence by taking types over rational approximations of $p_{\mathcal{X}}$ in $\Delta\mathcal{S}$, that converge sufficiently fast to the true distribution $p_{\mathcal{X}}$.

More specifically, we select rational distributions $p_n \in \Delta^{(n)}\mathcal{S}$ such that

$$|p_n - p_{\mathcal{X}}|_1 \leq \frac{|S_0|}{n}$$

As homogeneous spaces $\mathcal{H}_n$ we set $\mathcal{H}_n = T^{(n)}_{p_n}\mathcal{S}$. We will show that the Kolmogorov distance between $\mathcal{H}_n$ and $\mathcal{X}^{\otimes n}$ satisfies the required estimate (6.1).

First we apply slicing along the empirical two-fan

$$\mathcal{R}_n(\mathcal{X}) = \left(\mathcal{X}^{\otimes n} \leftarrow \tilde{\mathcal{X}}^{(n)} \to (\Delta\mathcal{S}, \tau_n)^{\mathbf{G}}\right)$$

defined in Section 3.3, Equation (3.7) on page 28.

For the estimate below we use the fact that

$$\mathcal{E}nt(\Delta\mathcal{S}, \tau_n) \leq \ln|\Delta^{(n)}\mathcal{S}| \leq |S_0| \cdot \ln(n+1)$$

By slicing (see Corollary 4.10(ii)) along the empirical two-fan we have

$$\begin{aligned}\mathbf{k}(T^{(n)}_{p_n}\mathcal{S}, \mathcal{X}^{\otimes n}) &\leq 2\cdot[\![\mathbf{G}]\!]\cdot\mathcal{E}nt(\Delta\mathcal{S}, \tau_n) + \int_{\Delta\mathcal{S}} \mathbf{k}(T^{(n)}_{p_n}\mathcal{S}, T^{(n)}_{\pi}\mathcal{S})\,\mathrm{d}\,\tau_n(\pi)\\ &\leq 2\cdot[\![\mathbf{G}]\!]\cdot|S_0|\cdot\ln(n+1) + \int_{\Delta\mathcal{S}} \mathbf{k}(T^{(n)}_{p_n}\mathcal{S}, T^{(n)}_{\pi}\mathcal{S})\,\mathrm{d}\,\tau_n(\pi)\end{aligned}$$

To estimate the integral we split the domain into a small divergence ball $B_{\varepsilon_n} = B_{\varepsilon_n}(p_{\mathcal{X}})$ around the "true" distribution and its complement

$$\begin{aligned}\int_{\Delta\mathcal{S}} \mathbf{k}(T^{(n)}_{p_n}\mathcal{S}, T^{(n)}_{\pi}\mathcal{S})\,\mathrm{d}\,\tau_n(\pi) &= \int_{\Delta\mathcal{S}\setminus B_{\varepsilon_n}} \mathbf{k}(T^{(n)}_{p_n}\mathcal{S}, T^{(n)}_{\pi}\mathcal{S})\,\mathrm{d}\,\tau_n(\pi)\\ &\quad + \int_{B_{\varepsilon_n}} \mathbf{k}(T^{(n)}_{p_n}\mathcal{S}, T^{(n)}_{\pi}\mathcal{S})\,\mathrm{d}\,\tau_n(\pi)\end{aligned} \tag{6.2}$$

and we set the radius ε_n equal to

$$\varepsilon_n := (|S_0| + 1)\frac{\ln(n+1)}{n}$$

To estimate the first integral on the right-hand side of equality (6.2) note that the distance between two types over the same configuration of sets can always be crudely estimated by

$$2\cdot\ln|S_0|\cdot[\![\mathbf{G}]\!]\cdot n$$

Moreover, by Sanov's theorem, Theorem 3.1, we can estimate the empirical measure of the complement of the divergence ball

$$\tau_n(\Delta\mathcal{S} \setminus B_{\varepsilon_n}) \leq \mathbf{e}^{-n\cdot\varepsilon_n + |S_0|\cdot\ln(n+1)} \leq \frac{1}{n}$$

where we used the definition of ε_n to conclude the last inequality. Therefore we obtain

$$\begin{aligned}\int_{\Delta\mathcal{S}\setminus B_{\varepsilon_n}} \mathbf{k}(T^{(n)}_{p_n}\mathcal{S}, T^{(n)}_{\pi}\mathcal{S})\,\mathrm{d}\,\tau_n(\pi) &\leq 2\cdot\ln|S_0|\cdot[\![\mathbf{G}]\!]\cdot n\cdot\tau_n(\Delta\mathcal{S}\setminus B_{\varepsilon_n})\\ &\leq 2\cdot\ln|S_0|\cdot[\![\mathbf{G}]\!]\end{aligned}$$

Define

$$\alpha_n = \frac{|S_0|}{n} + \sqrt{2\varepsilon_n}$$

if the right-hand side is smaller than 1 and set $\alpha_n = 1$ otherwise. Then every $\pi \in B_{\varepsilon_n}(p_{\mathcal{X}})$ satisfies $|p_n - \pi| \leq 2\alpha_n$ by Pinsker's inequality (Lemma 3.1, (i)), and the triangle inequality. Consequently, by the estimate on the distance between types in Proposition 5.2

$$\begin{aligned}&\int_{B_{\varepsilon_n}} \mathbf{k}(T^{(n)}_{p_n}\mathcal{S}, T^{(n)}_{\pi}\mathcal{S})\,\mathrm{d}\,\tau_n(\pi)\\ &\qquad\leq 2n\cdot[\![\mathbf{G}]\!]\cdot(\alpha_n\ln|S_0| + 2\,\mathcal{E}nt(\Lambda_{\alpha_n})) + 4\cdot[\![\mathbf{G}]\!]\cdot|S_0|\cdot\ln(n+1)\end{aligned}$$

Using the definition of α_n and ε_n we find that

$$\int_{B_{\varepsilon_n}} \mathbf{k}(T^{(n)}_{p_n}\mathcal{S}, T^{(n)}_{\pi}\mathcal{S})\,\mathrm{d}\,\tau_n(\pi) = \mathcal{O}\left(\sqrt{n\cdot\ln^3 n}\right)$$

and hence combining the above estimates

$$\frac{1}{n}\mathbf{k}(T^{(n)}_{p_n}\mathcal{S}, \mathcal{X}^{\otimes n}) = \mathcal{O}\left(\sqrt{\frac{\ln^3 n}{n}}\right).$$

A more precise check shows that for $n \geq |S_0|$, the constants appearing in $\mathcal{O}$ only depend on $|S_0|$ and $[\![\mathbf{G}]\!]$. ⊠

It is worth noting that each type considered as a subspace of the tensor power takes up only small probability. In fact its probability converges to zero with growing n. But as the calculation above shows, most (in terms of probability) of the configuration $\mathcal{X}^{\otimes n}$ consists of *polynomially many* types that are $\mathbf{k}$-similar to each other. Relative to the exponential growth of sizes of all the parts, "polynomially many" is as good as one. This is the difference with the setup used in Gromov's [Gro12]

7. EXTENSIONS

In the introduction we have already emphasized the close relationship between relative entropic sets and Information-Optimization problems. There, our definitions were restricted to extensions of two-fans to full configurations corresponding to three random variables. We will now generalize these definitions, and make the relationship between relative entropic sets and Information-Optimization problems explicit.

Further we will prove the Extension Lemma and use it to show that the relative entropic set associated to a full configuration depends continuously on the configuration.

7.1. **Information-Optimization and the relative entropic set.** In Section 2.4.1 we introduced (for $k \le l$) the restriction operator

$$R^*_{k,l} : \mathbf{Prob}\langle\mathbf{\Lambda}_l\rangle \to \mathbf{Prob}\langle\mathbf{\Lambda}_k\rangle$$

as follows. For a minimal full configuration $\mathcal{Y} = \langle Y_i\rangle_{i=1}^l$ we denote by

$$R^*_{k,l}\mathcal{Y} = \langle Y_i\rangle_{i\in\{1,\dots,k\}}$$

the restriction of $\mathcal{Y}$ to a minimal full configuration generated by Y_i, $i \in \{1,\dots,k\}$.

We call a minimal configuration $\mathcal{Y} \in \mathbf{Prob}\langle\Lambda_k\rangle$ an l-extension of a configuration $\mathcal{X}$ if

$$R^*_{k,l}\mathcal{Y} = \mathcal{X}$$

and we denote the class of all l-extensions by $\mathrm{Ext}_l(\mathcal{X})$.

Recall that for a full configuration $\mathcal{Y} \in \mathbf{Prob}\langle\Lambda_l\rangle$, we record the entropies of all its probability spaces in a vector in $\mathbb{R}^{2^{\{1,\dots,l\}}\smallsetminus\{\varnothing\}}$ that we denote by

$$\mathcal{E}nt_*(\mathcal{Y}) := \big(\mathcal{E}nt(Y_I)\big)_{I\in 2^{\{1,\dots,l\}}\smallsetminus\{\varnothing\}}$$

The entries in this vector are all nonnegative. To simplify notations we set

$$\mathbb{E}_l := (\mathbb{R}^{2^{\{1,\dots,l\}}\smallsetminus\{\varnothing\}}, |\cdot|_1)$$

and denote by $\mathbb{E}^*_l$ its dual vector-space.

As in the introduction, we introduce the *unstabilized relative entropic set*

$$\mathbf{\Gamma}^\circ_l(\mathcal{X}) := \{\mathcal{E}nt_*(\mathcal{Y}) \,\|\, \mathcal{Y} \in \mathrm{Ext}_l(\mathcal{X})\}$$

By the additivity property of the entropy with respect to tensor powers, there is the inclusion

$$\mathbf{\Gamma}^\circ_l(\mathcal{X}^{\otimes m}) + \mathbf{\Gamma}^\circ_l(\mathcal{X}^{\otimes n}) \subset \mathbf{\Gamma}^\circ_l(\mathcal{X}^{\otimes(m+n)}) \tag{7.1}$$

where the sum on the left hand side is the Minkowski sum. This allows us to define the limit

$$\lim_{n\to\infty}\frac{1}{n}\mathbf{\Gamma}^\circ_l(\mathcal{X}^{\otimes n}) := \bigcup_{n\in\mathbb{N}}\frac{1}{n}\mathbf{\Gamma}^\circ_l(\mathcal{X}^{\otimes n})$$

and we define the *stabilized relative entropic set* by

$$\overline{\mathbf{\Gamma}}_l(\mathcal{X}) := \mathrm{Closure}\left(\lim_{n\to\infty}\frac{1}{n}\mathbf{\Gamma}^\circ_l(\mathcal{X}^{\otimes n})\right)$$

which is a closed convex subset of $\mathbb{E}_l$ by property (7.1).

For a vector $c \in \mathbb{E}_l^*$, we define the *Information-Optimization problem*

$$\mathrm{IO}_c(\mathcal{X}) := \inf_{\mathcal{Y}\in\mathrm{Ext}_l(\mathcal{X})} \big\langle c\,,\, \mathcal{E}nt_*(\mathcal{Y})\,\big\rangle = \inf_{\mathcal{Y}\in\mathrm{Ext}_l(\mathcal{X})} \sum_{I\subset\{1,\dots,l\}} c_I \cdot \mathcal{E}nt(Y_I)$$

where c_I's are the coordinates of the vector c with respect to the basis in $\mathbb{E}_l^*$ dual to the standard basis in $\mathbb{E}_l$. Note that, equation (7.1) implies that the sequence

$$n \mapsto \mathrm{IO}_c\left(\mathcal{X}^{\otimes n}\right)$$

is subadditive. Hence, the limit

$$\lim_{n\to\infty} \frac{1}{n} \mathrm{IO}_c(\mathcal{X}^{\otimes n})$$

always exists (but may be equal to $-\infty$). If for all $n \in \mathbb{N}$,

$$\frac{1}{n} \mathrm{IO}_c(\mathcal{X}^{\otimes n}) = \mathrm{IO}_c(\mathcal{X})$$

we call the optimization problem associated to c *stable*. In general, we define the stabilized optimization problem

$$\mathrm{IO}_c^s(\mathcal{X}) := \lim_{n\to\infty} \frac{1}{n} \mathrm{IO}_c(\mathcal{X}_n^{\otimes n}).$$

As the stabilized relative entropic set is convex, it is the intersection of half-spaces that are defined by linear inequalities on entropies

$$\overline{\mathbf{\Gamma}}_k(\mathcal{X}) = \bigcap_{c\in\mathbb{E}_l^*} \{x \in \mathbb{E}_l \,\|\, \langle c, x\rangle \geq \mathrm{IO}_c^s(\mathcal{X})\}$$

In other words, the stabilized information optimization problems, that occur so often in practice, identify supporting hyper-planes of the convex set. The solution of all such linear problems determine the shape of the relative entropic set and vice versa.

7.2. **The entropic set and the entropic cone.** The definitions of relative entropic sets are motivated by the more classical notion of the entropic cone, which we will briefly discuss now. For $l \in \mathbb{N}$, the entropic set is defined as

$$\mathbf{\Gamma}_l^\circ := \{\mathcal{E}nt_*(\mathcal{Y}) \,\|\, \mathcal{Y} \in \mathbf{Prob}\,\langle \mathbf{\Lambda}_l\rangle\,,\, \mathcal{Y} \text{ is minimal}\}$$

Its closure is usually referred to as the entropic cone

$$\overline{\mathbf{\Gamma}}_l := \mathrm{Closure}(\mathbf{\Gamma}_l^\circ)$$

Indeed, the entropic cone $\overline{\mathbf{\Gamma}}_l$ is a closed, convex cone in $\mathbb{R}^{2^l-1}$ [Yeu12]. For $l \leq 3$, the entropic cone $\overline{\mathbf{\Gamma}}_l$ is polyhedral and completely described by Shannon inequalities. However, for $l \geq 4$, the situation is much more complicated. It is known that $\overline{\mathbf{\Gamma}}_l$ is not polyhedral for $l \geq 4$ [Mat07]. The shape of the entropic cone is not known as of the time of writing this article. It is an important open problem in information theory to find tight bounds on the entropic cone for $l \geq 4$. We hope that the techniques developed in this article will eventually lend itself to finding a useful characterization.

In fact, the entropic cone can be considered as the relative entropic set of an empty configuration $\not{O} \in \mathbf{Prob}\langle\varnothing\rangle$, that corresponds to the empty diagram category $\varnothing = \mathbf{\Lambda}_0$

$$\overline{\mathbf{\Gamma}}_l = \overline{\mathbf{\Gamma}}_l(\not{O})$$

For a diagram category $\mathbf{G}$ let us denote by $\{\bullet\} = \{\bullet\}^{\mathbf{G}}$ the constant $\mathbf{G}$-configuration of one-point spaces. Given an l-extension $\mathcal{Y} \in \mathbf{Prob}\langle\mathbf{\Lambda}_l\rangle$ of $\{\bullet\}^{\mathbf{\Lambda}_k}$ the restriction to the last $l-k$ terminal spaces induces a linear isomorphism

$$\overline{\mathbf{\Gamma}}_l(\{\bullet\}^{\mathbf{\Lambda}_k}) \cong \overline{\mathbf{\Gamma}}_{l-k} \tag{7.2}$$

7.3. **Extension lemma.** The Lipschitz continuity of relative entropic sets will follow from the following important proposition, which we will refer to as the Extension Lemma.

Proposition 7.1. ↓ (Extension Lemma) *Let $k, l \in \mathbb{N}$, $k \le l$ and let $\mathcal{X}, \mathcal{X}' \in \mathbf{Prob}\langle\mathbf{\Lambda}_k\rangle$ be minimal full configurations. For every $\mathcal{Y} \in \operatorname{Ext}_l \mathcal{X}$ there exists a $\mathcal{Y}' \in \operatorname{Ext}_l \mathcal{X}'$ such that*

$$\mathbf{k}(\mathcal{Y}', \mathcal{Y}) \le 2^{l-k}\,\mathbf{k}(\mathcal{X}', \mathcal{X})$$

⊠

The key behind the proof of the Extension Lemma, is that there is a full configuration $\mathcal{Z}$ that extends both $\mathcal{Y}$ and the optimal coupling between $\mathcal{X}$ and $\mathcal{X}'$. The configuration $\mathcal{Y}'$ can be chosen to be the restriction of $\mathcal{Z}$ to the full configuration generated by $\mathcal{X}'$ and the terminal spaces in $\mathcal{Y}$ which are not in $\mathcal{X}$. The estimate directly follows from Shannon inequalities. We present details at page 69.

It follows immediately from the Extension Lemma and the Lipschitz property of the entropy function $\mathcal{E}nt_*$ that asymptotically equivalent configurations have the same solutions to all Information-Optimization problems and, consequently, they have the same stabilized relative entropic set.

In fact, we have a much stronger statement. Both the unstabilized and stabilized relative entropic sets have a Lipschitz dependence on the configuration, if the distance between sets is measured by the Hausdorff distance.

Let us endow the collection of subsets of $\mathbb{E}_l$ with the Hausdorff metric with respect to the ℓ_1-distance. For two subsets S_1, S_2 of $\mathbb{E}_l$, define the Hausdorff distance between them by

$$\mathrm{d}_H(S_1, S_2) = \inf\{\varepsilon > 0 \,\|\, S_1 \subset S_2 + B_\varepsilon \text{ and } S_2 \subset S_1 + B_\varepsilon\}$$

where B_ε is the ℓ_1-ball of size ε around the origin in $\mathbb{E}_l$.

In fact, at this point the Hausdorff distance is only an extended pseudo-metric, in the sense, that it may take infinite values and it may vanish on pairs of non-identical points.

Suppose now that we are given two minimal full configurations $\mathcal{X}, \mathcal{X}' \in \mathbf{Prob}\langle\mathbf{\Lambda}_k\rangle$, and suppose a point $y \in \mathbb{E}_l$ lies in the unstabilized relative entropic set of $\mathcal{X}$, that is

$$y \in \mathbf{\Gamma}^\circ_l(\mathcal{X})$$

This means that there is an extension $\mathcal{Y} \in \mathrm{Ext}_l(\mathcal{X})$ such that

$$\mathcal{E}nt_*(\mathcal{Y}) = y$$

By the Extension Lemma, there exists a configuration $\mathcal{Y}' \in \mathrm{Ext}_l(\mathcal{X}')$ such that

$$\mathbf{k}(\mathcal{Y}, \mathcal{Y}') \le 2^{l-k}\, \mathbf{k}(\mathcal{X}, \mathcal{X}')$$

and by the 1-Lipschitz property of the entropy function the point $y' := \mathcal{E}nt_*(\mathcal{Y}')$ is close to the point y, that is

$$|y - y'|_1 = |\,\mathcal{E}nt_*(\mathcal{Y}) - \mathcal{E}nt_*(\mathcal{Y}')|_1 \le 2^{l-k}\, \mathbf{k}(\mathcal{X}, \mathcal{X}')$$

We have thus obtained the following corollary to the Extension Lemma.

Corollary 7.2. *Let $k \in \mathbb{N}$ and $\mathcal{X}, \mathcal{X}' \in \mathbf{Prob}\langle \mathbf{\Lambda}_k \rangle$. Then the Hausdorff distance between their unstabilized relative entropic sets satisfies the following Lipschitz estimate*

$$\mathrm{d}_H\left(\mathbf{\Gamma}_l^\circ(\mathcal{X}), \mathbf{\Gamma}_l^\circ(\mathcal{X}')\right) \le 2^{l-k}\, \mathbf{k}(\mathcal{X}, \mathcal{X}')$$

⊠

Note that in particular, the distance between unstabilized relative entropic sets is always finite and

$$\mathrm{d}_H\left(\mathbf{\Gamma}_l^\circ(\mathcal{X}), \mathbf{\Gamma}_l^\circ(\{\bullet\}^{\mathbf{\Lambda}_k})\right) \le 2^{l-k}\, \mathbf{k}(\mathcal{X}, \{\bullet\}^{\mathbf{\Lambda}_k}) = 2^{l-k} |\,\mathcal{E}nt_*(\mathcal{X})|_1.$$

Let us denote by $\mathbf{K}_{k,l}$ the metric space of closed convex sets K in $\mathbb{E}_l$ such that

$$d_H(K, \overline{\mathbf{\Gamma}}_l(\{\bullet\}^{\mathbf{\Lambda}_k})) < \infty$$

endowed with the Hausdorff distance.

Theorem 7.3. ↓ *Let $k \in \mathbb{N}$ and $\mathcal{X}, \mathcal{X}' \in \mathbf{Prob}\langle \mathbf{\Lambda}_k \rangle$. Then for all $l \in \mathbb{N}$, the Hausdorff distance between their stabilized relative entropic sets satisfies the Lipschitz estimate*

$$\mathrm{d}_H\left(\overline{\mathbf{\Gamma}}_l(\mathcal{X}), \overline{\mathbf{\Gamma}}_l(\mathcal{X}')\right) \le 2^{l-k}\, \boldsymbol{\kappa}(\mathcal{X}, \mathcal{X}')$$

In other words, the map $\overline{\mathbf{\Gamma}}_l$ from minimal full configurations in $\mathbf{Prob}\langle \mathbf{\Lambda}_k \rangle$ to $\mathbf{K}_{k,l}$ is 2^{l-k}-Lipschitz. ⊠

Finally, as a primer to Section 9, note that for any set $K \in \mathbf{K}_{k,l}$ the sequence

$$n \mapsto \frac{1}{n} K$$

converges in the Hausdorff distance to $\overline{\mathbf{\Gamma}}_l(\{\bullet\}^{\mathbf{\Lambda}_k})$. The set $K \subset \mathbb{E}_l$ can be viewed as a metric space itself, by just restricting the ℓ^1-metric to it. The above convergence can then be expressed by saying that the asymptotic cone of K equals $\overline{\mathbf{\Gamma}}_l(\{\bullet\}^{\mathbf{\Lambda}_k})$ and is isomorphic to $\overline{\mathbf{\Gamma}}_{l-k}$.

8. Mixtures

Mixtures provide some technical tools, which we will use in Section 9. The input data for the mixture operation is a family of **G**-configurations, parametrized by a probability space. As result one obtains another **G**-configuration with the pre-specified conditionals. One particular instance of a mixture is when one mixes two configurations $\mathcal{X}$ and $\{\bullet\}^{\mathbf{G}}$, the latter being a constant **G**-configuration of one-point probability spaces. This operation will be used as a substitute for taking radicals "$\mathcal{X}^{\otimes(1/n)}$" in Section 9 below.

8.1. Definition and elementary properties.

Let **G** be a complete diagram category and Θ be a probability space. Let $\{\mathcal{X}_\theta\}_{\theta\in\underline{\Theta}}$ be a family of **G**-configurations parametrized by Θ. The *mixture* of the family $\{\mathcal{X}_\theta\}$ is the reduction

$$\mathcal{M}ix\,\{\mathcal{X}_\theta\} = \big(\mathcal{Y} \longrightarrow \Theta^{\mathbf{G}}\big)$$

such that

$$\mathcal{Y}\lfloor\theta \cong \mathcal{X}_\theta \tag{8.1}$$

The mixture exists and is uniquely defined by property (8.1) up to an isomorphism which is identity on $\Theta^{\mathbf{G}}$.

We denote the top configuration of the mixture

$$\mathcal{Y} = \bigoplus_{\theta\in\Theta} \mathcal{X}_\theta$$

and also call it the mixture of the family $\{\mathcal{X}_\theta\}$.

When

$$\Theta = \Lambda_\alpha = \big(\,\{\square, \blacksquare\}\,; p(\blacksquare) = \alpha\big)$$

is a binary space we write simply

$$\mathcal{X}_\blacksquare \oplus_{\Lambda_\alpha} \mathcal{X}_\square$$

for the mixture. The configuration subindexed by the $\blacksquare$ will always be the first summand.

The entropy of the mixture can be evaluated by the following formula

$$\mathcal{E}nt_*\left(\bigoplus_{\theta\in\Theta}\mathcal{X}_\theta\right) = \int_\Theta \mathcal{E}nt_*(\mathcal{X}_\theta)\,\mathrm{d}\,p(\theta) + \mathcal{E}nt_*(\Theta^{\mathbf{G}})$$

Mixtures satisfy the distributive law with respect to the tensor product

$$\mathcal{M}ix(\{\mathcal{X}_\theta\}_{\theta\in\Theta}) \otimes \mathcal{M}ix(\{\mathcal{Y}_{\theta'}\}_{\theta'\in\Theta'}) \cong \mathcal{M}ix(\{\mathcal{X}_\theta\otimes\mathcal{Y}_{\theta'}\}_{(\theta,\theta')\in\Theta\otimes\Theta'})$$

$$\left(\bigoplus_{\theta\in\Theta}\mathcal{X}_\theta\right) \otimes \left(\bigoplus_{\theta'\in\Theta'}\mathcal{Y}_{\theta'}\right) \cong \bigoplus_{(\theta,\theta')\in\Theta\otimes\Theta'} (\mathcal{X}_\theta\otimes\mathcal{Y}_{\theta'})$$

8.2. **The distance estimates.** Recall that for a diagram category $\mathbf{G}$ we denote by $\{\bullet\} = \{\bullet\}^{\mathbf{G}}$ the constant $\mathbf{G}$-configuration of one-point spaces.

The mixture of a $\mathbf{G}$-configuration with $\{\bullet\}^{\mathbf{G}}$ may serve as an ersatz of taking radicals of the configuration. The following lemma provides a justification of this by some distance estimates related to mixtures and will be used in Section 9.

Lemma 8.1. ↓ *Let* $\mathbf{G}$ *be a complete diagram category and* $\mathcal{X}, \mathcal{Y} \in \mathbf{Prob}\langle\mathbf{G}\rangle$. *Then*

(i) $\boldsymbol{\kappa}(\mathcal{X}, \mathcal{X}^{\otimes n} \oplus_{\Lambda_{1/n}} \{\bullet\}) \leq \mathcal{E}nt(\Lambda_{1/n})$

(ii) $\boldsymbol{\kappa}\big(\mathcal{X}, (\mathcal{X} \oplus_{\Lambda_{1/n}} \{\bullet\})^{\otimes n}\big) \leq n \cdot \mathcal{E}nt(\Lambda_{1/n})$

(iii) $\boldsymbol{\kappa}\big((\mathcal{X} \otimes \mathcal{Y}) \oplus_{\Lambda_{1/n}} \{\bullet\}, (\mathcal{X} \oplus_{\Lambda_{1/n}} \{\bullet\}) \otimes (\mathcal{Y} \oplus_{\Lambda_{1/n}} \{\bullet\})\big) \leq 3\,\mathcal{E}nt(\Lambda_{1/n})$

(iv) $\boldsymbol{\kappa}\big((\mathcal{X} \oplus_{\Lambda_{1/n}} \{\bullet\}), (\mathcal{Y} \oplus_{\Lambda_{1/n}} \{\bullet\})\big) \leq \dfrac{1}{n}\boldsymbol{\kappa}(\mathcal{X}, \mathcal{Y})$

⊠

The proof can be found on page 71. Note, that the distance estimates in the lemma above are with respect to the asymptotic Kolmogorov distance. This is essential, since from the perspective of the intrinsic Kolmogorov distance mixtures are very badly behaved.

9. Tropical probability spaces and their configurations

In this section we introduce the notion of *tropical probability spaces* and their *configurations*. Configurations of tropical probability spaces are points in the asymptotic cone of the space $\mathbf{Prob}\langle\mathbf{G}\rangle$, that is they are "limits" of certain divergent sequences of "normal" configurations. We will first give the construction of an asymptotic cone in an abstract context. Next, we will apply the construction to the particular case of configurations of probability spaces. For some background on asymptotic cones, see for instance [BBI01].

9.1. **Asymptotic cones of metric spaces.** The *asymptotic cone* captures large-scale geometry of a metric space. Abstractly, the asymptotic cone of a pointed metric space is the pointed Gromov-Hausdorff limit of the sequence of spaces obtained from the given one by scaling down the metric. Of course, convergence is in general by no means assured. Sometimes a weaker type of convergence (using ultrafilters) is considered. Since, in our case, the asymptotic cone can be evaluated relatively explicitly we do not give the definition of Gromov-Hausdorff convergence or convergence with respect to an ultrafilter here, but instead give a construction.

We would like to understand asymptotic cones of the space of configurations of probability spaces, considered as a metric space with the pseudo-metric $\mathbf{k}$ or $\boldsymbol{\kappa}$. For a fixed complete diagram category $\mathbf{G}$ the space $\mathbf{Prob}\langle\mathbf{G}\rangle$ is a monoid with operation $\otimes$. It has the additional property that shifts are non-expanding maps. This simplifies the construction and analysis of its asymptotic cone. In fact, as we will see later the metric $\boldsymbol{\kappa}$ is already *asymptotic* relative to $\mathbf{k}$. The application of the asymptotic cone construction to the metric $\boldsymbol{\kappa}$ allows us to obtain a complete metric space with a simple description of points in it.

Note that even though the monoid $(\mathbf{Prob}\langle\mathbf{G}\rangle, \otimes)$ is not Abelian it has the property that for any $\mathcal{X}_1, \mathcal{X}_2 \in \mathbf{Prob}\langle\mathbf{G}\rangle$ one has

$$\mathbf{k}(\mathcal{X}_1 \otimes \mathcal{X}_2, \mathcal{X}_2 \otimes \mathcal{X}_1) = 0$$

Thus, from a metric perspective it is as good as being Abelian.

9.1.1. *Metrics versus pseudo-metrics.* A pseudo-metric $\boldsymbol{\delta}$ on a set X is a bivariate function satisfying all the axioms of a distance function except it is non-negative definite rather than positive definite. That is, the pseudo-distance function is allowed to vanish on pairs of non-identical points. A set equipped with a pseudo-metric will be called a pseudo-metric space. An isometry of such spaces is a distance-preserving map, such that for any point in the target space there is a point in the image, which is distance zero from it. Given such an pseudo-metric space $(X, \boldsymbol{\delta})$ one could always construct an isometric metric space $(X/_{\boldsymbol{\delta}=0}, \boldsymbol{\delta})$ by identifying all pairs of points that are distance zero apart.

Any property formulated in terms of the pseudo-metric holds simultaneously for a pseudo-metric space and its metric quotient. It will be convenient for us to construct pseudo-metrics on spaces instead of passing to the quotient spaces.

9.1.2. *Asymptotic cone of a metric Abelian monoid.* Let $(\Gamma, \otimes, \boldsymbol{\delta})$ be a monoid with a pseudo-metric $\boldsymbol{\delta}$, which satisfies the following properties

(i) The shifts

$$\cdot \otimes \gamma' : \Gamma \to \Gamma, \quad \gamma \mapsto \gamma \otimes \gamma'$$

are non-expanding for any $\gamma' \in \Gamma$

(ii) For any $\gamma, \gamma' \in \Gamma$ holds

$$\boldsymbol{\delta}(\gamma \otimes \gamma', \gamma' \otimes \gamma) = 0$$

We will call a monoid with pseudo-metric that satisfies these conditions a *metric Abelian monoid.* It follows from the shift-invariance property that for any $\gamma_1, \gamma_2, \gamma_3 \in \Gamma$ holds

$$\boldsymbol{\delta}(\gamma_1 \otimes \gamma_3, \gamma_2 \otimes \gamma_3) \leq \boldsymbol{\delta}(\gamma_1, \gamma_2) \tag{9.1}$$

and for any quadruple $\gamma_1, \gamma_2, \gamma_3, \gamma_4 \in \Gamma$ holds

$$\boldsymbol{\delta}(\gamma_1 \otimes \gamma_2, \gamma_3 \otimes \gamma_4) \leq \boldsymbol{\delta}(\gamma_1, \gamma_3) + \boldsymbol{\delta}(\gamma_2, \gamma_4) \tag{9.2}$$

and, in particular, the monoid operation is 1-Lipschitz with respect to each argument.

As a direct consequence, for every $n \in \mathbb{N}$, and $\gamma_1, \gamma_2 \in \Gamma$ also holds

$$\boldsymbol{\delta}(\gamma_1^{\otimes n}, \gamma_2^{\otimes n}) \leq n\, \boldsymbol{\delta}(\gamma_1, \gamma_2) \tag{9.3}$$

For a sequence $\overline{\gamma} = \{\gamma(i)\} \in \Gamma^{\mathbb{N}_0}$ define its *defect* with respect to the distance function $\boldsymbol{\delta}$ by

$$\mathrm{Defect}_{\boldsymbol{\delta}}(\overline{\gamma}) = \sup_{i,j \in \mathbb{N}_0} \boldsymbol{\delta}\left(\gamma(i+j), \gamma(i) \otimes \gamma(j)\right)$$

The sequence $\overline{\gamma}$ will be called $\boldsymbol{\delta}$*-linear* if $\operatorname{Defect}_{\boldsymbol{\delta}}(\overline{\gamma}) = 0$, and $\boldsymbol{\delta}$*-quasi-linear* if $\operatorname{Defect}_{\boldsymbol{\delta}}(\overline{\gamma}) < \infty$. Denote by $\mathsf{L}_{\boldsymbol{\delta}}(\Gamma)$ and $\mathsf{QL}_{\boldsymbol{\delta}}(\Gamma)$ the sets of all linear and, respectively, quasi-linear sequences in Γ with respect to the distance $\boldsymbol{\delta}$.

For two elements $\overline{\gamma}_1, \overline{\gamma}_2 \in \mathsf{QL}_d(\Gamma)$, define an asymptotic distance between them by

$$\hat{\boldsymbol{\delta}}(\overline{\gamma}_1, \overline{\gamma}_2) := \lim_{n\to\infty} \frac{1}{n} \boldsymbol{\delta}\left(\gamma_1(n), \gamma_2(n)\right)$$

Lemma 9.1. ↓ *For a pair* $\overline{\gamma}_1, \overline{\gamma}_2 \in \mathsf{QL}_{\boldsymbol{\delta}}(\Gamma)$ *the limit*

$$\lim_{n\to\infty} \frac{1}{n} \boldsymbol{\delta}\left(\gamma_1(n), \gamma_2(n)\right)$$

exists and is finite. ⊠

We provide the proof in Section T on page 72.

The bivariate function $\hat{\boldsymbol{\delta}}$ is a pseudo-distance on the set $\mathsf{QL}_{\boldsymbol{\delta}}(\Gamma)$. We call two sequences $\overline{\gamma}_1, \overline{\gamma}_2 \in \mathsf{QL}_{\boldsymbol{\delta}}(\Gamma)$ *asymptotically equivalent* if $\hat{\boldsymbol{\delta}}(\overline{\gamma}_1, \overline{\gamma}_2) = 0$ and write

$$\overline{\gamma}_1 \overset{\hat{\boldsymbol{\delta}}}{=} \overline{\gamma}_2$$

We will call a sequence $\overline{\gamma}$ *weakly quasi-linear*, if it is asymptotically equivalent to a quasi-linear sequence. Note that the space of all weakly quasi-linear sequences can also be endowed with the asymptotic distance and it is isometric to the space of quasi-linear sequences. As we will see later all the natural operations we consider are $\hat{\boldsymbol{\delta}}$-Lipschitz and therefore coincide for the asymptotically equivalent sequences. Thus given a weakly quasi-linear sequence we could always replace it by an equivalent quasi-linear sequence without any visible effect. Thus, we take the liberty to omit the adverb "weakly". Whenever we say quasi-linear sequence, we mean a weakly quasi-linear sequence, that is silently replaced by an asymptotically equivalent genuine quasi-linear sequence, if necessary.

The validity of the following constructions is very easy to verify, so we omit the proofs.

The set $\mathsf{QL}_{\boldsymbol{\delta}}(\Gamma)$ admits an action of the multiplicative semigroup $(\mathbb{R}_{\geq 0}, \cdot)$ defined in the following way. Let $\lambda \in \mathbb{R}_{\geq 0}$ and $\overline{\gamma} = \{\gamma(n)\} \in \mathsf{QL}_{\boldsymbol{\delta}}(\Gamma)$. Then define the action of λ on $\overline{\gamma}$ by

$$\overline{\gamma}^{\lambda} := \{\gamma(\lfloor \lambda \cdot n \rfloor)\}_{n\in\mathbb{N}_0} \tag{9.4}$$

This is only an action up to asymptotic equivalence. Similarly, in the constructions that follow we are tacitly assuming they are valid up to asymptotic equivalence.

The action

$$\cdot : \mathbb{R}_{\geq 0} \times (\mathsf{QL}_{\boldsymbol{\delta}}(\Gamma), \hat{\boldsymbol{\delta}}) \to (\mathsf{QL}_{\boldsymbol{\delta}}(\Gamma), \hat{\boldsymbol{\delta}})$$

is continuous with respect to $\hat{\boldsymbol{\delta}}$ and, moreover it is a homothety (dilation), that is

$$\hat{\boldsymbol{\delta}}(\overline{\gamma}_1^{\lambda}, \overline{\gamma}_2^{\lambda}) = \lambda \cdot \hat{\boldsymbol{\delta}}(\overline{\gamma}_1, \overline{\gamma}_2)$$

The group operation $\otimes$ on Γ induces a $\hat{\boldsymbol{\delta}}$-continuous (in fact, 1-Lipschitz) group operation on $\mathsf{QL}_{\boldsymbol{\delta}}(\Gamma)$ by multiplying sequences element-wise. The semigroup structure on $\mathsf{QL}_{\boldsymbol{\delta}}(\Gamma)$ is distributive with respect to the $\mathbb{R}_{\geq 0}$-action

$$(\overline{\gamma}_1 \otimes \overline{\gamma}_2)^\lambda = \overline{\gamma}_1^\lambda \otimes \overline{\gamma}_2^\lambda$$
$$\overline{\gamma}^{\lambda_1+\lambda_2} \overset{\hat{\boldsymbol{\delta}}}{=} \overline{\gamma}^{\lambda_1} \otimes \overline{\gamma}^{\lambda_2}$$

In particular for $n \in \mathbb{N}$

$$\overline{\gamma}^n \overset{\hat{\boldsymbol{\delta}}}{=} \overline{\gamma}^{\otimes n}$$

The path

$$[0,1] \ni \lambda \mapsto \overline{\gamma}_1^{1-\lambda} \otimes \overline{\gamma}_2^\lambda$$

will be called a convex interpolation and is a constant-speed $\hat{\boldsymbol{\delta}}$-geodesic between $\overline{\gamma}_1$ and $\overline{\gamma}_2$, that is for $\lambda \in [0,1]$,

$$\hat{\boldsymbol{\delta}}(\overline{\gamma}_1^{(1-\lambda)} \otimes \overline{\gamma}_2^\lambda, \overline{\gamma}_1) = \lambda\, \hat{\boldsymbol{\delta}}(\overline{\gamma}_1, \overline{\gamma}_2)$$
$$\hat{\boldsymbol{\delta}}(\overline{\gamma}_1^{(1-\lambda)} \otimes \overline{\gamma}_2^\lambda, \overline{\gamma}_2) = (1-\lambda)\, \hat{\boldsymbol{\delta}}(\overline{\gamma}_1, \overline{\gamma}_2)$$

9.1.3. *Conditions for completeness.* We would like to call

$$\Gamma_{\boldsymbol{\delta}}^{(\infty)} := (\mathsf{QL}_{\boldsymbol{\delta}}(\Gamma), \otimes, \cdot, \hat{\boldsymbol{\delta}})$$

the asymptotic cone of $(\Gamma, \otimes, \boldsymbol{\delta})$. However it is not clear in general, whether $\Gamma_{\boldsymbol{\delta}}^{(\infty)}$ is a complete space.

We can simply consider the metric completion, and call it the asymptotic cone of $(\Gamma, \otimes, \boldsymbol{\delta})$. We feel, however, that it adds just another level of obscurity as to what the points of $\Gamma_{\boldsymbol{\delta}}^{(\infty)}$ are.

Under some circumstances, however, the completeness of the space of quasi-linear sequences comes for free. This is the subject of the proposition below.

Suppose the metric Abelian monoid $(\Gamma, \otimes, \boldsymbol{\delta})$ has an additional property: There exists a constant $C > 0$, such that for any quasi-linear sequence $\overline{\gamma} \in \mathsf{QL}_{\boldsymbol{\delta}}(\Gamma)$, there exists an asymptotically equivalent quasi-linear sequence $\overline{\gamma}'$ with defect bounded by C. If this is the case, we say that the metric monoid $(\Gamma, \otimes, \boldsymbol{\delta})$ has the (C-)*uniformly bounded defect property.*

Proposition 9.2. ↓ *Suppose* $(\Gamma, \otimes, \boldsymbol{\delta})$ *is a metric Abelian monoid such that*

(i) *the distance function* $\boldsymbol{\delta}$ *is homogeneous, that is for any* $\gamma_1, \gamma_2 \in \Gamma$ *and* $n \in \mathbb{N}_0$

$$\boldsymbol{\delta}(\gamma_1^{\otimes n}, \gamma_2^{\otimes n}) = n \cdot \boldsymbol{\delta}(\gamma_1, \gamma_2)$$

(ii) $(\Gamma, \otimes, \boldsymbol{\delta})$ *has the uniformly bounded defect property.*

Then the space $(\mathsf{QL}_{\boldsymbol{\delta}}(\Gamma), \hat{\boldsymbol{\delta}})$ *is complete.* ⊠

The proof of the proposition can be found on page 72.

9.1.4. *On the density of linear sequences.* In Section 6 we have shown that Bernoulli sequences of configurations can be approximated by sequences of homogeneous configurations. The proposition below will allow us to extend this statement to a wider class of sequences. It gives a sufficient condition under which the linear sequences are dense in the quasi-linear sequences.

Proposition 9.3. ↓ *Suppose* $(\Gamma, \otimes, \boldsymbol{\delta})$ *has the* ε*-uniformly bounded defect property for every* $\varepsilon > 0$*. Then* $\mathsf{L}_{\boldsymbol{\delta}}(\Gamma)$ *is dense in* $\mathsf{QL}_{\boldsymbol{\delta}}(\Gamma)$ ⊠

See page 74 for the proof.

9.1.5. *Asymptotic metric on original semigroup.* Starting with an element $\gamma \in \Gamma$ one can construct a linear sequence $\vec{\gamma} = \{\gamma^{\otimes i}\}_{i \in \mathbb{N}_0}$. In view of inequality (9.3), this map is a contraction

$$\big(\Gamma, \boldsymbol{\delta}\,\big) \to \big(\,\mathsf{L}_{\boldsymbol{\delta}}(\Gamma), \hat{\boldsymbol{\delta}}\,\big) \tag{9.5}$$

By the inclusions in (9.5) we have an induced metric $\hat{\boldsymbol{\delta}}$ on Γ, satisfying for any $\gamma_1, \gamma_2 \in \Gamma$

$$\hat{\boldsymbol{\delta}}(\gamma_1, \gamma_2) \le \boldsymbol{\delta}(\gamma_1, \gamma_2) \tag{9.6}$$

and the following scale-invariance condition is gained

$$\hat{\boldsymbol{\delta}}(\gamma_1^{\otimes n}, \gamma_2^{\otimes n}) = n \cdot \hat{\boldsymbol{\delta}}(\gamma_1, \gamma_2) \tag{9.7}$$

for all $n \in \mathbb{N}_0$.

Note moreover that if $\boldsymbol{\delta}$ was scale-invariant to begin with, then $\hat{\boldsymbol{\delta}}$ coincides with $\boldsymbol{\delta}$ on Γ.

9.1.6. *Iteration of construction.* We may now iterate the constructions above, that is, we may apply them to $(\Gamma, \hat{\boldsymbol{\delta}})$ instead of $(\Gamma, \boldsymbol{\delta})$. One may wonder what is the purpose. However, we have already observed that $\hat{\boldsymbol{\delta}}$ satisfies the scale-invariance condition (9.7), which is one of the conditions going into a proof of completeness in Proposition 9.2. Moreover, when we will later apply the theory in this section to the particular case of $\Gamma = \mathbf{Prob}\,\langle \mathbf{G} \rangle$, we will see that

$$(\Gamma, \hat{\boldsymbol{\delta}}) = (\mathbf{Prob}\,\langle \mathbf{G} \rangle\,, \boldsymbol{\kappa})$$

and we will show that the latter space has the ε-uniformly bounded defect property for every $\varepsilon > 0$.

By virtue of the bound $\hat{\boldsymbol{\delta}} \le \boldsymbol{\delta}$, sequences that are quasi-linear with respect to $\hat{\boldsymbol{\delta}}$, are also quasi-linear with respect to $\boldsymbol{\delta}$. Since $\hat{\boldsymbol{\delta}}$ is scale-invariant, the associated asymptotic distance $\hat{\hat{\boldsymbol{\delta}}}$ coincides with $\hat{\boldsymbol{\delta}}$ on Γ. We will show (in Lemma 9.4 below) that $\hat{\hat{\boldsymbol{\delta}}}$ also corresponds to $\hat{\boldsymbol{\delta}}$ on $\boldsymbol{\delta}$-quasi-linear sequences.

In order to organize all these statements, and to be more precise, let us include the spaces in the following commutative diagram.

$$
\begin{tikzcd}
 & \big(\mathsf{L}_{\boldsymbol{\delta}}(\Gamma), \hat{\boldsymbol{\delta}}\big) \arrow[r, hook, "\jmath_1"] \arrow[d, hook, "\imath_1"] & \big(\mathsf{QL}_{\boldsymbol{\delta}}(\Gamma), \hat{\boldsymbol{\delta}}\big) \arrow[d, hook, "\imath_2"] \\
(\Gamma, \hat{\boldsymbol{\delta}}) \arrow[ur, "f"] \arrow[dr, "\varphi"'] & & \\
 & \big(\mathsf{L}_{\hat{\boldsymbol{\delta}}}(\Gamma), \hat{\hat{\boldsymbol{\delta}}}\big) \arrow[r, hook, "\jmath_2"] & \big(\mathsf{QL}_{\hat{\boldsymbol{\delta}}}(\Gamma), \hat{\hat{\boldsymbol{\delta}}}\big)
\end{tikzcd}
\tag{9.8}
$$

The maps f, φ and $\imath_1$ are isometries. The maps $\jmath_1$ and $\jmath_2$ are isometric embeddings. The next lemmas show that $\imath_2$ is also an isometric embedding, and it has dense image.

Lemma 9.4. ↓ *The natural inclusion*

$$\imath_2 : (\mathsf{QL}_{\boldsymbol{\delta}}(\Gamma), \hat{\boldsymbol{\delta}}) \hookrightarrow (\mathsf{QL}_{\hat{\boldsymbol{\delta}}}(\Gamma), \hat{\hat{\boldsymbol{\delta}}})$$

is an isometric embedding. ⊠

Lemma 9.5. ↓ *The image of the isometric embedding*

$$\imath_2 : (\mathsf{QL}_{\boldsymbol{\delta}}(\Gamma), \hat{\boldsymbol{\delta}}) \hookrightarrow (\mathsf{QL}_{\hat{\boldsymbol{\delta}}}(\Gamma), \hat{\hat{\boldsymbol{\delta}}})$$

is dense in $(\mathsf{QL}_{\hat{\boldsymbol{\delta}}}(\Gamma), \hat{\hat{\boldsymbol{\delta}}})$ ⊠

The proofs of the two lemmas above are to be found on page 75.

9.2. **Tropical probability spaces and configurations.** Now we apply the above construction to the space of complete configurations with fixed combinatorial type $\mathbf{G}$.

Fix a complete diagram category $\mathbf{G}$ and consider the space $\mathbf{Prob}\langle\mathbf{G}\rangle$ of configurations modeled on $\mathbf{G}$. It carries the following structures:

(i) A pseudo-metric $\mathbf{k}$ or $\boldsymbol{\kappa}$.
(ii) A 1-Lipschitz tensor product $\otimes$.
(iii) A 1-Lipschitz entropy function $\mathcal{E}nt_* : \mathbf{Prob}\langle\mathbf{G}\rangle \to \mathbb{R}^{[\![\mathbf{G}]\!]}$.

The tensor product of configurations is commutative from a metric perspective. Recall that in Corollary 4.5 the subadditivity of both $\mathbf{k}$ and $\boldsymbol{\kappa}$ was established, namely for any $\mathcal{X}, \mathcal{Y}, \mathcal{U}, \mathcal{V} \in \mathbf{Prob}\langle\mathbf{G}\rangle$ holds

$$\mathbf{k}(\mathcal{X} \otimes \mathcal{U}, \mathcal{Y} \otimes \mathcal{V}) \le \mathbf{k}(\mathcal{X}, \mathcal{Y}) + \mathbf{k}(\mathcal{U}, \mathcal{V}).$$

and

$$\boldsymbol{\kappa}(\mathcal{X} \otimes \mathcal{U}, \mathcal{Y} \otimes \mathcal{V}) \le \boldsymbol{\kappa}(\mathcal{X}, \mathcal{Y}) + \boldsymbol{\kappa}(\mathcal{U}, \mathcal{V}).$$

The space $(\mathbf{Prob}\langle\mathbf{G}\rangle, \otimes, \mathbf{k})$ is a metric Abelian monoid. Note also that $\hat{\mathbf{k}} = \boldsymbol{\kappa}$ on $\mathbf{Prob}\langle\mathbf{G}\rangle$, along the lines of Section 9.1.5.

However, the metric $\mathbf{k}$ is not scale-invariant. Moreover, it is unclear whether the metric semigroup $(\mathbf{Prob}\langle\mathbf{G}\rangle, \otimes, \mathbf{k})$ has the uniformly bounded defect property. This is why we iterate the construction, as announced in Section 9.1.6, and consider the space of $\boldsymbol{\kappa}$-quasi-linear sequences instead.

Lemma 9.6. ↓ *For a complete diagram category, and for every $\varepsilon > 0$, the space $(\mathbf{Prob}\langle\mathbf{G}\rangle, \otimes, \boldsymbol{\kappa})$ has the ε-uniformly bounded defect property, that is for any $\boldsymbol{\kappa}$-quasi-linear sequence $\overline{\mathcal{X}} \in \mathsf{QL}_{\boldsymbol{\kappa}}(\mathbf{Prob}\langle\mathbf{G}\rangle)$ there exists an asymptotically equivalent sequence $\overline{\mathcal{Y}}$ with defect not exceeding ε.* ⊠

By applying the general setup in the previous section to the metric semi-groups $(\mathbf{Prob}\langle\mathbf{G}\rangle, \otimes, \mathbf{k})$ and $(\mathbf{Prob}\langle\mathbf{G}\rangle, \otimes, \boldsymbol{\kappa})$ and as a corollary to Lemma 9.6 we obtain the following theorem.

Theorem 9.7. *Consider the commutative diagram*
(9.9)

$$\begin{tikzcd}
 & \big(\mathsf{L}_{\mathbf{k}}(\mathbf{Prob}\langle\mathbf{G}\rangle), \hat{\mathbf{k}}\big) \arrow[r, hook, "\jmath_1"] \arrow[d, hook, "\imath_1"] & \big(\mathsf{QL}_{\mathbf{k}}(\mathbf{Prob}\langle\mathbf{G}\rangle), \hat{\mathbf{k}}\big) \arrow[d, hook, "\imath_2"] \\
(\mathbf{Prob}\langle\mathbf{G}\rangle, \boldsymbol{\kappa}) \arrow[ur, "f"] \arrow[dr, "\varphi"] & & \\
 & \big(\mathsf{L}_{\boldsymbol{\kappa}}(\mathbf{Prob}\langle\mathbf{G}\rangle), \hat{\boldsymbol{\kappa}}\big) \arrow[r, hook, "\jmath_2"] & \big(\mathsf{QL}_{\boldsymbol{\kappa}}(\mathbf{Prob}\langle\mathbf{G}\rangle), \hat{\boldsymbol{\kappa}}\big)
\end{tikzcd}$$

Then the following statements hold:

(i) *The maps $f, \varphi, \imath_1$ are isometries.*
(ii) *The maps $\imath_2, \jmath_1, \jmath_2$ are isometric embeddings and each map has a dense image in the corresponding target space.*
(iii) *The space in the lower-right corner, $\big(\mathsf{QL}_{\boldsymbol{\kappa}}(\mathbf{Prob}\langle\mathbf{G}\rangle), \hat{\boldsymbol{\kappa}}\big)$, is complete.*

⊠

We may finally define the space of *tropical* $\mathbf{G}$*-configurations*, as the space in the lower-right corner of the diagram

$$\mathbf{Prob}\langle\mathbf{G}\rangle^{(\infty)} := \big(\mathsf{QL}_{\boldsymbol{\kappa}}(\mathbf{Prob}\langle\mathbf{G}\rangle), \otimes, \cdot, \hat{\boldsymbol{\kappa}}\big)$$

By the Theorem 9.7 above, this space is complete.

The entropy function $\mathcal{E}nt_* : \mathbf{Prob}\langle\mathbf{G}\rangle \to \mathbb{R}^{[\![\mathbf{G}]\!]}$ extends to a linear functional

$$\mathcal{E}nt_* : \mathbf{Prob}\langle\mathbf{G}\rangle^{(\infty)} \to (\mathbb{R}^{[\![\mathbf{G}]\!]}, |\cdot|_1)$$

of norm one, defined by

$$\mathcal{E}nt_*(\overline{\mathcal{X}}) = \lim_{n\to\infty} \frac{1}{n} \mathcal{E}nt_*\big(\mathcal{X}(n)\big)$$

9.2.1. *Sequences of homogeneous configurations are dense.* Let $\tilde{\mathsf{L}}_{\mathbf{k}}(\mathbf{Prob}\langle\mathbf{G}\rangle_{\mathbf{h}})$ stand for the weakly linear sequences of homogeneous configurations, that is those sequences, that are asymptotically equivalent to a linear sequence (not necessarily of homogeneous spaces).

For a sequence of homogeneous spaces $\overline{\mathcal{H}} \in \tilde{\mathsf{L}}_{\mathbf{k}}(\mathbf{Prob}\langle\mathbf{G}\rangle_{\mathbf{h}})$ define $\mathbf{aep}(\overline{\mathcal{H}})$ to be a $\mathbf{k}$-linear sequence asymptotically equivalent to $\overline{\mathcal{H}}$.

Now we can extend the commutative diagram (9.9) as follows

$$\begin{tikzcd}
\tilde{\mathsf{L}}_{\mathbf{k}}(\mathbf{Prob}\langle\mathbf{G}\rangle_{\mathbf{h}}) \arrow[r, "\mathbf{aep}"] & \mathsf{L}_{\mathbf{k}}(\mathbf{Prob}\langle\mathbf{G}\rangle) \arrow[r, hook, "\jmath_1"] \arrow[dd, hook, "\imath_1"] & \mathsf{QL}_{\mathbf{k}}(\mathbf{Prob}\langle\mathbf{G}\rangle) \arrow[dd, hook, "\imath_2"] \\
(\mathbf{Prob}\langle\mathbf{G}\rangle, \boldsymbol{\kappa}) \arrow[ur, "f"] \arrow[dr, "\varphi"'] & & \\
& \mathsf{L}_{\boldsymbol{\kappa}}(\mathbf{Prob}\langle\mathbf{G}\rangle) \arrow[r, hook, "\jmath_2"] & \mathbf{Prob}\langle\mathbf{G}\rangle^{(\infty)}
\end{tikzcd} \tag{9.10}$$

By the Asymptotic Equipartition Property for configurations, Theorem 6.1, the map **aep** is an isometry, hence we have the following theorem.

Theorem 9.8. *The map*

$$\jmath_2 \circ \imath_1 \circ \mathbf{aep} : \tilde{\mathsf{L}}_{\mathbf{k}}(\mathbf{Prob}\langle\mathbf{G}\rangle_{\mathbf{h}}) \to \mathbf{Prob}\langle\mathbf{G}\rangle^{(\infty)}$$

is an isometric embedding with dense image. ⊠

Let $\mathbf{Prob}\langle\mathbf{G}\rangle^{(\infty)}_{\mathbf{h}} \subset \mathbf{Prob}\langle\mathbf{G}\rangle^{(\infty)}$ denote the space of weakly quasi-linear sequences of configurations $\overline{\mathcal{H}} \in \mathbf{Prob}\langle\mathbf{G}\rangle^{(\infty)}$, such that for every $n \in \mathbb{N}_0$, the configuration $\mathcal{H}(n)$ is homogeneous. We will refer to $\mathbf{Prob}\langle\mathbf{G}\rangle^{(\infty)}_{\mathbf{h}}$ as the space of homogeneous tropical configurations.

Denote by **aep** the embedding

$$\mathbf{aep} : \mathbf{Prob}\langle\mathbf{G}\rangle^{(\infty)}_{\mathbf{h}} \hookrightarrow \mathbf{Prob}\langle\mathbf{G}\rangle^{(\infty)}$$

Theorem 9.9. (Asymptotic Equipartition Theorem for tropical configurations) *Let* $\mathbf{G}$ *be a complete diagram category. Then the map*

$$\mathbf{aep} : \mathbf{Prob}\langle\mathbf{G}\rangle^{(\infty)}_{\mathbf{h}} \hookrightarrow \mathbf{Prob}\langle\mathbf{G}\rangle^{(\infty)}$$

is an isometry. ⊠

Proof: We need to show that for every tropical configuration $\overline{\mathcal{X}} \in \mathbf{Prob}\langle\mathbf{G}\rangle^{(\infty)}$, there exists a homogeneous tropical configuration $\overline{\mathcal{H}} \in \mathbf{Prob}\langle\mathbf{G}\rangle^{(\infty)}_{\mathbf{h}}$ such that

$$\hat{\boldsymbol{\kappa}}(\overline{\mathcal{H}}, \overline{\mathcal{X}}) = 0$$

By Lemma 9.6 and Proposition 9.3, for every $j \in \mathbb{N}$ there exists a sequence $\overline{\mathcal{Y}}_j \in \mathsf{L}_{\boldsymbol{\kappa}}(\mathbf{Prob}\langle\mathbf{G}\rangle) \cong \mathsf{L}_{\mathbf{k}}(\mathbf{Prob}\langle\mathbf{G}\rangle)$ such that

$$\hat{\boldsymbol{\kappa}}(\overline{\mathcal{Y}}_j, \overline{\mathcal{X}}) \leq \frac{1}{j}$$

By the Asymptotic Equipartition Property for configurations, Theorem 6.1, there are sequences of homogeneous configurations $\overline{\mathcal{H}}_j$ such that

$$\hat{\boldsymbol{\kappa}}(\overline{\mathcal{Y}}_j, \overline{\mathcal{H}}_j) = 0$$

Define $\mathbf{i}(j)$ such that for all $k \geq \mathbf{i}(j)$

$$\frac{1}{k}\boldsymbol{\kappa}(\mathcal{Y}_j(k), \mathcal{H}_j(k)) \leq \frac{1}{j}$$

and moreover

$$\frac{1}{k}\boldsymbol{\kappa}(\mathcal{Y}_j(k),\mathcal{X}(k)) \leq \frac{2}{j}$$

The function $\mathbf{i}$ can be chosen monotonically increasing. For every $i \in \mathbb{N}_0$ there is a unique $\mathbf{j}(i) \in \mathbb{N}_0$ such that

$$\mathbf{i}\big(\mathbf{j}(i)\big) \leq i < \mathbf{i}\big(\mathbf{j}(i)+1\big)$$

Define then

$$\mathcal{H}(k) = \mathcal{H}_{\mathbf{j}(k)}(k)$$

It follows that for $k > \mathbf{i}(1)$,

$$\frac{1}{k}\boldsymbol{\kappa}(\mathcal{H}(k),\mathcal{X}(k)) \leq \frac{3}{\mathbf{j}(k)}$$

Since $\mathbf{j}$ is a non-decreasing, divergent sequence, the theorem follows. ⊠

Thus we have shown that all arrows in diagram (9.10) are isometric embeddings with dense images. We would like to conjecture, that, in fact, they are all isometries. In any case, the difference between metrics $\hat{\boldsymbol{\kappa}}$ and $\boldsymbol{\kappa}$ is so small ($\boldsymbol{\kappa}$ is defined on the dense subset of the domain of definition of $\hat{\boldsymbol{\kappa}}$ and they coincide whenever both are defined), that we will not write the hat anymore and just use notation $\boldsymbol{\kappa}$ for the metric on $\mathbf{Prob}\langle\mathbf{G}\rangle^{(\infty)}$.

9.3. **Tropical probability spaces and tropical chains.** In this section we evaluate the spaces $\mathbf{Prob}^{(\infty)}$ and $\mathbf{Prob}\langle\mathbf{C}_n\rangle^{(\infty)}$, where $\mathbf{C}_n$ is a chain, which is the diagram category introduced in 2.1.2 on page 14.

Recall that a finite probability space U is homogeneous if $\mathrm{Aut}(U)$ acts transitively on the support of the measure. The property of being homogeneous is invariant under isomorphism and every homogeneous space is isomorphic to a probability space with the uniform distribution.

Homogeneous chains also have a very simple description. A chain of reductions is homogeneous, if and only if all the individual spaces are homogeneous.

This simple description allows us to evaluate explicitly the Kolmogorov distance on the spaces of weakly linear sequences of homogeneous chains and consequently the space of tropical chains.

Theorem 9.10.

(i) $\mathbf{Prob}^{(\infty)} \cong (\mathbb{R}_{\geq 0}, |\cdot - \cdot|, +, \cdot)$

(ii) $\mathbf{Prob}\langle\mathbf{C}_n\rangle^{(\infty)} \cong \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n \;\middle\|\; 0 \leq x_n \leq \cdots \leq x_1 \right\}$

where the right-hand side is a cone in $(\mathbb{R}^n, |\cdot|_1)$.

⊠

To prove the Theorem 9.10 we evaluate first the isometry class of the space of weakly linear sequences of homogeneous spaces (or chains). We will only present an argument for single spaces, since the argument for chains is very similar.

Lemma 9.11.
$$\tilde{\mathsf{L}}_{\mathbf{k}}(\mathbf{Prob}_{\mathbf{h}}) \cong (\mathbb{R}_{\geq 0}, |\cdot - \cdot|, +, \cdot)$$

$\boxtimes$

Note that the right-hand side is a complete metric space, thus the Asymptotic Equipartition Theorem for tropical configurations, Theorem 9.8, together with Lemma 9.11, imply Theorem 9.10.

To prove Lemma 9.11 we need to evaluate the Kolmogorov distance between two homogeneous spaces, or chains of homogeneous spaces. This is the subject of the next lemma, from which Lemma 9.11 follows immediately.

Lemma 9.12. $\downarrow$ *Denote by U_n a finite uniform probability space of cardinality n, then*

(i)
$$\mathbf{k}(U_n, U_m) \leq 2 \ln 2 + \left| \ln \frac{n}{m} \right|$$

(ii)
$$\boldsymbol{\kappa}(U_n, U_m) = |\mathcal{E}nt(U_n) - \mathcal{E}nt(U_m)|$$

$\boxtimes$

9.4. **Stochastic processes.** Often, stochastic processes naturally give rise to $\boldsymbol{\kappa}$-quasi-linear sequences. We include this last subsection as an indication that our statements, together with the construction of the tropical cone, have a much larger reach than sequences of independent random variables. We will be brief, and come back to the topic in a subsequent article.

For a minimal diamond configuration

$$\begin{tikzcd}
 & C \arrow[dl] \arrow[dr] \arrow[dd] & \\
A \arrow[dr] & & B \arrow[dl] \\
 & D &
\end{tikzcd}$$

we define the conditional mutual information between A and B given D by

$$\mathcal{I}(A; B \rfloor D) := \mathcal{E}nt(A) + \mathcal{E}nt(B) - \mathcal{E}nt(C) - \mathcal{E}nt(D)$$

Shannon's inequality (1.2) says that the conditional mutual information is always non-negative. Any minimal two-fan $A \leftarrow C \rightarrow B$ can be completed to a diamond with the one-point probability space $\{\bullet\}$ as the terminal vertex, and the mutual information between A and B is defined as

$$\mathcal{I}(A; B) = \mathcal{I}(A; B \rfloor \{\bullet\}) = \mathcal{E}nt(A) + \mathcal{E}nt(B) - \mathcal{E}nt(C)$$

Let

$$\dots, X_{-1}, X_0, X_1, \dots$$

be a stationary stochastic process with finite state space $\underline{X}$. Thus, for any $I = \{k, k+1, \dots, l\} \subset \mathbb{Z}$ we have jointly distributed random variables $X_k, \dots, X_l$, that generate a full configuration

$$\mathcal{X}_I = \langle X_k, \dots, X_l \rangle = \{X_J\}_{J \subset I}$$

as explained in Section 2.1.4. This collection of full configurations is consistent in the sense that for $k \le k' \le l' \le l$ there are canonical isomorphisms

$$\mathcal{X}_{I'} \cong R^*_{I',I} \mathcal{X}_I$$

where $I := \{k, \dots, l\}$, $I' := \{k', \dots, l'\}$ and $R^*_{I',I}$ is the restriction operator introduced in Section 2.4.1.

The property of being stationary means that there are canonical isomorphisms for any finite subset $I \subset \mathbb{Z}$ and $l \in \mathbb{Z}$

$$\mathcal{X}_I \xrightarrow{\cong} \mathcal{X}_{I+l}$$

For $I = \{k, k+1, \dots, l\}$ we call the initial space X_I of the configuration $\mathcal{X}_I$ the space of trajectories of the process over I and denote it X_k^l.

Note that by stationarity, for every $m \in \mathbb{Z}$, $k \in \mathbb{N}$ and $l \in \mathbb{N}_0$,

$$\mathcal{I}(X_m^{m+k-1}; X_{m+k}^{m+k+l-1}) = \mathcal{I}(X_{-k+1}^0; X_1^l)$$

Moreover the right-hand side is an increasing function of both k and l. We make the following important observation. The defect of the sequence $n \mapsto X_1^n$ is equal to

$$\begin{aligned}
\mathrm{Defect}_{\boldsymbol{\kappa}}(\{X_1^n\}) &= \sup_{m,n\in\mathbb{N}_0} \boldsymbol{\kappa}(X_1^{m+n}, X_1^m \otimes X_1^n) \\
&= \sup_{m,n\in\mathbb{N}_0} \left|\mathcal{E}nt(X_1^m \otimes X_1^n) - \mathcal{E}nt(X_1^{m+n})\right| \\
&= \sup_{m,n\in\mathbb{N}_0} \left|\mathcal{E}nt(X_{-m+1}^0 \otimes X_1^n) - \mathcal{E}nt(X_1^{m+n})\right| \\
&= \sup_{m,n\in\mathbb{N}_0} \mathcal{I}(X_{-m+1}^0, X_1^n)
\end{aligned}$$

Therefore, the sequence $n \mapsto X_1^n$ is $\boldsymbol{\kappa}$-quasi-linear if and only if

$$\lim_{k,l\to\infty} \mathcal{I}(X_{-k+1}^0, X_1^l) < \infty \tag{9.11}$$

Once condition (9.11) is satisfied for a stochastic process, it defines a tropical probability space $\overline{X} \in \mathbf{Prob}$.

Note that condition (9.11) is satisfied for any stationary, finite-state Markov chains.

T. Technical proofs

This section contains some proofs that did not make it into the main text.

T.2. Statements from the section "Configurations".

Lemma 2.1. ↑ *Let* **G** *be a diagram category. Then*

(i) *A two-fan $\mathcal{F} = (\mathcal{X} \leftarrow \mathcal{Z} \rightarrow \mathcal{Y})$ of* **G**-*configurations is minimal, if and only if the constituent two-fans of probability spaces $\mathcal{F}_i = (X_i \leftarrow Z_i \rightarrow Y_i)$ are all minimal.*

(ii) *For any two-fan $\mathcal{F} = (\mathcal{X} \leftarrow \mathcal{Z} \rightarrow \mathcal{Y})$ of* **G**-*configurations its minimal reduction exists, that is, there exists a minimal two-fan $\mathcal{F}' = (\mathcal{X} \leftarrow \mathcal{Z}' \rightarrow \mathcal{Y})$ included in the following diagram*

$$\begin{tikzcd}
& \mathcal{Z} \arrow[d] \arrow[ddl] \arrow[ddr] & \\
& \mathcal{Z}' \arrow[dl] \arrow[dr] & \\
\mathcal{X} & & \mathcal{Y}
\end{tikzcd}$$

⊠

Proof: We will need the following lemma

Lemma T.1. *Suppose we are given two two-fans of probability spaces*

$$\mathcal{F} = (X \xleftarrow{\alpha} Z \xrightarrow{\beta} Y)$$

$$\mathcal{F}'' = (X'' \xleftarrow{\alpha''} Z'' \xrightarrow{\beta''} Y'')$$

such that $\mathcal{F}''$ is minimal. Let

$$\mathcal{F} \xrightarrow{\mu} \mathcal{F}' = (X \xleftarrow{\alpha'} Z' \xrightarrow{\beta'} Y)$$

be a minimal reduction of $\mathcal{F}$. Then for any reduction $\rho : \mathcal{F} \rightarrow \mathcal{F}''$, there exists a reduction $\rho' : \mathcal{F}' \rightarrow \mathcal{F}''$ such that $\rho = \rho' \circ \mu$ ⊠

Proof: We define ρ' on the terminal spaces of $\mathcal{F}'$ to coincide with ρ.

To prove the lemma we just need to provide a dashed arrow that makes the following diagram commutative

$$\begin{tikzcd}
& Z \arrow[d, dashed, "\mu"] \arrow[ddl, "\alpha"'] \arrow[ddr, "\beta"] \arrow[drrr, "\rho"] & & & & \\
& Z' \arrow[dl, dashed, "\alpha'"] \arrow[dr, dashed, "\beta'"'] \arrow[rrr, dashed, "\rho'"] & & & Z'' \arrow[dl, dashed, "\alpha''"'] \arrow[dr, dashed, "\beta''"] & \\
X \arrow[rrr, bend right, "\rho=\rho'"'] & & Y \arrow[rrr, bend right, "\rho=\rho'"'] & X'' & & Y''
\end{tikzcd}$$

The reduction ρ' is constructed by simple diagram chasing and by using the minimality of $\mathcal{F}''$. Suppose $z' \in \underline{Z'}$ and $z_1, z_2 \in \underline{Z}$ are such that $z' = \mu(z_1) = \mu(z_2)$. By commutativity of the solid arrows in the diagram above, we have

$$\alpha'' \circ \rho(z_1) = \rho \circ \alpha' \circ \mu(z_1) = \rho \circ \alpha' \circ \mu(z_2) = \alpha'' \circ \rho(z_2)$$

Similarly

$$\beta'' \circ \rho(z_1) = \beta'' \circ \rho(z_2)$$

Thus by minimality of $\mathcal{F}''$ it follows that $\rho(z_1) = \rho(z_2)$. Hence, ρ' can be constructed by setting $\rho'(z') = \rho(z_1)$. ⊠

Now we proceed to prove claim (i) of Lemma 2.1. Let $\mathbf{G} = \{O_i; m_{ij}\}$ be a diagram category, $\mathcal{X}, \mathcal{Y}, \mathcal{Z} \in \mathbf{Prob}\langle\mathbf{G}\rangle$ be three $\mathbf{G}$-configurations and $\mathcal{F} = (\mathcal{X} \leftarrow \mathcal{Z} \rightarrow \mathcal{Y})$ be a two-fan. Recall that it can also be considered as a $\mathbf{G}$-configuration of two-fans

$$\mathcal{F} = \{\mathcal{F}_i; f_{ij}\}$$

Any minimizing reduction

$$\mathcal{F} = (\mathcal{X} \leftarrow \mathcal{Z} \rightarrow \mathcal{Y}) \longrightarrow \mathcal{F}' = (\mathcal{X} \leftarrow \mathcal{Z}' \rightarrow \mathcal{Y})$$

induces reductions

$$\mathcal{F}_i = (X_i \leftarrow Z_i \rightarrow Y_i) \longrightarrow \mathcal{F}_i = (X_i \leftarrow Z'_i \rightarrow Y_i)$$

for all i in the index set I. It follows that if all $\mathcal{F}_i$'s are minimal, then so is $\mathcal{F}$.

Now we prove the implication in the other direction. Suppose $\mathcal{F}$ is minimal. We have to show that all $\mathcal{F}_i$ are minimal as well. Suppose there exist a non-minimal fan among $\mathcal{F}_i$'s. For an index $i \in I$ let

$$\check{J}(i) := \{ j \in I \,\|\, \mathrm{Hom}_{\mathbf{G}}(O_j, O_i) \neq \varnothing \}$$
$$\hat{J}(i) := \{ j \in I \,\|\, \mathrm{Hom}_{\mathbf{G}}(O_i, O_j) \neq \varnothing \}$$

Choose an index i_0 such that

(i) $\mathcal{F}_{i_0}$ is not minimal

(ii) for any $j \in \hat{J}(i_0) \backslash \{i_0\}$ the two-fan $\mathcal{F}_j$ is minimal.

Consider now the minimal reduction $\mu : \mathcal{F}_{i_0} \to \mathcal{F}'_{i_0}$ and construct a two-fan $\mathcal{G} = \{\mathcal{G}_i; g_{ij}\}$ of $\mathbf{G}$-configurations by setting

$$\mathcal{G}_i := \begin{cases} \mathcal{F}'_i & \text{if } i = i_0 \\ \mathcal{F}_i & \text{otherwise} \end{cases}$$

and

$$g_{ij} := \begin{cases} \mu \circ f_{ij} & \text{if } j = i_0 \text{ and } i \in \check{J}(i_0) \\ f'_{ij} & \text{if } i = i_0 \text{ and } j \in \hat{J}(i_0) \\ f_{ij} & \text{otherwise} \end{cases}$$

where $f'_{i_0 j}$ is the reduction provided by the Lemma T.1 applied to the diagram

$$\begin{tikzcd} \mathcal{F}_{i_0} \arrow[d, "\mu"] \arrow[dr, "f_{i_0 j}"] & \\ \mathcal{F}'_{i_0} \arrow[r, dashed, "f'_{i_0 j}"] & \mathcal{F}_j \end{tikzcd}$$

We thus constructed a non-trivial reduction $\mathcal{F} \to \mathcal{G}$ which is identity on the terminal $\mathbf{G}$-configurations $\mathcal{X}$ and $\mathcal{Y}$. This contradicts the minimality of $\mathcal{F}$.

To address the second assertion of the Lemma 2.1 observe that the argument above gives an algorithm for the construction of a minimal reduction of any two-fan of **G**-configurations. ⊠

T.4. Statements from the section "Kolmogorov distance".

Proposition 4.1. ↑ *Let* **G** *be a complete diagram category. Then the bivariate function*

$$\mathbf{k} : \mathbf{Prob}\langle \mathbf{G}\rangle \times \mathbf{Prob}\langle \mathbf{G}\rangle \to \mathbb{R}$$

is a pseudo-distance on $\mathbf{Prob}\langle \mathbf{G}\rangle$.
Moreover, two configurations $\mathcal{X},\mathcal{Y} \in \mathbf{Prob}\langle \mathbf{G}\rangle$ *satisfy* $\mathbf{k}(\mathcal{X},\mathcal{Y}) = 0$ *if and only if* $\mathcal{X}$ *is isomorphic to* $\mathcal{Y}$ *in* $\mathbf{Prob}\langle \mathbf{G}\rangle$. ⊠

Proof: The symmetry of **k** is immediate. The non-negativity of **k** follows from the fact that entropy of the target space of a reduction is not greater then the entropy of the domain, which is a particular instance of the Shannon inequality (1.2).

We proceed to prove the triangle inequality. We will make use of the following lemma

Lemma T.2. *For a minimal full configuration of probability spaces*

$$\langle X, Y, Z\rangle = \left(\begin{tikzcd} & & XYZ \arrow[dll] \arrow[d] \arrow[drr] & & \\ XY \arrow[d] \arrow[drr] & & XZ \arrow[dll] \arrow[drr] & & YZ \arrow[dll] \arrow[d] \\ X & & Y & & Z \end{tikzcd}\right)$$

holds

$$\mathrm{kd}(X \leftarrow XZ \to Z) \le \mathrm{kd}(X \leftarrow XY \to Y) + \mathrm{kd}(Y \leftarrow YZ \to Z)$$

⊠

Proof: By Shannon inequality (1.2) on page 13 we have

$$\mathcal{E}nt(X\lfloor Z) \le \mathcal{E}nt(XY\lfloor Z) \le \mathcal{E}nt(X\lfloor Y) + \mathcal{E}nt(Y\lfloor Z)$$

Similarly,

$$\mathcal{E}nt(Z\lfloor X) \le \mathcal{E}nt(Z\lfloor Y) + \mathcal{E}nt(Y\lfloor X)$$

and therefore

$$\mathrm{kd}(X \leftarrow XZ \to Z) \le \mathrm{kd}(X \leftarrow XY \to Y) + \mathrm{kd}(Y \leftarrow YZ \to Z)$$

⊠

Now we continue with the proof of Proposition 4.1.

Let **G** be an arbitrary complete reduction category. Suppose $\mathcal{X} = \{X_i; f_{ij}\}$, $\mathcal{Y} = \{Y_i; g_{ij}\}$ and $\mathcal{Z} = \{Z_i; h_{ij}\}$ are **G**-configurations, with initial spaces being X_0, Y_0 and Z_0, respectively. Let

$$\hat{\mathcal{F}} = (\mathcal{X} \leftarrow \mathcal{F} \to \mathcal{Y})$$
$$\hat{\mathcal{G}} = (\mathcal{Y} \leftarrow \mathcal{G} \to \mathcal{Z})$$

be two optimal minimal two-fans satisfying

$$\mathbf{k}(\mathcal{X},\mathcal{Y}) = \mathrm{kd}(\hat{\mathcal{F}})$$
$$\mathbf{k}(\mathcal{Y},\mathcal{Z}) = \mathrm{kd}(\hat{\mathcal{G}})$$

Recall that each two-fan of $\mathbf{G}$-configurations is a $\mathbf{G}$-configuration of two-fans between the individual spaces, that is

$$\mathcal{F} = \left\{\mathcal{F}_i = (X_i \leftarrow F_i \rightarrow Y_i)\right\}$$
$$\mathcal{G} = \left\{\mathcal{G}_i = (Y_i \leftarrow G_i \rightarrow Z_i)\right\}$$

We construct a coupling $\mathcal{H}$ between $\mathcal{X}$ and $\mathcal{Z}$ in the following manner. Starting with the two-tents configuration between the initial spaces, we use adhesion to extend it to a full configuration, thus constructing a coupling between X_0 and Z_0. This full configuration could then be "pushed down" and provides full extensions of two-tents on all lower levels. Thus we could "compose" couplings $\mathcal{F}$ and $\mathcal{G}$ and use a Shannon inequality to establish the triangle inequality for the Kolmogorov distance. Details are as follows.

Consider a two-tents configuration

$$X_0 \leftarrow F_0 \rightarrow Y_0 \leftarrow G_0 \rightarrow Z_0$$

and extend it by adhesion, as described in Section 2.5 to a $\mathbf{\Lambda}_3$-configuration

$$\begin{tikzcd}
 & A_0 \arrow[dl] \arrow[d] \arrow[dr] & \\
F_0 \arrow[d] \arrow[dr] & H_0 \arrow[dl] \arrow[dr] & G_0 \arrow[d] \arrow[dl] \\
X_0 & Y_0 & Z_0
\end{tikzcd}$$

Together with the reductions

$$(X_0)^{\mathbf{G}} \rightarrow \mathcal{X}$$
$$(Y_0)^{\mathbf{G}} \rightarrow \mathcal{Y}$$
$$(Z_0)^{\mathbf{G}} \rightarrow \mathcal{Z}$$

it gives rise to a $\mathbf{\Lambda}_3$-configuration of $\mathbf{G}$-configurations

$$\begin{tikzcd}
 & (A_0)^{\mathbf{G}} \arrow[dl] \arrow[d] \arrow[dr] & \\
(F_0)^{\mathbf{G}} \arrow[d] \arrow[dr] & (H_0)^{\mathbf{G}} \arrow[dl] \arrow[dr] & (G_0)^{\mathbf{G}} \arrow[d] \arrow[dl] \\
\mathcal{X} & \mathcal{Y} & \mathcal{Z}
\end{tikzcd} \tag{T.1}$$

Note that the minimal reductions of the two-fan subconfigurations of (T.1)

$$\mathcal{X} \leftarrow (F_0)^{\mathbf{G}} \rightarrow \mathcal{Y}$$
$$\mathcal{Y} \leftarrow (G_0)^{\mathbf{G}} \rightarrow \mathcal{Z}$$

are the two-fans $\hat{\mathcal{F}}$ and $\hat{\mathcal{G}}$, respectively, by Lemma 2.1.

Now consider the "minimization" of the above configuration

$$
\begin{tikzcd}
& \mathcal{A} \arrow[dl] \arrow[d] \arrow[dr] & \\
\mathcal{F} \arrow[d] \arrow[dr] & \mathcal{H} \arrow[dl] \arrow[dr] & \mathcal{G} \arrow[d] \arrow[dl] \\
\mathcal{X} & \mathcal{Y} & \mathcal{Z}
\end{tikzcd}
\tag{T.2}
$$

It could also be viewed as a $\mathbf{G}$-configuration of $\mathbf{\Lambda}_3$ configurations,

$$
\begin{tikzcd}
& A_i \arrow[dl] \arrow[d] \arrow[dr] & \\
F_i \arrow[d] \arrow[dr] & H_i \arrow[dl] \arrow[dr] & G_i \arrow[d] \arrow[dl] \\
X_i & Y_i & Z_i
\end{tikzcd}
$$

each of which is minimal by Corollary 2.2.

Now we can apply Lemma T.2 to each level to conclude that

$$
\begin{aligned}
\mathbf{k}(\mathcal{X}, \mathcal{Z}) &\leq \mathrm{kd}(\mathcal{X} \leftarrow \mathcal{H} \rightarrow \mathcal{Z}) \\
&\leq \mathrm{kd}(\mathcal{X} \leftarrow \mathcal{F} \rightarrow \mathcal{Y}) + \mathrm{kd}(\mathcal{X} \leftarrow \mathcal{G} \rightarrow \mathcal{Y}) \\
&= \mathbf{k}(\mathcal{X}, \mathcal{Y}) + \mathbf{k}(\mathcal{Y}, \mathcal{Z})
\end{aligned}
$$

Finally, if $k(\mathcal{X}, \mathcal{Y}) = 0$, then there is a two-fan $\mathcal{F}$ of $\mathbf{G}$-configurations between $\mathcal{X}$ and $\mathcal{Y}$ with $\mathrm{kd}(\mathcal{F}) = 0$, from which it follows that $\mathcal{X}$ and $\mathcal{Y}$ are isomorphic.⊠

Proposition 4.3. ↑ *Let* $\mathbf{G}$ *be a complete diagram category. Then with respect to the Kolmogorov distance on* $\mathbf{Prob}\langle\mathbf{G}\rangle$ *the tensor product*

$$
\otimes : (\mathbf{Prob}\langle\mathbf{G}\rangle, \mathbf{k})^2 \rightarrow (\mathbf{Prob}\langle\mathbf{G}\rangle, \mathbf{k})
$$

is 1-Lipschitz in each variable, that is, for every triple $\mathcal{X}, \mathcal{Y}, \mathcal{Y}' \in \mathbf{Prob}\langle\mathbf{G}\rangle$ *the following bound holds*

$$
\mathbf{k}(\mathcal{X} \otimes \mathcal{Y}, \mathcal{X} \otimes \mathcal{Y}') \leq \mathbf{k}(\mathcal{Y}, \mathcal{Y}')
$$

⊠

Proof: The claim follows easily from the additivity of entropy in equation (1.3). Suppose that $\mathcal{X} = \{X_i; f_{ij}\}$, $\mathcal{Y} = \{Y_i; g_{ij}\}$ and $\mathcal{Y}' = \{Y_i'; g_{ij}'\}$ are three $\mathbf{G}$-configurations and

$$
\mathcal{F} = (\mathcal{Y} \leftarrow \mathcal{Z} \rightarrow \mathcal{Y}')
$$

is an optimal fan, so that

$$
\mathbf{k}(\mathcal{Y}, \mathcal{Y}') = \sum_i \left[2\,\mathcal{E}nt(Z_i) - \mathcal{E}nt(Y_i) - \mathcal{E}nt(Y_i') \right]
$$

Consider the fan

$$
\mathcal{G} = (\mathcal{X} \otimes \mathcal{Y} \leftarrow \mathcal{X} \otimes \mathcal{Z} \rightarrow \mathcal{X} \otimes \mathcal{Y}')
$$

Then, by additivity of entropy, in equation (1.3), we have

$$\begin{aligned}\operatorname{kd}(\mathcal{F}) &= \sum_i \left[2\,\mathcal{E}nt(X_i \otimes Z_i) - \mathcal{E}nt(X_i \otimes Y_i) - \mathcal{E}nt(X_i \otimes Y_i')\right]\\ &= \sum_i \left[2\,\mathcal{E}nt(Z_i) - \mathcal{E}nt(Y_i) - \mathcal{E}nt(Y_i')\right]\\ &= \operatorname{kd}(\mathcal{G})\end{aligned}$$

and, therefore,

$$\mathbf{k}(\mathcal{X} \otimes \mathcal{Y}, \mathcal{X} \otimes \mathcal{Y}') \leq \operatorname{kd}(\mathcal{G}) = \operatorname{kd}(\mathcal{F}) = \mathbf{k}(\mathcal{Y}, \mathcal{Y}')$$

Thus, the tensor product of probability spaces is 1-Lipschitz with respect to each argument. ⊠

Proposition 4.7. ↑ *Suppose* $\mathbf{G}$ *is a complete diagram category and* $\boldsymbol{\delta} = \mathbf{k}, \boldsymbol{\kappa}$ *is either Kolmogorov distance or asymptotic Kolmogorov distance on* $\mathbf{Prob}\langle\mathbf{G}\rangle$*. Then the entropy function*

$$\mathcal{E}nt_* : (\mathbf{Prob}\langle\mathbf{G}\rangle, \boldsymbol{\delta}) \to (\mathbb{R}^{[\![\mathbf{G}]\!]}, |\cdot|_1), \quad \mathcal{X} = \{X_i, f_{ij}\} \mapsto (\mathcal{E}nt X_i)_i \in \mathbb{R}^{[\![\mathbf{G}]\!]}$$

is 1-Lipschitz. ⊠

Proof:
Let $\mathcal{X}, \mathcal{Y} \in \mathbf{Prob}\langle\mathbf{G}\rangle$ and let

$$\mathcal{G} = (\mathcal{X} \leftarrow \mathcal{Z} \to \mathcal{Y})$$

be an optimal fan with components

$$\mathcal{G}_i = (X_i \leftarrow Z_i \to Y_i)$$

For a fixed index i we can estimate the difference of entropies

$$\mathcal{E}nt(X_i) - \mathcal{E}nt(Y_i) = 2\Big(\mathcal{E}nt(X_i) - \mathcal{E}nt(Z_i)\Big) + \operatorname{kd}(\mathcal{G}_i) \leq \operatorname{kd}(\mathcal{G}_i)$$

By symmetry we then have

$$|\,\mathcal{E}nt(X_i) - \mathcal{E}nt(Y_i)| \leq \operatorname{kd}(\mathcal{G}_i)$$

Adding above inequalities for all i we have

$$|\,\mathcal{E}nt_*(\mathcal{X}) - \mathcal{E}nt_*(\mathcal{Y})|_1 \leq \operatorname{kd}(\mathcal{G}) = \mathbf{k}(\mathcal{X}, \mathcal{Y})$$

By the additivity of entropy we also obtain the 1-Lipschitz property of the entropy function with respect to the asymptotic Kolmogorov distance $\boldsymbol{\kappa}$. ⊠

Proposition 4.8. ↑ *Suppose* $R : \mathbf{G}' \to \mathbf{G}$ *is a functor between two complete diagram categories and* $\boldsymbol{\delta}$ *stands for either Kolmogorov or asymptotic Kolmogorov distance. Then the restriction operator*

$$R^* : (\mathbf{Prob}\langle\mathbf{G}\rangle, \boldsymbol{\delta}) \to (\mathbf{Prob}\langle\mathbf{G}'\rangle, \boldsymbol{\delta}), \quad \mathcal{X} \mapsto \mathcal{X} \circ R$$

is Lipschitz. ⊠

Proof: The claim follows from the functoriality of the restriction operator. We argue as follows.

Suppose that $R : \mathbf{G}' \to \mathbf{G}$ is a functor and R^* is the corresponding restriction operator. For $\mathcal{X}_1, \mathcal{X}_2 \in \mathbf{Prob}\langle\mathbf{G}\rangle$ let

$$\mathcal{F} = (\mathcal{X}_1 \leftarrow \mathcal{Y} \rightarrow \mathcal{X}_2)$$

be an optimal fan. Then

$$\mathcal{F}' := (R^*\mathcal{X}_1 \leftarrow R^*\mathcal{Y} \rightarrow R^*\mathcal{X}_2)$$

is a fan with the terminal vertices being the restrictions of $\mathcal{X}_1$ and $\mathcal{X}_2$. It can be considered as a $\mathbf{G}'$-configuration of two-fans over individual spaces in $R^*\mathcal{X}_1$ and $R^*\mathcal{X}_2$ each of which also appears as a fan in $\mathcal{F}$.

Thus, we obtain the rough estimate

$$\mathbf{k}(R^*\mathcal{X}_1, R^*\mathcal{X}_2) \leq [\![\mathbf{G}']\!] \cdot \mathbf{k}(\mathcal{X}_1, \mathcal{X}_2)$$

Since the restriction operator commutes with tensor powers, the same estimate also holds for the asymptotic Kolmogorov distance $\boldsymbol{\kappa}$. ⊠

Proposition 4.9. ↑ (Slicing Lemma) *Suppose* $\mathbf{G}$ *is a complete diagram category and we are given* $\mathcal{X}, \hat{\mathcal{X}}, \mathcal{Y}, \hat{\mathcal{Y}} \in \mathbf{Prob}\langle\mathbf{G}\rangle$ *– four* $\mathbf{G}$*-configurations and* $U, V, W \in \mathbf{Prob}$ *– probability spaces, that are included into the following three-tents configuration*

$$\begin{tikzcd}[column sep=small] & \hat{\mathcal{X}} \arrow[dl] \arrow[dr] & & W \arrow[dl] \arrow[dr] & & \hat{\mathcal{Y}} \arrow[dl] \arrow[dr] & \\ \mathcal{X} & & U & & V & & \mathcal{Y} \end{tikzcd}$$

such that the two-fan $(U \leftarrow W \rightarrow V)$ *is minimal. Then the following estimate holds*

$$\begin{aligned} \mathbf{k}(\mathcal{X}, \mathcal{Y}) \leq & \int_W \mathbf{k}(\mathcal{X}\lfloor u, \mathcal{Y}\lfloor v) \, \mathrm{d}\, p_W(u, v) \\ & + [\![\mathbf{G}]\!] \cdot \mathrm{kd}(U \leftarrow W \rightarrow V) \\ & + \sum_i \left[\mathcal{E}nt(U\lfloor X_i) + \mathcal{E}nt(V\lfloor Y_i) \right] \end{aligned}$$

⊠

Proof: Since the two-fan $(U \leftarrow W \rightarrow V)$ is minimal the probability space W could be considered having underlying set to be a subset of the Cartesian product of the underlying sets of U and V. For any pair $(u, v) \in \underline{W}$ with a positive weight consider an optimal two-fan

$$\mathcal{G}_{uv} = (\mathcal{X}\lfloor u \xleftarrow{\pi_{\mathcal{X}}} \mathcal{Z}_{uv} \xrightarrow{\pi_{Y}} \mathcal{Y}\lfloor v) \tag{T.3}$$

where $\mathcal{Z}_{uv} = \{Z_{uv,i}; \rho_{ij}\}$. Let $p_{uv,i}$ be the probability distributions on $Z_{uv,i}$ – the individual spaces in the configuration $\mathcal{Z}_{uv}$. The next step is to take a convex combination of distributions $p_{uv,i}$ weighted by p_W to construct a coupling $\mathcal{X} \leftarrow \mathcal{Z} \rightarrow \mathcal{Y}$.

First we extend the 7-vertex configuration to a full $\mathbf{\Lambda}_4$-configuration of $\mathbf{G}$-configurations, such that the top vertex has the distribution

$$p_i(x,y,u,v) \coloneqq p_{uv,i}(x,y) p_W(u,v) \tag{T.4}$$

as described in the Section 2.1.4.

If we integrate over y, we obtain

$$\sum_y p_i(x,u,v,y) = ((\pi_{\mathcal{X},i})_* p_{uv,i})(x) p_W(u,v)$$

Then we use that by (T.3) it holds that $(\pi_{\mathcal{X},i})_* p_{uv,i} = p_{X_i}(\cdot \lfloor u)$ and therefore

$$\sum_y p_i(x,y,u,v) = p_{X_i}(x \lfloor u) p_W(u,v).$$

In the same way,

$$\sum_x p_i(x,y,u,v) = p_{Y_i}(y \lfloor v) p_W(u,v).$$

Note that this exactly corresponds to adhesion, as described in Section 2.5. It follows that

$$\mathcal{X}\lfloor uv = \mathcal{X}\lfloor u \quad \text{and} \quad \mathcal{Y}\lfloor uv = \mathcal{Y}\lfloor v \tag{T.5}$$

and

$$\mathcal{E}nt(X_i\lfloor UV) = \mathcal{E}nt(X_i\lfloor U) \quad \text{and} \quad \mathcal{E}nt(Y_i\lfloor UV) = \mathcal{E}nt(Y_i\lfloor V) \tag{T.6}$$

The extended configuration contains a two-fan of configurations $\mathcal{F} = (\mathcal{X} \leftarrow \mathcal{Z} \to \mathcal{Y})$ with terminal vertices $\mathcal{X}$ and $\mathcal{Y}$. We call its initial vertex $\mathcal{Z} = \{XY_i, f_{ij}\}$.

The following estimates conclude the proof the the Slicing Lemma. First we use the definitions of intrinsic Kolmogorov distance $\mathbf{k}$ and of $\operatorname{kd}(\mathcal{F})$ to estimate

$$\begin{aligned}
\mathbf{k}(\mathcal{X},\mathcal{Y}) &\le \operatorname{kd}(\mathcal{F}) \\
&= \sum_i \operatorname{kd}(\mathcal{F}_i) \\
&= \sum_i \big[2\,\mathcal{E}nt(XY_i) - \mathcal{E}nt(X_i) - \mathcal{E}nt(Y_i)\big]
\end{aligned}$$

Next, we apply the definition of the conditional entropy to rewrite the right-hand side

$$\begin{aligned}
\mathbf{k}(\mathcal{X},\mathcal{Y}) \le \sum_i \big[&2\,\mathcal{E}nt(XY_i\lfloor UV) + 2\,\mathcal{E}nt(UV) - 2\,\mathcal{E}nt(UV\lfloor XY_i) \\
&- \mathcal{E}nt(X_i\lfloor U) - \mathcal{E}nt(U) + \mathcal{E}nt(U\lfloor X_i) \\
&- \mathcal{E}nt(Y_i\lfloor V) - \mathcal{E}nt(V) + \mathcal{E}nt(V\lfloor Y_i)\big]
\end{aligned}$$

We now use (T.6) and rearrange terms to obtain

$$\begin{aligned}
\mathbf{k}(\mathcal{X},\mathcal{Y}) \le \sum_i \big[&2\,\mathcal{E}nt(XY_i\lfloor UV) - \mathcal{E}nt(X_i\lfloor UV) - \mathcal{E}nt(Y_i\lfloor UV) \\
&+ 2\,\mathcal{E}nt(UV) - \mathcal{E}nt(U) - \mathcal{E}nt(V) \\
&- 2\,\mathcal{E}nt(UV\lfloor XY_i) + \mathcal{E}nt(U\lfloor X_i) + \mathcal{E}nt(V\lfloor Y_i)\big]
\end{aligned}$$

By the integral formula for conditional entropy (1.4) applied to the first three terms we get

$$\sum_i \big[2\,\mathcal{E}nt(XY_i\lfloor UV) - \mathcal{E}nt(X_i\lfloor UV) - \mathcal{E}nt(Y_i\lfloor UV)\big]$$
$$= \int_{UV} \mathbf{k}(\mathcal{X}\lfloor uv, \mathcal{Y}\lfloor uv)\,\mathrm{d}\,p_W(u,v)$$

However, because of (T.5) this simplifies to

$$\int_{UV} \mathbf{k}(\mathcal{X}\lfloor uv, \mathcal{Y}\lfloor uv)\,\mathrm{d}\,p_W(u,v) = \int_{UV} \mathbf{k}(\mathcal{X}\lfloor u, \mathcal{Y}\lfloor v)\,\mathrm{d}\,p_W(u,v)$$

Therefore,

$$\begin{aligned}\mathbf{k}(\mathcal{X},\mathcal{Y}) \le \int_{UV} \mathbf{k}(\mathcal{X}\lfloor u, \mathcal{Y}\lfloor v)\,\mathrm{d}\,p_W(u,v) + [\![\mathbf{G}]\!]\cdot \mathrm{kd}(U \leftarrow W \rightarrow V)\\ + \sum_i \big[\,\mathcal{E}nt(U\lfloor X_i) + \mathcal{E}nt(V\lfloor Y_i)\big]\end{aligned}$$

⊠

Proposition 4.11. ↑ *Let* $\mathcal{S} = \{S_i, f_{ij}\} \in \mathbf{Set}\langle\mathbf{G}\rangle$ *be a configuration of sets modeled on a complete diagram category* $\mathbf{G}$ *with the initial set* S_0*. Let* $p, q \in \Delta\mathcal{S}$ *be two probability distributions. Denote* $\mathcal{X} := (\mathcal{S}, p)$*,* $\mathcal{Y} := (\mathcal{S}, q)$ *and* $\alpha = \frac{1}{2}|p_0 - q_0|_1$*. Then*

$$\mathbf{k}(\mathcal{X},\mathcal{Y}) \le 2\cdot[\![\mathbf{G}]\!]\cdot\big(\alpha\cdot\ln|S_0| + \mathcal{E}nt(\Lambda_\alpha)\big)$$

⊠

Proof: We will need the following obvious rough estimate of the Kolmogorov distance that holds for any $p, q \in \Delta\mathcal{S}$:

$$2\cdot\mathbf{k}(\mathcal{X},\mathcal{Y}) \le 2[\![\mathbf{G}]\!]\cdot\ln|S_0| \tag{T.7}$$

It can be obtained by taking a tensor product for the coupling between $\mathcal{X}$ and $\mathcal{Y}$.

Our goal now is to write p and q as the convex combination of three other distributions $\hat{p}$, p^+ and q^+ as in

$$\begin{aligned}p &= (1-\alpha)\cdot\hat{p} + \alpha\cdot p^+\\ q &= (1-\alpha)\cdot\hat{p} + \alpha\cdot q^+\end{aligned}$$

with the smallest possible $\alpha \in [0,1]$.

We could do it the following way. Let $\alpha := \frac{1}{2}|p_0 - q_0|$. If $\alpha = 1$ then the proposition follows from the rough estimate (T.7), so from now on we assume that $\alpha < 1$. Define three probability distributions $\hat{p}_0$, p_0^+ and q_0^+ on S_0 by setting for every $x \in S_0$

$$\hat{p}_0(x) := \frac{1}{1-\alpha} \min\{p_0(x), q_0(x)\}$$
$$p_0^+ := \frac{1}{\alpha}\big(p_0 - (1-\alpha)\hat{p}_0\big)$$
$$q_0^+ := \frac{1}{\alpha}\big(q_0 - (1-\alpha)\hat{p}_0\big)$$

Denote by $\hat{p}, p^+, q^+ \in \Delta\mathcal{S}$ the distributions corresponding to $\hat{p}_0, p_0^+, q_0^+ \in \Delta S_0$ under isomorphism (3.5). Thus we have

$$p = (1-\alpha)\hat{p} + \alpha \cdot p^+$$
$$q = (1-\alpha)\hat{p} + \alpha \cdot q^+$$

Now we construct a "two-tents" configuration of $\mathbf{G}$-configurations

$$\mathcal{X} \leftarrow \tilde{\mathcal{X}} \rightarrow \Lambda_\alpha \leftarrow \tilde{\mathcal{Y}} \rightarrow \mathcal{Y} \tag{T.8}$$

by setting

$$\tilde{X}_i := \Big(S_i \times \underline{\Lambda}_\alpha;\ \tilde{p}_i(s,\square) = (1-\alpha)\hat{p}_i(s),\ \tilde{p}_i(s,\blacksquare) = \alpha \cdot p_i^+(s)\Big)$$
$$\tilde{Y}_i := \Big(S_i \times \underline{\Lambda}_\alpha;\ \tilde{q}_i(s,\square) = (1-\alpha)\hat{p}_i(s),\ \tilde{q}_i(s,\blacksquare) = \alpha \cdot q_i^+(s)\Big)$$

and

$$\tilde{\mathcal{X}} := \big\{\tilde{X}_i;\ f_{ij} \times \mathrm{Id}\big\}$$
$$\tilde{\mathcal{Y}} := \big\{\tilde{Y}_i;\ f_{ij} \times \mathrm{Id}\big\}$$

The reductions in the "two-tents" sub-configurations of (T.8) are given by coordinate projections. Note that the following isomorphisms hold

$$\mathcal{X}\lfloor\square \cong (\mathcal{S}, \hat{p})$$
$$\mathcal{X}\lfloor\blacksquare \cong (\mathcal{S}, p^+)$$
$$\mathcal{Y}\lfloor\square \cong (\mathcal{S}, \hat{p}) \cong \mathcal{X}\lfloor\square$$
$$\mathcal{Y}\lfloor\blacksquare \cong (\mathcal{S}, q^+)$$

Now we apply part (i) of Corollary 4.10 to obtain the desired inequality

$$\begin{aligned}\mathbf{k}(\mathcal{X},\mathcal{Y}) &\leq (1-\alpha)\,\mathbf{k}(\mathcal{X}\lfloor\square, \mathcal{Y}\lfloor\square) + \alpha \cdot \mathbf{k}(\mathcal{X}\lfloor\blacksquare, \mathcal{Y}\lfloor\blacksquare) \\ &\quad + \sum_i \Big[\,\mathcal{E}nt(\Lambda_\alpha\lfloor X_i) + \mathcal{E}nt(\Lambda_\alpha\lfloor Y_i)\Big] \\ &\leq 2 \cdot [\![\mathbf{G}]\!] \cdot \big(\alpha \cdot \ln|S_0| + \mathcal{E}nt(\Lambda_\alpha)\big)\end{aligned}$$

⊠

T.7. **Statements from the section "Extensions".**

Proposition 7.1. ↑ (Extension Lemma) *Let $k, l \in \mathbb{N}$, $k \le l$ and let $\mathcal{X}, \mathcal{X}' \in$ $\mathbf{Prob}\langle \mathbf{\Lambda}_k \rangle$ be minimal full configurations. For every $\mathcal{Y} \in \operatorname{Ext}_l \mathcal{X}$ there exists a $\mathcal{Y}' \in \operatorname{Ext}_l \mathcal{X}'$ such that*

$$\mathbf{k}(\mathcal{Y}', \mathcal{Y}) \le 2^{l-k}\, \mathbf{k}(\mathcal{X}', \mathcal{X})$$

⊠

Proof: Denote by X_0, X_0' the initial spaces in the configurations $\mathcal{X}$, $\mathcal{X}'$, respectively. Let Y_0 be the initial space of the full sub-configuration of $\mathcal{Y}$ generated by $Y_{k+1}, \dots, Y_l$. Let K_0 be the initial space in the optimal coupling between $\mathcal{X}$ and $\mathcal{X}'$

$$\mathcal{F} = (\mathcal{X}' \leftarrow \mathcal{K} \rightarrow \mathcal{X})$$

Recall that X_0 could be considered as the Cartesian product of the underlying sets of spaces generating $\mathcal{X}$ with some distribution on it. A similar view holds for X_0', Y_0 and K_0. Thus we have in particular $\underline{K}_0 = \underline{X}_0' \times \underline{X}_0$

Define a full minimal configuration $\mathcal{Z} \in \mathbf{Prob}\langle \mathbf{\Lambda}_{2k+l} \rangle$ by providing a distribution on

$$\underline{X}_0' \times \underline{X}_0 \times \underline{Y}_0 = \underline{K}_0 \times \underline{Y}_0$$

as explained in the Section 2.1.4. The distribution will be defined by

$$p(\mathbf{x}', \mathbf{x}, \mathbf{y}) \coloneqq p_{\mathcal{F}}(\mathbf{x}', \mathbf{x}) \cdot p_{\mathcal{Y}}(\mathbf{x}, \mathbf{y}) / p_{\mathcal{X}'}(\mathbf{x})$$

It is clear that $\mathcal{Z}$ contains both the coupling $\mathcal{F}$ and configuration $\mathcal{Y}$ as restrictions. It also contains the minimal full configuration

$$\mathcal{Y}' = \langle X_1', \dots, X_k', Y_{k+1}, \dots, Y_l \rangle$$

and a coupling $\mathcal{G}$ between $\mathcal{Y}$ and $\mathcal{Y}'$.

For a pair of spaces A and B in $\mathcal{Z}$ we denote by AB the initial space of a minimal fan in $\mathcal{Z}$ with the terminal spaces A and B. The two-fan of $\mathbf{\Lambda}_{k+l}$-configurations $\mathcal{G}$ can be considered as a $\mathbf{\Lambda}_{k+l}$-configuration of two-fans

$$\mathcal{G}_{IJ} \coloneqq (X_I Y_J \longleftarrow G_{IJ} \longrightarrow X_I' Y_J)$$

Using this notation we estimate for $I \subset \{1, \dots, k\}$ and $J \subset \{k+1, \dots, l\}$

$$\begin{aligned}
\mathbf{k}(\mathcal{Y}, \mathcal{Y}') &\le \operatorname{kd} \mathcal{G} = \sum_{I,J} \operatorname{kd}(\mathcal{G}_{IJ}) \\
&= \sum_{I,J} \Big[2\,\mathcal{E}nt(G_{IJ}) - \mathcal{E}nt(X_I Y_J) - \mathcal{E}nt(X_I' Y_J) \Big] \\
&\le \sum_{I,J} \Big[2\,\mathcal{E}nt(X_I X_I') - \mathcal{E}nt(X_I) - \mathcal{E}nt(X_I') + \\
&\qquad + \big(2\,\mathcal{E}nt(Y_J \lfloor X_I X_I') - \mathcal{E}nt(Y_J \lfloor X_I) - \mathcal{E}nt(Y_J \lfloor X_I') \big) \Big] \\
&\le 2^{k-l} \sum_I \operatorname{kd}(\mathcal{F}_I) \\
&\le 2^{k-l} \operatorname{kd}(\mathcal{F}) = 2^{k-l}\, \mathbf{k}(\mathcal{X}, \mathcal{X}')
\end{aligned}$$

⊠

Theorem 7.3. ↑ *Let $k \in \mathbb{N}$ and $\mathcal{X}, \mathcal{X}' \in \mathbf{Prob}\langle \mathbf{\Lambda}_k \rangle$. Then for all $l \in \mathbb{N}$, the Hausdorff distance between their stabilized relative entropic sets satisfies the Lipschitz estimate*

$$\mathrm{d}_H\left(\overline{\mathbf{\Gamma}}_l(\mathcal{X}), \overline{\mathbf{\Gamma}}_l(\mathcal{X}')\right) \leq 2^{l-k}\,\boldsymbol{\kappa}(\mathcal{X}, \mathcal{X}')$$

In other words, the map $\overline{\mathbf{\Gamma}}_l$ from minimal full configurations in $\mathbf{Prob}\langle \mathbf{\Lambda}_k \rangle$ to $\mathbf{K}_{k,l}$ is 2^{l-k}-Lipschitz. ⊠

Proof: Note that by Corollary 7.2, or more directly by Proposition 7.1, for $n \in \mathbb{N}$

$$\frac{1}{n}\,\mathrm{d}_H\left(\mathbf{\Gamma}_l^\circ\big((\mathcal{X})^{\otimes n}\big), \mathbf{\Gamma}_l^\circ\big((\mathcal{X}')^{\otimes n}\big)\right) \leq 2^{l-k}\frac{1}{n}\,\mathbf{k}\big((\mathcal{X})^{\otimes n}, (\mathcal{X}')^{\otimes n}\big)$$

Hence, by the scaling properties of the Hausdorff distance

$$\mathrm{d}_H\left(\frac{1}{n}\mathbf{\Gamma}_l^\circ\big((\mathcal{X})^{\otimes n}\big), \frac{1}{n}\mathbf{\Gamma}_l^\circ\big((\mathcal{X}')^{\otimes n}\big)\right) \leq 2^{l-k}\frac{1}{n}\,\mathbf{k}\big((\mathcal{X})^{\otimes n}, (\mathcal{X}')^{\otimes n}\big)$$

For convenience, we introduce the notation

$$K_n = \mathrm{Closure}\left(\frac{1}{n}\mathbf{\Gamma}_l^\circ\big((\mathcal{X})^{\otimes n}\big)\right) \qquad K = \mathbf{\Gamma}_l^\circ(\mathcal{X})$$
$$K_n' = \mathrm{Closure}\left(\frac{1}{n}\mathbf{\Gamma}_l^\circ\big((\mathcal{X}')^{\otimes n}\big)\right) \qquad K' = \overline{\mathbf{\Gamma}}_l(\mathcal{X}')$$

Recall that by definition,

$$K = \mathrm{Closure}\left(\bigcup_{n\in\mathbb{N}} K_n\right) \qquad K' = \mathrm{Closure}\left(\bigcup_{n\in\mathbb{N}} K_n'\right)$$

Note that by the superadditivity property of the unstabilized relative entropic sets (see inclusion (7.1)) the sequences $n \mapsto K_{n!}$ and $n \mapsto K'_{n!}$ are monotonically increasing sequences of sets, and

$$\bigcup_{i=1}^n K_i \subset K_{n!} \qquad \bigcup_{i=1}^n K_i' \subset K'_{n!}$$

Now select a large radius $R > 0$. Let $B_R(0)$ denote the ball of radius R around the origin in $\mathbb{R}^{2^{\{1,\dots,l\}}}$. By compactness and the definition of the stabilized relative entropic set

$$\mathrm{d}_H(K_{n!} \cap B_R(0), K \cap B_R(0)) \to 0$$
$$\mathrm{d}_H(K'_{n!} \cap B_R(0), K' \cap B_R(0)) \to 0$$

as $n \to \infty$. Therefore also

$$\mathrm{d}_H\left(K \cap B_R(0), K' \cap B_R(0)\right) \leq 2^{l-k}\,\boldsymbol{\kappa}\left(\mathcal{X}, \mathcal{X}'\right)$$

Because this inequality holds for every $R > 0$, the estimate in the lemma follows. ⊠

T.8. Statements from the section "Mixtures".

Lemma 8.1. ↑ *Let* $\mathbf{G}$ *be a complete diagram category and* $\mathcal{X}, \mathcal{Y} \in \mathbf{Prob}\langle\mathbf{G}\rangle$. *Then*

(i) $\boldsymbol{\kappa}(\mathcal{X}, \mathcal{X}^{\otimes n} \oplus_{\Lambda_{1/n}} \{\bullet\}) \leq \mathcal{E}nt(\Lambda_{1/n})$

(ii) $\boldsymbol{\kappa}\big(\mathcal{X}, (\mathcal{X} \oplus_{\Lambda_{1/n}} \{\bullet\})^{\otimes n}\big) \leq n \cdot \mathcal{E}nt(\Lambda_{1/n})$

(iii) $\boldsymbol{\kappa}\big((\mathcal{X} \otimes \mathcal{Y}) \oplus_{\Lambda_{1/n}} \{\bullet\}, (\mathcal{X} \oplus_{\Lambda_{1/n}} \{\bullet\}) \otimes (\mathcal{Y} \oplus_{\Lambda_{1/n}} \{\bullet\})\big) \leq 3\,\mathcal{E}nt(\Lambda_{1/n})$

(iv) $\boldsymbol{\kappa}\big((\mathcal{X} \oplus_{\Lambda_{1/n}} \{\bullet\}), (\mathcal{Y} \oplus_{\Lambda_{1/n}} \{\bullet\})\big) \leq \dfrac{1}{n}\boldsymbol{\kappa}(\mathcal{X}, \mathcal{Y})$

⊠

Proof: Recall that for the empirical reduction

$$\mathbf{q} : \Lambda_{1/n}^{\otimes N} \to \Delta\Lambda_{1/n}$$

the quantity $N \cdot \mathbf{q}(\lambda)(\blacksquare)$ counts the number of black squares in the sequence λ. It is a binomially distributed random variable with the mean N/n and variance $\frac{N}{n}(1 - \frac{1}{n})$.

The first claim is then proven by the following calculation

$$\begin{aligned}
&\boldsymbol{\kappa}(\mathcal{X}, \mathcal{X}^{\otimes n} \oplus_{\Lambda_{1/n}} \{\bullet\}) \\
&\quad= \lim_{N\to\infty} \frac{1}{N}\mathbf{k}\big(\mathcal{X}^{\otimes N}, (\mathcal{X}^{\otimes n} \oplus_{\Lambda_{1/n}} \{\bullet\})^{\otimes N}\big) \\
&\quad= \lim_{N\to\infty} \frac{1}{N}\mathbf{k}\left(\mathcal{X}^{\otimes N}, \bigoplus_{\lambda\in\Lambda_{1/n}^{\otimes N}} \mathcal{X}^{\otimes n\cdot N\cdot\mathbf{q}(\lambda)(\blacksquare)}\right) \\
&\quad\leq \mathcal{E}nt(\Lambda_{1/n}) + \lim_{N\to\infty} \frac{1}{N}\int_{\lambda\in\Lambda_{1/n}^{\otimes n}} \mathbf{k}(\mathcal{X}^{\otimes N}, \mathcal{X}^{\otimes(N\cdot n\cdot\mathbf{q}(\lambda)(\blacksquare))})\,\mathrm{d}\,p(\lambda) \\
&\quad\leq \mathcal{E}nt(\Lambda_{1/n}) + |\mathcal{E}nt_*(\mathcal{X})|_1 \cdot \lim_{N\to\infty} \frac{n}{N} \cdot \int_{\lambda\in\Lambda_{1/n}^{\otimes N}} \big|N/n - N\cdot\mathbf{q}(\lambda)(\blacksquare)\big|\,\mathrm{d}\,p(\lambda) \\
&\quad\leq \mathcal{E}nt(\Lambda_{1/n}) + |\mathcal{E}nt_*(\mathcal{X})|_1 \cdot \lim_{N\to\infty} \frac{n}{N} \cdot \sqrt{N\cdot\frac{1}{n}(1-\frac{1}{n})} \\
&\quad= \mathcal{E}nt(\Lambda_{1/n})
\end{aligned}$$

The second claim is proven similarly and the third follows from the second and the 1-Lipschitz property of the tensor product. Finally, the fourth follows from Corollary 4.10(iv), by slicing both arguments along $\Lambda_{1/n}$. ⊠

T.9. Statements from the section "Tropical Probability".

Lemma T.3. *Suppose the sequence* $\{a(i)\}_{i\in\mathbb{N}_0}$ *of real numbers is bounded from below and is quasi-subadditive, that is there is a constant* $C \in \mathbb{R}$ *such that for any* $i, j \in \mathbb{N}_0$ *holds*

$$a(i + j) \leq a(i) + a(j) + C$$

Then the limit

$$\lim_{i\to\infty} \frac{1}{i}a(i)$$

exists and is finite. ⊠

Proof: The lemma is standard and is sometimes refered to as Fekete's subadditive lemma. We include a proof for the convenience of the reader. Assume first that $C = 0$. Then the sequence satisfies $a(k \cdot i) \leq k \cdot a(i)$ and in particular $a(i) \leq i \cdot a(1)$. Let $l := \liminf \frac{1}{i} a(i) \in [0, \infty)$. Choose $\varepsilon > 0$. Then we can find $k \in \mathbb{N}$ such that $\frac{1}{k} a(k) \leq l + \varepsilon$. For $n \in \mathbb{N}$ let q, r be the quotient and the reminder of the integer division of n by k, that is

$$n = q \cdot k + r, \quad 0 \leq r < k$$

Then

$$\frac{1}{n} a(n) \leq \frac{1}{n} (q \cdot a(k) + a(r)) \leq \frac{1}{q \cdot k + r} (q \cdot a(k)) + \frac{1}{n} a(r) \leq l + \varepsilon + \varepsilon = l + 2\varepsilon$$

The last inequality holds once n is sufficiently large, specifically when

$$n \geq \frac{1}{\varepsilon} \max_{0 \leq i \leq k} a(i)$$

Therefore

$$\lim_{i \to \infty} \frac{1}{i} a(i) = l$$

Now if $C > 0$ then the sequence $b(i) := a(i) + C$ is subadditive and $\frac{1}{i} b(i)$ converges by the previous argument. Thus we have

$$\lim_{i \to \infty} \frac{1}{i} b(i) = \lim_{i \to \infty} \frac{1}{i} (a(i) + C) = \lim_{i \to \infty} \frac{1}{i} a(i)$$

⊠

Lemma 9.1. ↑ *For a pair $\overline{\gamma}_1, \overline{\gamma}_2 \in \mathsf{QL}_{\boldsymbol{\delta}}(\Gamma)$ the limit*

$$\lim_{n \to \infty} \frac{1}{n} \boldsymbol{\delta} \big(\gamma_1(n), \gamma_2(n) \big)$$

exists and is finite. ⊠

Proof: Suppose $\overline{\gamma}_1$ and $\overline{\gamma}_2$ are two quasi-linear sequences of elements of Γ, then for any $i, j \in \mathbb{N}_0$

$$\begin{aligned}
\boldsymbol{\delta} \big(\gamma_1(i+j), &\gamma_2(i+j) \big) \\
&\leq \boldsymbol{\delta} \big(\gamma_1(i+j), \gamma_1(i) \otimes \gamma_1(j) \big) + \boldsymbol{\delta} \big(\gamma_2(i+j), \gamma_2(i) \otimes \gamma_2(j) \big) \\
&\quad + \boldsymbol{\delta} \big(\gamma_1(i) \otimes \gamma_1(j), \gamma_2(i) \otimes \gamma_2(j) \big) \\
&\leq \mathrm{Defect}_{\boldsymbol{\delta}}(\overline{\gamma}_1) + \mathrm{Defect}_{\boldsymbol{\delta}}(\overline{\gamma}_2) + \boldsymbol{\delta} \big(\gamma_1(i), \gamma_2(i) \big) + \boldsymbol{\delta} \big(\gamma_1(j), \gamma_2(j) \big)
\end{aligned}$$

Thus the sequence $\boldsymbol{\delta}(\gamma_1(i), \gamma_2(i))$ is quasi-subadditive and by Lemma T.3 the limit

$$\lim_{i \to \infty} \frac{1}{i} \boldsymbol{\delta} \big(\gamma_1(i), \gamma_2(i) \big)$$

exists and is finite. ⊠

Proposition 9.2. ↑ *Suppose $(\Gamma, \otimes, \boldsymbol{\delta})$ is a metric Abelian monoid such that*

(i) *the distance function* $\boldsymbol{\delta}$ *is homogeneous, that is for any* $\gamma_1, \gamma_2 \in \Gamma$ *and* $n \in \mathbb{N}_0$

$$\boldsymbol{\delta}(\gamma_1^{\otimes n}, \gamma_2^{\otimes n}) = n \cdot \boldsymbol{\delta}(\gamma_1, \gamma_2)$$

(ii) $(\Gamma, \otimes, \boldsymbol{\delta})$ *has the uniformly bounded defect property.*

Then the space $(\mathsf{QL}_{\boldsymbol{\delta}}(\Gamma), \hat{\boldsymbol{\delta}})$ *is complete.* ⊠

Proof: Given a Cauchy sequence $\{\overline{\gamma}_i\}$ of elements in $(\mathsf{QL}_{\boldsymbol{\delta}}(\Gamma), \hat{\boldsymbol{\delta}})$ we need to find a limiting element $\overline{\varphi} \in \mathsf{QL}_{\boldsymbol{\delta}}(\Gamma)$. We will do that by a version of the diagonal process, that is we define $\varphi(n)$ to have value $\gamma_i(n)$ for i sufficiently large depending on n. The quasi-linearity of $\overline{\varphi}$ would follow from the fact that for a fixed n and all sufficiently large i the set $\{\gamma_i(n)\}$ is uniformly bounded.

Now we give the detailed argument. First we replace each element of the sequence $\{\overline{\gamma}_i\}$ by an asymptotically equivalent element with defect bounded by the constant C according to assumption (ii) of the lemma. We will still call the new sequence $\{\overline{\gamma}_i\}$. The Cauchy sequence $\{\overline{\gamma}_i\}$ satisfies

$$\sup_{i,j \geq \mathbf{i}} \hat{\boldsymbol{\delta}}(\overline{\gamma}_i, \overline{\gamma}_j) \to 0 \quad \text{as} \quad \mathbf{i} \to \infty$$

By assumption (ii) of the lemma for any $n, k \in \mathbb{N}_0$ holds

$$\begin{aligned} k \cdot \boldsymbol{\delta}\left(\gamma_i(n), \gamma_j(n)\right) &= \boldsymbol{\delta}\left(\gamma_i(n)^{\otimes k}, \gamma_j(n)^{\otimes k}\right) \\ &\leq \boldsymbol{\delta}\left(\gamma_i(kn), \gamma_j(kn)\right) + 2k \cdot C \end{aligned}$$

Dividing by k we obtain

$$\boldsymbol{\delta}(\gamma_i(n), \gamma_j(n)) \leq \frac{1}{k} \boldsymbol{\delta}(\gamma_i(kn), \gamma_j(kn)) + 2C$$

Now we pass to the limit sending k to infinity, while keeping n fixed:

$$\boldsymbol{\delta}(\gamma_i(n), \gamma_j(n)) \leq n \cdot \hat{\boldsymbol{\delta}}(\overline{\gamma}_i, \overline{\gamma}_j) + 2C$$

Given n let $\mathbf{i}(n)$ be a number such that for any $i, j \geq \mathbf{i}(n)$ holds

$$\hat{\boldsymbol{\delta}}(\overline{\gamma}_i, \overline{\gamma}_j) \leq \frac{1}{n}$$

We may assume that $\mathbf{i}(n)$ is nondecreasing as a function of n. Then for any $i, j, n \in \mathbb{N}$ with $i, j \geq \mathbf{i}(n)$ we have the following bound

$$\boldsymbol{\delta}\left(\gamma_i(n), \gamma_j(n)\right) \leq 2C + 1 \tag{T.9}$$

Now we are ready to define the limiting sequence $\overline{\varphi}$ by setting

$$\varphi(n) := \gamma_{\mathbf{i}(n)}(n)$$

First we verify that $\overline{\varphi}$ is quasi-linear

$$\begin{aligned}
\boldsymbol{\delta}\big(\varphi(n+m),\varphi(n)\otimes\varphi(m)\big) &= \boldsymbol{\delta}\left(\gamma_{\mathbf{i}(n+m)}(n+m),\gamma_{\mathbf{i}(n)}(n)\otimes\gamma_{\mathbf{i}(m)}(m)\right)\\
&\le \boldsymbol{\delta}\left(\gamma_{\mathbf{i}(n+m)}(n+m),\gamma_{\mathbf{i}(n+m)}(n)\otimes\gamma_{\mathbf{i}(n+m)}(m)\right)\\
&\quad+\boldsymbol{\delta}\left(\gamma_{\mathbf{i}(n+m)}(n)\otimes\gamma_{\mathbf{i}(n+m)}(m),\gamma_{\mathbf{i}(n)}(n)\otimes\gamma_{\mathbf{i}(m)}(m)\right)\\
&\le C+\boldsymbol{\delta}\left(\gamma_{\mathbf{i}(n+m)}(n),\gamma_{\mathbf{i}(n)}(n)\right)+\boldsymbol{\delta}\left(\gamma_{\mathbf{i}(n+m)}(m),\gamma_{\mathbf{i}(m)}(m)\right)\\
&\le C+2(2C+1)=5C+2=:C'
\end{aligned}$$

The convergence of $\overline{\gamma}_i$ to $\overline{\varphi}$ is shown as follows. For $n,k\in\mathbb{N}$ let $q,r\in\mathbb{N}_0$ be the quotient and the remainder of the division of n by k, that is $n=q\cdot k+r$ and $0\le r<k$. Fix $k\in\mathbb{N}$ and let $i\ge\mathbf{i}(k)$, then

$$\begin{aligned}
\hat{\boldsymbol{\delta}}(\overline{\gamma}_i,\overline{\varphi}) &= \lim_{n\to\infty}\frac{1}{n}\boldsymbol{\delta}\left(\gamma_i(n),\varphi(n)\right)\\
&= \lim_{n\to\infty}\frac{1}{n}\boldsymbol{\delta}\left(\gamma_i(q\cdot k+r),\gamma_{\mathbf{i}(n)}(q\cdot k+r)\right)\\
&\le \lim_{n\to\infty}\frac{1}{n}\left(q\cdot\boldsymbol{\delta}\left(\gamma_i(k),\gamma_{\mathbf{i}(n)}(k)\right)+\boldsymbol{\delta}\left(\gamma_i(r),\gamma_{\mathbf{i}(n)}(r)\right)+2qC'+2C'\right)\\
&\le \lim_{n\to\infty}\frac{1}{n}\big((3q+3)\cdot C'\big)\\
&= \frac{3C'}{k}
\end{aligned}$$

Since $k\in\mathbb{N}$ is arbitrary we have

$$\lim_{i\to\infty}\hat{\boldsymbol{\delta}}(\overline{\gamma}_i,\overline{\varphi})=0$$

⊠

Proposition 9.3. ↑ *Suppose* $(\Gamma,\otimes,\boldsymbol{\delta})$ *has the* ε*-uniformly bounded defect property for every* $\varepsilon>0$*. Then* $\mathsf{L}_{\boldsymbol{\delta}}(\Gamma)$ *is dense in* $\mathsf{QL}_{\boldsymbol{\delta}}(\Gamma)$ ⊠

Proof: Let $\overline{\gamma}=\{\gamma(n)\}$ be a $\boldsymbol{\delta}$-quasi-linear sequence. We need to approximate it with linear sequences. For $i\in\mathbb{N}$, let $\overline{\gamma}_i$ be a sequence asymptotically equivalent to γ and satisfying

$$\operatorname{Defect}_{\boldsymbol{\delta}}\overline{\gamma}_i\le 1/i$$

as provided by the $1/i$-uniformly bounded defect property.

Define a $\boldsymbol{\delta}$-linear sequence $\overline{\eta}_i$ by

$$\eta_i(n):=\gamma_i(1)^{\otimes n}$$

Then

$$\begin{aligned}\hat{\boldsymbol{\delta}}(\overline{\gamma},\overline{\eta}_i) &= \hat{\boldsymbol{\delta}}(\overline{\gamma}_i,\overline{\eta}_i)\\ &= \lim_{n\to\infty}\frac{1}{n}\,\boldsymbol{\delta}(\gamma_i(n),\eta_i(n))\\ &= \lim_{n\to\infty}\frac{1}{n}\,\boldsymbol{\delta}(\gamma_i(n),\gamma_i(1)^{\otimes n})\\ &\leq \lim_{n\to\infty}\frac{1}{n}\cdot n\cdot \mathrm{Defect}_{\boldsymbol{\delta}}(\overline{\gamma}_i)\\ &\leq \frac{1}{i}\end{aligned}$$

Thus $\lim \overline{\eta}_i = \gamma$. ⊠

Lemma 9.4. ↑ *The natural inclusion*

$$\imath_2 : (\mathsf{QL}_{\boldsymbol{\delta}}(\Gamma),\hat{\boldsymbol{\delta}}) \hookrightarrow (\mathsf{QL}_{\hat{\boldsymbol{\delta}}}(\Gamma),\hat{\hat{\boldsymbol{\delta}}})$$

is an isometric embedding. ⊠

Proof: Let $\overline{\gamma}_1,\overline{\gamma}_2 \in \mathsf{QL}_{\boldsymbol{\delta}}(\Gamma)$ be two sequences of $\boldsymbol{\delta}$-quasi-linear sequences. We have to show that the two numbers

$$\hat{\boldsymbol{\delta}}(\overline{\gamma}_1,\overline{\gamma}_2) = \lim_{n\to\infty}\frac{1}{n}\,\boldsymbol{\delta}\left(\gamma_1(n),\gamma_2(n)\right)$$

and

$$\hat{\hat{\boldsymbol{\delta}}}(\overline{\gamma}_1,\overline{\gamma}_2) = \lim_{n\to\infty}\frac{1}{n}\,\hat{\boldsymbol{\delta}}\left(\gamma_1(n),\gamma_2(n)\right)$$

are equal. Since shifts are non-expanding maps, we have $\hat{\boldsymbol{\delta}} \leq \boldsymbol{\delta}$ and it follows immediately that

$$\hat{\hat{\boldsymbol{\delta}}}(\overline{\gamma}_1,\overline{\gamma}_2) \leq \hat{\boldsymbol{\delta}}(\overline{\gamma}_1,\overline{\gamma}_2)$$

and we are left to show the opposite inequality. We will do it as follows. Fix $n > 0$, then

$$\begin{aligned}\hat{\boldsymbol{\delta}}(\overline{\gamma}_1,\overline{\gamma}_2) &= \lim_{k\to\infty}\frac{1}{kn}\,\boldsymbol{\delta}\left(\gamma_1(kn),\gamma_2(kn)\right)\\ &\leq \lim_{k\to\infty}\frac{1}{kn}\left(\boldsymbol{\delta}\left(\gamma_1(n)^{\otimes k},\gamma_2(n)^{\otimes k}\right) + k\cdot\left(\mathrm{Defect}_{\boldsymbol{\delta}}(\overline{\gamma}_1)+\mathrm{Defect}_{\boldsymbol{\delta}}(\overline{\gamma}_2)\right)\right)\\ &\leq \frac{1}{n}\,\hat{\boldsymbol{\delta}}\left(\gamma_1(n),\gamma_2(n)\right) + \frac{1}{n}\left(\mathrm{Defect}_{\boldsymbol{\delta}}(\overline{\gamma}_1)+\mathrm{Defect}_{\boldsymbol{\delta}}(\overline{\gamma}_2)\right)\end{aligned}$$

Passing to the limit with respect to n gives required inequality

$$\hat{\boldsymbol{\delta}}(\overline{\gamma}_1,\overline{\gamma}_2) \leq \hat{\hat{\boldsymbol{\delta}}}(\overline{\gamma}_1,\overline{\gamma}_2)$$

⊠

Lemma 9.5. ↑ *The image of the isometric embedding*

$$\imath_2 : (\mathsf{QL}_{\boldsymbol{\delta}}(\Gamma),\hat{\boldsymbol{\delta}}) \hookrightarrow (\mathsf{QL}_{\hat{\boldsymbol{\delta}}}(\Gamma),\hat{\hat{\boldsymbol{\delta}}})$$

is dense in $(\mathsf{QL}_{\hat{\boldsymbol{\delta}}}(\Gamma),\hat{\hat{\boldsymbol{\delta}}})$ ⊠

Proof: Given an element $\overline{\gamma} = \{\gamma(n)\}$ in $\mathsf{QL}_{\hat{\boldsymbol{\delta}}}(\Gamma)$ we have to find a $\hat{\hat{\boldsymbol{\delta}}}$-approximating sequence $\overline{\gamma}_i = \{\gamma_i(n)\}$ in $\mathsf{QL}_{\boldsymbol{\delta}}(\Gamma)$. Define

$$\gamma_i(n) := \gamma(i)^{\otimes\lfloor\frac{n}{i}\rfloor}$$

We have to show that each $\overline{\gamma}_i$ is $\boldsymbol{\delta}$-quasi-linear and that $\hat{\hat{\boldsymbol{\delta}}}(\overline{\gamma}_i,\overline{\gamma}) \xrightarrow{i\to\infty} 0$. These follow from

$$\begin{aligned}\hat{\boldsymbol{\delta}}\left(\gamma_i(m+n),\gamma_i(m)\otimes\gamma_i(n)\right) &= \hat{\boldsymbol{\delta}}\left(\gamma(i)^{\otimes\lfloor\frac{m+n}{i}\rfloor},\gamma(i)^{\otimes\lfloor\frac{m}{i}\rfloor}\otimes\gamma(i)^{\otimes\lfloor\frac{n}{i}\rfloor}\right)\\ &\leq \hat{\boldsymbol{\delta}}\left(\gamma(i),\mathbf{1}\right)\end{aligned}$$

and

$$\begin{aligned}\hat{\hat{\boldsymbol{\delta}}}(\overline{\gamma}_i,\overline{\gamma}) &= \lim_{n\to\infty}\frac{1}{n}\,\hat{\boldsymbol{\delta}}\left(\gamma_i(n),\gamma(n)\right)\\ &= \lim_{n\to\infty}\frac{1}{n}\,\hat{\boldsymbol{\delta}}\left(\gamma(i)^{\otimes\lfloor\frac{n}{i}\rfloor},\gamma(n)\right)\\ &\leq \lim_{n\to\infty}\left[\frac{1}{n}\,\hat{\boldsymbol{\delta}}\left(\gamma\left(i\lfloor\tfrac{n}{i}\rfloor\right),\gamma(n)\right)+\frac{1}{n}\lfloor\tfrac{n}{i}\rfloor\operatorname{Defect}_{\hat{\boldsymbol{\delta}}}(\overline{\gamma})\right]\\ &\leq \lim_{n\to\infty}\left[\frac{1}{n}\max_{k=0,\dots,i-1}\hat{\boldsymbol{\delta}}\left(\mathbf{1},\gamma(k)\right)+\frac{i}{n}\operatorname{Defect}_{\hat{\boldsymbol{\delta}}}(\overline{\gamma})\right]+\frac{1}{i}\operatorname{Defect}_{\hat{\boldsymbol{\delta}}}(\overline{\gamma})\\ &\leq \frac{1}{i}\operatorname{Defect}_{\hat{\boldsymbol{\delta}}}(\overline{\gamma})\end{aligned}$$

It is worth noting, that the defect of $\overline{\gamma}_i$ need not to be uniformly bounded with respect to i. ⊠

Lemma 9.6. ↑ *For a complete diagram category, and for every $\varepsilon > 0$, the space $(\mathbf{Prob}\langle\mathbf{G}\rangle,\otimes,\boldsymbol{\kappa})$ has the ε-uniformly bounded defect property, that is for any $\boldsymbol{\kappa}$-quasi-linear sequence $\overline{\mathcal{X}}\in\mathsf{QL}_{\boldsymbol{\kappa}}(\mathbf{Prob}\langle\mathbf{G}\rangle)$ there exists an asymptotically equivalent sequence $\overline{\mathcal{Y}}$ with defect not exceeding ε.* ⊠

Proof: Let $\overline{\mathcal{X}} = \{\mathcal{X}(i)\}$ be a quasi-linear sequence and let $\varepsilon > 0$. We will find an asymptotically equivalent sequence with defect less than ε.

Define a new sequence $\overline{\mathcal{Y}} = \{\mathcal{Y}(i)\}$ by

$$\mathcal{Y}(i) := \left[\mathcal{X}(k\cdot i)\right]\oplus_{\Lambda_{1/k}}\{\bullet\}$$

where the number $k\in\mathbb{N}$ will be chosen later. First we verify that the sequences $\overline{\mathcal{X}}$ and $\overline{\mathcal{Y}}$ are asymptotically equivalent, that is

$$\hat{\boldsymbol{\kappa}}(\overline{\mathcal{X}},\overline{\mathcal{Y}}) := \lim_{i\to\infty}\frac{1}{i}\,\boldsymbol{\kappa}\left(\mathcal{X}(i),\mathcal{Y}(i)\right) = 0$$

We estimate the asymptotic distance between individual members of sequences $\overline{\mathcal{X}}$ and $\overline{\mathcal{Y}}$ using Lemma 8.1 as follows

$$\begin{aligned}\boldsymbol{\kappa}(\mathcal{X}(i),\mathcal{Y}(i)) &= \boldsymbol{\kappa}\left(\mathcal{X}(i),\mathcal{X}(k\cdot i)\oplus_{\Lambda_{1/k}}\{\bullet\}\right)\\ &\leq \boldsymbol{\kappa}\left(\mathcal{X}(i),\mathcal{X}(i)^{\otimes k}\oplus_{\Lambda_{1/k}}\{\bullet\}\right)+\boldsymbol{\kappa}\left(\mathcal{X}(i)^{\otimes k}\oplus_{\Lambda_{1/k}}\{\bullet\},\mathcal{X}(k\cdot i)\oplus_{\Lambda_{1/k}}\{\bullet\}\right)\\ &\leq \mathcal{E}nt(\Lambda_{1/k})+\operatorname{Defect}_{\boldsymbol{\kappa}}(\overline{\mathcal{X}})\end{aligned}$$

Thus $\hat{\boldsymbol{\kappa}}(\overline{\mathcal{X}},\overline{\mathcal{Y}})=0$ and the two sequences are asymptotically equivalent.

Next we show that the sequence $\overline{\mathcal{Y}}$ is $\boldsymbol{\kappa}$-quasi-linear and evaluate its defect using Lemma 8.1. Let $i,j\in\mathbb{N}$, then

$$\begin{aligned}
&\boldsymbol{\kappa}\big(\mathcal{Y}(i+j),\mathcal{Y}(i)\otimes\mathcal{Y}(j)\big)\\
&\quad=\boldsymbol{\kappa}\Big(\mathcal{X}(k\cdot i+k\cdot j)\oplus_{\Lambda_{1/k}}\{\bullet\},\big[\mathcal{X}(k\cdot i)\oplus_{\Lambda_{1/k}}\{\bullet\}\big]\otimes\big[\mathcal{X}(k\cdot j)\oplus_{\Lambda_{1/k}}\{\bullet\}\big]\Big)\\
&\quad\leq\boldsymbol{\kappa}\Big(\big[\mathcal{X}(k\cdot i)\otimes\mathcal{X}(k\cdot j)\big]\oplus_{\Lambda_{1/k}}\{\bullet\},\big[\mathcal{X}(k\cdot i)\oplus_{\Lambda_{1/k}}\{\bullet\}\big]\otimes\big[\mathcal{X}(k\cdot j)\oplus_{\Lambda_{1/k}}\{\bullet\}\big]\Big)\\
&\qquad+\frac{1}{k}\operatorname{Defect}(\overline{\mathcal{X}})\\
&\quad\leq 3\,\mathcal{E}nt(\Lambda_{1/k})+\frac{1}{k}\operatorname{Defect}_{\boldsymbol{\kappa}}(\overline{\mathcal{X}})
\end{aligned}$$

Thus, by choosing k to be a solution to the inequality

$$3\,\mathcal{E}nt(\Lambda_{1/k})+\frac{1}{k}\operatorname{Defect}_{\boldsymbol{\kappa}}(\overline{\mathcal{X}})\leq\varepsilon$$

we can make sure that

$$\operatorname{Defect}_{\boldsymbol{\kappa}}(\overline{\mathcal{Y}})\leq\varepsilon$$

⊠

Lemma 9.12. ↑ *Denote by U_n a finite uniform probability space of cardinality n, then*

(i)
$$\mathbf{k}(U_n,U_m)\leq 2\ln 2+\left|\ln\frac{n}{m}\right|$$

(ii)
$$\boldsymbol{\kappa}(U_n,U_m)=|\,\mathcal{E}nt(U_n)-\mathcal{E}nt(U_m)|$$

⊠

Proof: Consider a two-fan $U_n\xleftarrow{f}U_{nm}\xrightarrow{g}U_m$. To construct specific reductions f and g we identify U_{nm}, U_n and U_m with the cyclic groups of the corresponding order

$$\begin{aligned}
U_{nm}&\leftrightarrow\mathbb{Z}_{nm}\\
U_n&\leftrightarrow\mathbb{Z}_n\\
U_m&\leftrightarrow\mathbb{Z}_m
\end{aligned}$$

Consider the short exact sequences

$$\{0\}\longrightarrow\mathbb{Z}_n\xrightarrow{\times m}\mathbb{Z}_{nm}\xrightarrow{\bmod m}\mathbb{Z}_m\longrightarrow\{0\}$$
$$\{0\}\longrightarrow\mathbb{Z}_m\xrightarrow{\times n}\mathbb{Z}_{nm}\xrightarrow{\bmod n}\mathbb{Z}_n\longrightarrow\{0\}$$

Choose for f the left splitting in the first exact sequence, and for g the left splitting in the second exact sequence.

Now that we constructed a two-fan $U_n \xleftarrow{f} U_{nm} \xrightarrow{g} U_m$, let $U_n \leftarrow Z \rightarrow U_m$ be its minimal reduction. Now we estimate $|Z| \le n+m$, which implies that

$$\begin{aligned}
\mathbf{k}(U_n, U_m) &\le 2\,\mathcal{E}nt(Z) - \mathcal{E}nt(U_n) - \mathcal{E}nt(U_m) \\
&\le 2\ln(n+m) - \ln n - \ln m \\
&\le 2\ln 2 + 2\ln \max\{n, m\} - \ln n - \ln m \\
&= 2\ln 2 + \left|\ln \frac{n}{m}\right|
\end{aligned}$$

To prove the second assertion note that entropy is a $\mathbf{k}$-1-Lipschitz function. Therefore we have

$$|\,\mathcal{E}nt(U_n) - \mathcal{E}nt(U_m)| \le \mathbf{k}(U_n, U_m) \le |\,\mathcal{E}nt(U_n) - \mathcal{E}nt(U_m)| + 2\ln 2$$

Substituting in the definition of asymptotic Kolmogorov distance we obtain the required equality. ⊠

References

[ABD+08] Nihat Ay, Nils Bertschinger, Ralf Der, Frank Güttler, and Eckehard Olbrich. Predictive information and explorative behavior of autonomous robots. *The European Physical Journal B*, 63(3):329–339, 2008.

[ABK+15] Samson Abramsky, Rui Soares Barbosa, Kohei Kishida, Raymond Lal, and Shane Mansfield. Contextuality, cohomology and paradox. *arXiv preprint arXiv:1502.03097*, 2015.

[BBI01] Dmitri Burago, Yuri Burago, and Sergei Ivanov. *A course in metric geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001.

[BFL11] John C. Baez, Tobias Fritz, and Tom Leinster. A characterization of entropy in terms of information loss. *Entropy*, 13(11):1945–1957, 2011.

[BRO+14] Nils Bertschinger, Johannes Rauh, Eckehard Olbrich, Jürgen Jost, and Nihat Ay. Quantifying unique information. *Entropy*, 16(4):2161–2183, 2014.

[Csi98] Imre Csiszár. The method of types. *IEEE Trans. Inform. Theory*, 44(6):2505–2523, 1998. Information theory: 1948–1998.

[CT91] Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. Wiley Series in Telecommunications. John Wiley & Sons, Inc., New York, 1991. A Wiley-Interscience Publication.

[DFZ11] Randall Dougherty, Chris Freiling, and Kenneth Zeger. Non-shannon information inequalities in four random variables. *arXiv preprint arXiv:1104.3602*, 2011.

[Fri09] Karl Friston. The free-energy principle: a rough guide to the brain? *Trends in cognitive sciences*, 13(7):293–301, 2009.

[Gro12] Misha Gromov. In a search for a structure, part 1: On entropy. Preprint available at http://www.ihes.fr/gromov, 2012.

[Mat07] Frantisek Matus. Infinitely many information inequalities. In *Information Theory, 2007. ISIT 2007. IEEE International Symposium on*, pages 41–44. IEEE, 2007.

[SA15] Bastian Steudel and Nihat Ay. Information-theoretic inference of common ancestors. *Entropy*, 17(4):2304–2327, 2015.

[Sin76] Ya. G. Sinai. *Introduction to ergodic theory*. Princeton University Press, Princeton, N.J., 1976. Translated by V. Scheffer, Mathematical Notes, 18.

[VDP13] Sander G Van Dijk and Daniel Polani. Informational constraints-driven organization in goal-directed behavior. *Advances in Complex Systems*, 16(02n03):1350016, 2013.

[Yeu12] Raymond W Yeung. *A first course in information theory*. Springer Science & Business Media, 2012.