

Conditioning in Tropical Probability Theory

Rostislav Matveev^{1,*}  and Jacobus W. Portegies^{2,*} ¹ Max Planck Institute for Mathematics in the Sciences, 04103 Leipzig, Germany² Department of Mathematics and Computer Science, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands

* Correspondence: matveev@mis.mpg.de (R.M.); j.w.portegies@tue.nl (J.W.P.)

Abstract: We define a natural operation of conditioning of tropical diagrams of probability spaces and show that it is Lipschitz continuous with respect to the asymptotic entropy distance.

Keywords: tropical probability

1. Introduction

In [1,2], we have initiated the study of tropical probability spaces and their diagrams. In [1], we endowed (commutative) diagrams of probability spaces with the intrinsic entropy distance and, in [2], we defined tropical diagrams as points in the asymptotic cone of the metric space. They are represented by certain sequences of diagrams of probability spaces.

We expect that tropical diagrams will be helpful in the study of information optimization problems, such as the ones considered in [3–8], and we have indeed applied them to derive a dimension-reduction result for the shape of the entropic cone in [9].

In this present article, we introduce the notion of conditioning on a space in a tropical diagram and show that the operation is Lipschitz continuous with respect to the asymptotic entropy distance.

It is a rather technical result, and we have, therefore, decided to treat it in this separate article, but it is an important ingredient in the theory and, in particular, we need it for the dimension-reduction result mentioned before.

Given a tuple of finite-valued random variables $(X_i)_{i=1}^n$ and a random variable Y , one may "condition" the collection (X_i) on Y . The result of this operation is a family of n -tuples of random variables denoted $(X_i|Y)_{i=1}^n$ parameterized by those values of Y that have positive probability. Each tuple of random variables in this family is defined on a separate probability space.

When passing to the tropical setting, the situation is different in the sense that when we condition a tropical diagram $[\mathcal{X}]$ on a space $[Y]$, the result is again a tropical diagram $[\mathcal{X}|Y]$ rather than a family. After recalling some preliminaries in Section 2, we describe the operation of conditioning and prove that the result depends in a Lipschitz way on the original diagram in Section 3.

2. Preliminaries

Our main objects of study are commutative diagrams of probability spaces and their tropical counterparts. In this section, we recall briefly the main definitions and results.

2.1. Probability Spaces and Their Diagrams

2.1.1. Probability Spaces

By a *finite probability space*, we mean a set with a probability measure that has finite support. A *reduction* from one probability space to another is an equivalence class of measure-preserving maps. Two maps are equivalent if they coincide on a set of full measures. We call a point x in a probability space $X = (\underline{X}, p)$ an *atom* if it has positive



Citation: Matveev, R.; Portegies, J.W. Conditioning in Tropical Probability Theory. *Entropy* **2023**, *25*, 1641. <https://doi.org/10.3390/e25121641>

Academic Editor: Daniel Chicharro

Received: 6 November 2023

Revised: 4 December 2023

Accepted: 7 December 2023

Published: 9 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

weight, and we write $x \in X$ to mean x is an atom in X (as opposed to $x \in \underline{X}$ for points in the underlying set). For a probability space X , we denote by $|X|$ the cardinality of the support of the probability measure.

2.1.2. Indexing Categories

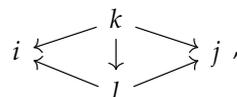
To record the combinatorial structure of commutative diagrams of probability spaces and reductions, we use an object that we call an *indexing category*. By an indexing category, we mean a finite category \mathbf{G} such that for any pair of objects $i, j \in \mathbf{G}$, there is at most one morphism between them either way. In addition, we will assume it satisfies one additional property that we will describe after introducing some terminology. For a pair of objects $i, j \in \mathbf{G}$ such that there is a morphism $\gamma_{ij} : i \rightarrow j$, object i will be called an *ancestor* of j and object j will be called a *descendant* of i . The subcategory of all descendants of an object $i \in \mathbf{G}$ is called an *ideal* generated by i and will be denoted $[i]$, while we will call the subcategory consisting of all ancestors of i together with all the morphisms in it a *co-ideal* generated by i and denote it by $\lceil i \rceil$. (The term *filter* is also used for a co-ideal in the literature about lattices).

The additional property that an indexing category has to satisfy is that for any pair of objects $i, j \in \mathbf{G}$, there exists a *minimal common ancestor* \hat{i} , and \hat{i} is an ancestor for both i and j and any other ancestor of them both is also an ancestor of \hat{i} ; in other words, \mathbf{G} is an upper semi-lattice.

An equivalent formulation of the property above is the following: the intersection of the co-ideals generated by two objects $i, j \in \mathbf{G}$ is also a co-ideal generated by some object $\hat{i} \in \mathbf{G}$.

Any indexing category \mathbf{G} is necessarily *initial*, which means that there exists an *initial object*, that is an object i_0 such that $\mathbf{G} = [i_0]$.

A *fan* in a category is a pair of morphisms with the same domain. A fan $(i \leftarrow k \rightarrow j)$ is called *minimal* if for any other fan $(i \leftarrow l \rightarrow j)$ included in a commutative diagram



the vertical arrow must be an isomorphism; in other words, k is a minimal common ancestor of i and j .

For any pair of objects i, j in an indexing category \mathbf{G} , there exists a unique minimal fan $(i \leftarrow \hat{i} \rightarrow j)$ in \mathbf{G} .

2.1.3. Diagrams

We denote by \mathbf{Prob} the category of finite probability spaces and reductions, i.e., the equivalence classes of measure-preserving maps. For an indexing category $\mathbf{G} = \{i; \gamma_{ij}\}$, a \mathbf{G} -diagram is a functor $\mathcal{X} : \mathbf{G} \rightarrow \mathbf{Prob}$. A reduction f from one \mathbf{G} -diagram $\mathcal{X} = \{X_i; \chi_{ij}\}$ to another $\mathcal{Y} = \{Y_i; v_{ij}\}$ is a natural transformation between the functors. It amounts to a collection of reductions $f_i : X_i \rightarrow Y_i$, such that the big diagram consisting of all spaces X_i, Y_i and all morphisms χ_{ij}, v_{ij} and f_i is commutative. The category of \mathbf{G} -diagrams and reductions will be denoted as $\mathbf{Prob}\langle \mathbf{G} \rangle$. The construction of diagrams could be iterated; thus, we can consider \mathbf{H} -diagrams of \mathbf{G} -diagrams and denote the corresponding category $\mathbf{Prob}\langle \mathbf{G} \rangle\langle \mathbf{H} \rangle = \mathbf{Prob}\langle \mathbf{G}, \mathbf{H} \rangle$. Every \mathbf{H} -diagram of a \mathbf{G} -diagram can also be considered as \mathbf{G} -diagrams of \mathbf{H} -diagrams; thus, there is a natural equivalence of categories $\mathbf{Prob}\langle \mathbf{G}, \mathbf{H} \rangle \cong \mathbf{Prob}\langle \mathbf{H}, \mathbf{G} \rangle$.

A \mathbf{G} -diagram \mathcal{X} will be called *minimal* if it maps minimal fans in \mathbf{G} to minimal fans in the target category. The subspace of all minimal \mathbf{G} -diagrams will be denoted $\mathbf{Prob}\langle \mathbf{G} \rangle_m$. In [1], we have shown that for any fan in \mathbf{Prob} or in $\mathbf{Prob}\langle \mathbf{G} \rangle$, its minimization exists and is unique up to isomorphism.

2.1.4. Tensor Product

The tensor product of two probability spaces $X = (\underline{X}, p)$ and $Y = (\underline{Y}, q)$ is their independent product $X \otimes Y := (\underline{X} \times \underline{Y}, p \otimes q)$. For two \mathbf{G} -diagrams $\mathcal{X} = \{X_i; \chi_{ij}\}$ and $\mathcal{Y} = \{Y_i; \nu_{ij}\}$, we define their tensor product to be $\mathcal{X} \otimes \mathcal{Y} = \{X_i \otimes Y_j; \chi_{ij} \times \nu_{ij}\}$.

2.1.5. Constant Diagrams

Given an indexing category \mathbf{G} and a probability space, we can form a *constant* diagram $X^{\mathbf{G}}$ that has all spaces equal to X and all reductions equal to the identity isomorphism. Sometimes, when such a constant diagram is included in a diagram with other \mathbf{G} -diagrams (such as, for example, a reduction $\mathcal{X} \rightarrow X^{\mathbf{G}}$), we will write simply X in place of $X^{\mathbf{G}}$.

2.1.6. Entropy

Evaluating entropy on every space in a \mathbf{G} -diagram, we obtain a tuple of non-negative numbers indexed by objects in \mathbf{G} ; thus, entropy gives a map

$$\text{Ent}_* : \mathbf{Prob}\langle \mathbf{G} \rangle \rightarrow \mathbb{R}^{\mathbf{G}},$$

where the target space $\mathbb{R}^{\mathbf{G}}$ is a space of real-valued functions on the set of objects in \mathbf{G} endowed with the ℓ^1 -norm. Entropy is a homomorphism in that it satisfies

$$\text{Ent}_*(\mathcal{X} \otimes \mathcal{Y}) = \text{Ent}_*(\mathcal{X}) + \text{Ent}_*(\mathcal{Y}).$$

2.1.7. Entropy Distance

Let \mathbf{G} be an indexing category and $\mathcal{K} = (\mathcal{X} \leftarrow \mathcal{Z} \rightarrow \mathcal{Y})$ be a fan of \mathbf{G} -diagrams. We define the *entropy distance* as

$$\text{kd}(\mathcal{K}) := \|\text{Ent}_* \mathcal{Z} - \text{Ent}_* \mathcal{X}\|_1 + \|\text{Ent}_* \mathcal{Z} - \text{Ent}_* \mathcal{Y}\|_1.$$

The *intrinsic entropy distance* between two \mathbf{G} -diagrams is defined to be the infimal entropy distance of all fans with terminal diagrams \mathcal{X} and \mathcal{Y} :

$$\mathbf{k}(\mathcal{X}, \mathcal{Y}) := \inf\{\text{kd}(\mathcal{K}) : \mathcal{K} = (\mathcal{X} \leftarrow \mathcal{Z} \rightarrow \mathcal{Y})\}.$$

The intrinsic entropy distance was introduced in [10,11] for probability spaces.

In [1], it is shown that the infimum is attained, that the optimal fan is minimal, that \mathbf{k} is a pseudo-distance, which vanishes if, and only if, \mathcal{X} and \mathcal{Y} are isomorphic, and that Ent_* is a 1-Lipschitz linear functional with respect to \mathbf{k} .

2.2. Diagrams of Sets, Distributions, and Empirical Reductions

2.2.1. Distributions on Sets

For a set S , we denote by ΔS the collection of all finitely-supported probability distributions on S . For a pair of distributions $\pi_1, \pi_2 \in \Delta S$, we denote by $\|\pi_1 - \pi_2\|_1$ the *total variation distance* between them.

For a map $f : S \rightarrow S'$ between two sets, we denote by $f_* : \Delta S \rightarrow \Delta S'$ the induced affine map (the map-preserving convex combinations).

For $n \in \mathbb{N}$, we define the *empirical map* $\mathbf{q} : S^n \rightarrow \Delta S$ by the assignment below. For $\bar{s} = (s_1, \dots, s_n) \in S^n$ and $A \subset S$, we define

$$\mathbf{q}(\bar{s})(A) := \frac{1}{n} \cdot |\{k : s_k \in A\}|.$$

For a finite probability space $X = (S, p)$, the *empirical distribution* on ΔX is the push-forward $\tau_n := \mathbf{q}_* p^{\otimes n}$. Thus,

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

is a reduction of finite probability spaces. The construction of empirical reduction is functorial, which is for a reduction between two probability spaces $f : X \rightarrow Y$, the diagram of the reductions

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

commutes.

2.2.2. Distributions on Diagrams of Sets

Let **Set** denote the category of sets and surjective maps. For an indexing category **G**, we denote by **Set(G)** the category of **G**-diagrams in **Set**. The objects in **Set(G)** are commutative diagrams of sets indexed by **G**, and the spaces in such a diagram are sets, where the arrows represent surjective maps, subject to commutativity relations.

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

$$\Delta \mathcal{S} := \left\{ (\pi_i) \in \prod_i \Delta S_i : (\sigma_{ij})_* \pi_i = \pi_j \right\}.$$

If S_0 is the initial set of \mathcal{S} , then there is an isomorphism of

$$\begin{aligned} \Delta S_0 &\xrightarrow{\cong} \Delta \mathcal{S} & (1) \\ \Delta S_0 \ni \pi_0 &\mapsto \{(\sigma_{0i})_* \pi_0\} \in \Delta \mathcal{S} \\ \Delta S_0 \ni \pi_0 &\mapsto \{\pi_i\} \in \Delta \mathcal{S}. \end{aligned}$$

Given a **G**-diagram of sets $\mathcal{S} = \{S_i; \sigma_{ij}\}$ and an element $\pi \in \Delta \mathcal{S}$, we can construct a **G**-diagram of probability spaces $(\mathcal{S}, \pi) := \{(S_i, \pi_i); \sigma_{ij}\}$. Note that any diagram \mathcal{X} of probability spaces has this form.

2.3. Conditioning

Consider a **G**-diagram of probability spaces $\mathcal{X} = (\mathcal{S}, \pi)$, where \mathcal{S} is a diagram of sets and $\pi \in \Delta \mathcal{S}$. Let $X_0 = (S_0, \pi_0)$ be the initial space in \mathcal{X} and $U := X_i$ be another space in \mathcal{X} . Since S_0 is initial, there is a map $\sigma_{0,i} : S_0 \rightarrow S_i$. Fix an atom $u \in U$ and define the conditioned distribution $\pi_0(\cdot|u)$ on S_0 as the distribution supported in $\sigma_{0,i}^{-1}(u)$ and for every $s \in \sigma_{0,i}^{-1}(u)$ defined by

$$\pi_0(s|u) := \frac{\pi_0(s)}{\pi_0(\sigma_{0,i}^{-1}(u))}.$$

Let $\pi(\cdot|u) \in \Delta \mathcal{S}$ be the distribution corresponding to $\pi_0(\cdot|u)$ under the isomorphism in (1). We define the *conditioned G*-diagram as $\mathcal{X}|u := (\mathcal{S}, \pi(\cdot|u))$.

2.4. The Slicing Lemma

In [1], we prove the so-called Slicing Lemma that allows us to estimate the intrinsic entropy distance between two diagrams in terms of distances between conditioned diagrams. Among the corollaries of the Slicing Lemma is the following inequality.

Proposition 1. *Let $(\mathcal{X} \leftarrow \hat{\mathcal{X}} \rightarrow U^{\mathbf{G}}) \in \mathbf{Prob}(\mathbf{G}, \Lambda_2)$ be a fan of **G**-diagrams of probability spaces and $\mathcal{Y} \in \mathbf{Prob}(\mathbf{G})$ be another diagram. Then,*

$$\mathbf{k}(\mathcal{X}, \mathcal{Y}) \leq \int_U \mathbf{k}(\mathcal{X}|u, \mathcal{Y}) dp(u) + 2\llbracket \mathbf{G} \rrbracket \cdot \text{Ent } U.$$

The fan in the assumption of the proposition above can often be constructed in the following manner. Suppose \mathcal{X} is a **G**-diagram and $U := X_\iota$ is a space in it for some $\iota \in \mathbf{G}$. We can construct a fan $(\mathcal{X} \xleftarrow{f} \hat{\mathcal{X}} \xrightarrow{g} U^{\mathbf{G}}) \in \mathbf{Prob}(\mathbf{G}, \Lambda_2)$ by assigning \hat{X}_ι to be the initial space

of the (unique) minimal fan in \mathcal{X} with terminal spaces X_i and U and f_i and g_i to be left and right reductions in that fan for any $i \in \mathbf{G}$.

2.5. Tropical Diagrams

A detailed discussion of the topics in this section can be found in [2].

The asymptotic entropy distance between two diagrams of the same combinatorial type is defined by

$$\kappa(\mathcal{X}, \mathcal{Y}) := \lim \frac{1}{n} \mathbf{k}(\mathcal{X}^n, \mathcal{Y}^n).$$

A tropical \mathbf{G} -diagram is an equivalence class of certain sequences of \mathbf{G} -diagrams of probability spaces. Below, we describe the type of sequences and the equivalence relation.

A function $\varphi : \mathbb{R} \geq 1 \rightarrow \mathbb{R} \geq 0$ is called an *admissible function* if φ is non-decreasing and there is a constant D_φ , such that for any $t \geq 1$:

$$t \cdot \int_t^\infty \frac{\varphi(s)}{s^2} ds \leq D_\varphi \cdot \varphi(t).$$

An example of an admissible function will be $\varphi(t) = t^\alpha$ for $\alpha \in [0, 1)$.

A sequence $\vec{\mathcal{X}} = (\mathcal{X}(n) : n \in \mathbb{N}_0)$ of diagrams of probability spaces will be called *quasi-linear* with the *defect* bounded by an admissible function φ if for some $C > 0$ and all $m, n \in \mathbb{N}$, it satisfies

$$\kappa(\mathcal{X}(n+m), \mathcal{X}(n) \otimes \mathcal{X}(m)) \leq C \cdot \varphi(n+m).$$

For example for a diagram \mathcal{X} , the sequence $\vec{\mathcal{X}} := (\mathcal{X}^n : n \in \mathbb{N}_0)$ is φ -quasi-linear for $\varphi \equiv 0$ (and for any admissible φ). Sequences with zero defect are called *linear*, and the space of all linear sequences in $\mathbf{Prob}(\mathbf{G})$ is denoted by $L(\mathbf{Prob}(\mathbf{G}))$.

The asymptotic entropy distance between two φ -quasi-linear sequences $\vec{\mathcal{X}} = (\mathcal{X}(n) : n \in \mathbb{N}_0)$ and $\vec{\mathcal{Y}} = (\mathcal{Y}(n) : n \in \mathbb{N}_0)$ is defined to be

$$\kappa(\vec{\mathcal{X}}, \vec{\mathcal{Y}}) := \lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{k}(\mathcal{X}(n), \mathcal{Y}(n)),$$

and the sequences are called *asymptotically equivalent* if $\kappa(\vec{\mathcal{X}}, \vec{\mathcal{Y}}) = 0$. An equivalence class of a sequence $\vec{\mathcal{X}}$ will be denoted as $[\vec{\mathcal{X}}]$, and the totality of all the classes as $\mathbf{Prob}[\mathbf{G}]$. We have shown in [2] that the space of equivalence classes of φ -quasi-linear sequences does not depend on the choice of a non-zero admissible function φ .

The sum of two such equivalence classes is defined to be the equivalence class of the sequence obtained by tensor-multiplying representative sequences of the summands term-wise. In addition, there is a doubly transitive action of $\mathbb{R}_{\geq 0}$ on $\mathbf{Prob}[\mathbf{G}]$. In [2], the following theorem is proven.

Theorem 1. *Let \mathbf{G} be an indexing category. Then:*

1. *The space $\mathbf{Prob}[\mathbf{G}]$ does not depend on the choice of a positive admissible function φ up to isometry.*
2. *The space $\mathbf{Prob}[\mathbf{G}]$ is metrically complete.*
3. *The map $\mathcal{X} \mapsto \vec{\mathcal{X}}$ is a κ - κ -isometric embedding. The space of linear sequences, i.e., the image of the map above, is dense in $\mathbf{Prob}[\mathbf{G}]$.*
4. *There is a distance-preserving homomorphism from $\mathbf{Prob}[\mathbf{G}]$ into a Banach space B , whose image is a closed convex cone in B .*
5. *The entropy functional*

$$\begin{aligned} \text{Ent}_* : \mathbf{Prob}[\mathbf{G}] &\rightarrow \mathbb{R}^{\mathbf{G}} \\ [(\mathcal{X}(n))_{n \in \mathbb{N}_0}] &\mapsto \lim_{n \rightarrow \infty} \frac{1}{n} \text{Ent}_* \mathcal{X}(n) \end{aligned}$$

is a well-defined 1-Lipschitz linear map.

2.6. Asymptotic Equipartition Property for Diagrams

Among all \mathbf{G} -diagrams, there is a special class of maximally symmetric ones. We call such diagrams *homogeneous*; see below for the definition. Homogeneous diagrams come in very handy in many considerations, because their structure is easier to describe than that of general diagrams. We show below that among the tropical diagrams, those that have homogeneous representatives are dense. It means, in particular, that when considering continuous functionals in the space of diagrams, it suffices to only study them in the space of all homogeneous diagrams.

2.6.1. Homogeneous Diagrams

A \mathbf{G} -diagram \mathcal{X} is called *homogeneous* if the automorphism group $\text{Aut}(\mathcal{X})$ acts transitively on every space in \mathcal{X} , by which we mean that the action is transitive on the support of the probability measure. Homogeneous probability spaces are isomorphic to uniform spaces. For more complex indexing categories, this simple description is not sufficient.

2.6.2. Tropical Homogeneous Diagrams

The subcategory of all homogeneous \mathbf{G} -diagrams will be denoted $\mathbf{Prob}\langle\mathbf{G}\rangle_{\text{h}}$, and we write $\mathbf{Prob}\langle\mathbf{G}\rangle_{\text{h,m}}$ for the category of minimal homogeneous \mathbf{G} -diagrams. These spaces are invariant under the tensor product; thus, they are metric Abelian monoids, and the general "tropicalization" described in [2] can be performed. Passing to the tropical limit, we obtain spaces of tropical (minimal) homogeneous diagrams, which we denote by $\mathbf{Prob}[\mathbf{G}]_{\text{h}}$ and $\mathbf{Prob}[\mathbf{G}]_{\text{h,m}}$, respectively.

2.6.3. Asymptotic Equipartition Property

For an indexing category \mathbf{G} , denote by $[\![\mathbf{G}]\!]$ the number of objects in \mathbf{G} . In [1], the following theorem is proven.

Theorem 2. *Suppose $\mathcal{X} \in \mathbf{Prob}\langle\mathbf{G}\rangle$ is a \mathbf{G} -diagram of probability spaces for some fixed indexing category \mathbf{G} . Then, there exists a sequence $\mathcal{H} = (\mathcal{H}_n)_{n=0}^{\infty}$ of homogeneous \mathbf{G} -diagrams, such that*

$$\frac{1}{n} \mathbf{k}(\mathcal{X}^n, \mathcal{H}_n) \leq C(|X_0|, [\![\mathbf{G}]\!]) \cdot \sqrt{\frac{\ln^3 n}{n}}, \tag{2}$$

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

The approximating sequence of homogeneous diagrams is evidently quasi-linear with the defect bounded by the admissible function

$$\varphi(t) := 2C(|X_0|, [\![\mathbf{G}]\!]) \cdot t^{3/4} \geq 2C(|X_0|, [\![\mathbf{G}]\!]) \cdot t^{1/2} \cdot \ln^{3/2} t.$$

Thus, Theorem 2 above states that $L(\mathbf{Prob}\langle\mathbf{G}\rangle) \subset \mathbf{Prob}[\mathbf{G}]_{\text{h}}$. On the other hand, we have shown in [2] that the space of linear sequences $L(\mathbf{Prob}\langle\mathbf{G}\rangle)$ is dense in $\mathbf{Prob}[\mathbf{G}]$. Combining the two statements, we obtain the following theorem.

Theorem 3. *For any indexing category \mathbf{G} , the space $\mathbf{Prob}[\mathbf{G}]_{\text{h}}$ is dense in $\mathbf{Prob}[\mathbf{G}]$. Similarly, the space $\mathbf{Prob}[\mathbf{G}]_{\text{h,m}}$ is dense in $\mathbf{Prob}[\mathbf{G}]_{\text{m}}$.*

3. Conditioning of Tropical Diagrams

3.1. Motivation

Let $\mathcal{X} \in \mathbf{Prob}\langle\mathbf{G}\rangle$ be a \mathbf{G} -diagram of probability spaces containing probability space $U = X_{i_0}$ indexed by an object $i_0 \in \mathbf{G}$.

Given an atom $u \in U$, we can define a conditioned diagram $\mathcal{X}|u$. If the diagram \mathcal{X} is homogeneous, then the isomorphism class of $\mathcal{X}|u$ is independent of u , so that $(\mathcal{X}|u : u \in U)$ is a constant family. On the other hand, we have shown that the power of any diagram can be approximated by homogeneous diagrams, thus suggesting that in the tropical setting $\mathcal{X}|U$ should be a well-defined tropical diagram, rather than a family. Below, we give a definition of the tropical conditioning operation and prove its consistency.

3.2. Classical-Tropical Conditioning

Here, we define the operation of conditioning of the classical diagram, such that the result is a tropical diagram. Let \mathcal{X} be a \mathbf{G} -diagram of probability spaces and U be a space in \mathcal{X} . We define the conditioning map

$$[\cdot] : \mathbf{Prob}(\mathbf{G}) \rightarrow \mathbf{Prob}[\mathbf{G}]$$

by conditioning \mathcal{X} by $u \in U$ and averaging the corresponding tropical diagrams:

$$[\mathcal{X}|U] := \int_{u \in U} \overrightarrow{(\mathcal{X}|u)} dp_U(u),$$

where $\overrightarrow{(\mathcal{X}|u)}$ is the tropical diagram represented by a linear sequence generated by $\mathcal{X}|u$; see Section 2.5. Note that the integral on the right-hand side is just a finite convex combination of tropical diagrams. Expanding all the definitions, we will obtain for $[\mathcal{Y}] := [\mathcal{X}|U]$, the representative sequence

$$\mathcal{Y}(n) = \bigotimes_{u \in U} (\mathcal{X}|u)^{[n \cdot p(u)]}.$$

3.3. Properties

3.3.1. Conditioning of Homogeneous Diagrams

If the diagram \mathcal{X} is *homogeneous*, then for any atom $u \in U$, with a positive weight,

$$[\mathcal{X}|U] \overset{\check{}}{=} \overrightarrow{(\mathcal{X}|u)}.$$

3.3.2. Entropy

By definition, the conditioned entropy is

$$\text{Ent}_*(\mathcal{X}|U) := \int_U \text{Ent}_*(\mathcal{X}|u) dp_U(u).$$

Now that $[\mathcal{X}|U]$ is a tropical diagram, the expression $\text{Ent}_*(\mathcal{X}|U)$ can be interpreted in two, a priori different, ways: by the formula above and as the entropy of the object introduced in the previous subsection. Fortunately, the numeric value of it does not depend on the interpretation since the entropy is a linear functional on $\mathbf{Prob}[\mathbf{G}]$.

3.3.3. Additivity

If \mathcal{X} and \mathcal{Y} are two \mathbf{G} -diagrams with $U := X_\iota, V := Y_\iota$ for some $\iota \in \mathbf{G}$, then

$$[(\mathcal{X} \otimes \mathcal{Y})|(U \otimes V)] = [\mathcal{X}|U] + [\mathcal{Y}|V].$$

Proof.

$$\begin{aligned} [(\mathcal{X} \otimes \mathcal{Y})|(U \otimes V)] &= \int_{U \otimes V} \overrightarrow{(\mathcal{X} \otimes \mathcal{Y})|(u, v)} dp(u) dp(v) \\ &= \int_{U \otimes V} (\overrightarrow{(\mathcal{X}|u)} + \overrightarrow{(\mathcal{Y}|v)}) dp(u) dp(v) = \int_U \overrightarrow{(\mathcal{X}|u)} dp(u) + \int_V \overrightarrow{(\mathcal{Y}|v)} dp(v) \\ &= [\mathcal{X}|U] + [\mathcal{Y}|V] \end{aligned}$$

□

3.3.4. Homogeneity

It follows that for any diagram \mathcal{X} , a space U in \mathcal{X} and $n \in \mathbb{N}_0$ holds

$$[\mathcal{X}^n|U^n] = n \cdot [\mathcal{X}|U].$$

3.4. Continuity and Lipschitz Property

Proposition 2. Let \mathbf{G} be an indexing category, $\mathcal{X}, \mathcal{Y} \in \mathbf{Prob}(\mathbf{G})$ be two \mathbf{G} diagrams, and $U := X_\iota$ and $V := Y_\iota$ be two spaces in \mathcal{X} and \mathcal{Y} , respectively, indexed by some $\iota \in \mathbf{G}$. Then,

$$\kappa([\mathcal{X}|U], [\mathcal{Y}|V]) \leq (2 \cdot \llbracket \mathbf{G} \rrbracket + 1) \cdot \mathbf{k}(\mathcal{X}, \mathcal{Y}).$$

Using the homogeneity property of conditioning, Section 3.3.4, we can obtain the following stronger inequality.

Corollary 1. In the setting of Proposition 2, the following holds:

$$\kappa([\mathcal{X}|U], [\mathcal{Y}|V]) \leq (2 \cdot \llbracket \mathbf{G} \rrbracket + 1) \cdot \kappa(\mathcal{X}, \mathcal{Y}).$$

Before we prove Proposition 2, we will need some preparatory lemmas.

Lemma 1. Let \mathcal{A} be a \mathbf{G} -diagram of probability spaces and E be a space in it. Let $\mathbf{q} : E^n \rightarrow (\Delta E, \tau_n)$ be the empirical reduction. Then, for any $n \in \mathbb{N}$ and any $\bar{e}, \bar{e}' \in E^n$

$$\mathbf{k}(\mathcal{A}^n|\bar{e}, \mathcal{A}^n|\bar{e}') \leq n \cdot \|\text{Ent}_*(\mathcal{A})\|_1 \cdot \|\mathbf{q}(\bar{e}) - \mathbf{q}(\bar{e}')\|_1.$$

Proof. To prove the lemma, we construct a coupling between $\mathcal{A}^n|\bar{e}$ and $\mathcal{A}^n|\bar{e}'$ in the following manner. Note that there exists a permutation $\sigma \in S_n$, such that

$$|\{i : e_i \neq e'_{\sigma i}\}| = \frac{n}{2} \cdot \|\mathbf{q}(\bar{e}) - \mathbf{q}(\bar{e}')\|_1.$$

Let

$$\begin{aligned} I &= \{i : e_i = e'_{\sigma i}\} \\ \tilde{I} &= \{i : e_i \neq e'_{\sigma i}\}. \end{aligned}$$

Using that $|\tilde{I}| = \frac{n}{2} \cdot \|\mathbf{q}(\bar{e}) - \mathbf{q}(\bar{e}')\|_1$, we can estimate

$$\begin{aligned} \mathbf{k}(\mathcal{A}^n|\bar{e}, \mathcal{A}^n|\bar{e}') &= \mathbf{k}\left(\bigotimes_{i=1}^n (\mathcal{A}|e_i), \bigotimes_{i=1}^n (\mathcal{A}|e'_{\sigma i})\right) \\ &\leq \sum_{i \in I} \text{kd}(\mathcal{A}|e_i \overset{\cong}{\leftarrow} \mathcal{A}|e'_{\sigma i}) + \sum_{i \in \tilde{I}} \text{kd}(\mathcal{A}|e_i \overset{\otimes}{\leftarrow} \mathcal{A}|e'_{\sigma i}) \\ &\leq n \cdot \|\text{Ent}_*(\mathcal{A})\|_1 \cdot \|\mathbf{q}(\bar{e}) - \mathbf{q}(\bar{e}')\|_1, \end{aligned}$$

where $\mathcal{A} \overset{\cong}{\leftarrow} \mathcal{B}$ denotes the isomorphism coupling of two naturally isomorphic diagrams, while $\mathcal{A} \overset{\otimes}{\leftarrow} \mathcal{B}$ denotes the ‘‘independence’’ coupling. \square

Lemma 2. Let \mathcal{A} be a \mathbf{G} -diagram of probability spaces and E be a space in \mathcal{A} . Then,

$$\int_{E^n} \mathbf{k}(\mathcal{A}^n, \mathcal{A}^n|\bar{e}) \, dp(\bar{e}) \leq 2n \cdot \llbracket \mathbf{G} \rrbracket \cdot \text{Ent}(E) + o(n).$$

Proof. First, we apply Proposition 1 slicing the first argument:

$$\int_{E^n} \mathbf{k}(\mathcal{A}^n, \mathcal{A}^n | \bar{e}) dp(\bar{e}) \leq \int_{E^n} \int_{E^n} \mathbf{k}(\mathcal{A}^n | \bar{e}', \mathcal{A}^n | \bar{e}) dp(\bar{e}') dp(\bar{e}) + 2n \cdot \llbracket \mathbf{G} \rrbracket \cdot \text{Ent}(E).$$

We will argue now that the double integral on the right-hand side grows sub-linearly with n . We estimate the double integral by applying Lemma 1 to the integrand

$$\begin{aligned} \int_{E^n} \int_{E^n} \mathbf{k}(\mathcal{A}^n | \bar{e}', \mathcal{A}^n | \bar{e}) dp(\bar{e}') dp(\bar{e}) &\leq \int_{E^n} \int_{E^n} n \cdot \llbracket \mathbf{G} \rrbracket \cdot |\text{Ent}_*(\mathcal{A})|_1 \cdot |\mathbf{q}(\bar{e}) - \mathbf{q}(\bar{e}')|_1 dp(\bar{e}') dp(\bar{e}) \\ &= n \cdot \llbracket \mathbf{G} \rrbracket \cdot |\text{Ent}_*(\mathcal{A})|_1 \cdot \int_{\Delta E} \int_{\Delta E} |\pi - \pi'|_1 d\tau_n(\pi) d\tau_n(\pi') = o(n), \end{aligned}$$

where the convergence to zero of the last double integral follows from Sanov’s theorem. \square

Corollary 2. Let \mathcal{A} be a \mathbf{G} -diagram and E a probability space included in \mathcal{A} . Then,

$$\kappa(\vec{\mathcal{A}}, [\mathcal{A}|E]) \leq 2\llbracket \mathbf{G} \rrbracket \cdot \text{Ent}(E).$$

Proof. Let $n \in \mathbb{N}$. Then,

$$\begin{aligned} \kappa(\vec{\mathcal{A}}, [\mathcal{A}|E]) &= \frac{1}{n} \kappa(\vec{\mathcal{A}}^n, [\mathcal{A}^n|E^n]) \\ &= \frac{1}{n} \kappa(\vec{\mathcal{A}}^n, \int_{E^n} \vec{\mathcal{A}}^n | \bar{e} dp(\bar{e})) \\ &\leq \frac{1}{n} \int_{E^n} \kappa(\vec{\mathcal{A}}^n, \vec{\mathcal{A}}^n | \bar{e}) dp(\bar{e}) \\ &= \frac{1}{n} \int_{E^n} \kappa(\mathcal{A}^n, \mathcal{A}^n | \bar{e}) dp(\bar{e}) \\ &\leq 2 \cdot \llbracket \mathbf{G} \rrbracket \cdot \text{Ent}(E) + o(n^0), \end{aligned}$$

where we used Lemma 2 and the fact that $\kappa \leq \mathbf{k}$ in the last line. We finish the proof by taking the limit $n \rightarrow \infty$. \square

Proof of Proposition 2. We start with a note on general terminology: a reduction $f : A \rightarrow B$ of probability spaces can also be considered as a fan $\mathcal{F} := (A \xleftarrow{\bar{f}} A \xrightarrow{f} B)$. Then, the entropy distance of f is

$$\text{kd}(f) := \text{kd}(\mathcal{F}) = \text{Ent } A - \text{Ent } B.$$

If the reduction f is a part of a bigger diagram containing also space U , then the following inequality holds:

$$\int_U \text{kd}(f|u) dp(u) \leq \text{kd}(f).$$

Let $\mathcal{K} \in \text{Prob}(\mathbf{G}, \Lambda_2)$

$$\mathcal{K} = \left(\mathcal{X} \xleftarrow{f} \mathcal{Z} \xrightarrow{g} \mathcal{Y} \right) \in \text{Prob}(\mathbf{G}, \Lambda_2) = \text{Prob}(\Lambda_2, \mathbf{G})$$

be an optimal coupling between \mathcal{X} and \mathcal{Y} . It can also be viewed as a \mathbf{G} -diagram of two fans, $\mathcal{K} = \{\mathcal{K}_i\}_{i \in \mathbf{G}}$, each of which is a minimal coupling between X_i and Y_i . Among them is the minimal fan $\mathcal{W} := \mathcal{K}_l = (U \xleftarrow{f_l} W \xrightarrow{g_l} V)$.

We use the triangle inequality to bound the distance $\kappa([\mathcal{X}|U], [\mathcal{Y}|V])$ by four summands as follows:

$$\begin{aligned} \kappa([\mathcal{X}|U], [\mathcal{Y}|V]) &\leq \kappa([\mathcal{X}|U], [\mathcal{Z}|U]) + \kappa([\mathcal{Z}|U], [\mathcal{Z}|W]) + \\ &\quad \kappa([\mathcal{Z}|W], [\mathcal{Z}|V]) + \kappa([\mathcal{Z}|V], [\mathcal{Y}|V]). \end{aligned}$$

We will estimate each of the four summands separately. The bound for the first one is as follows:

$$\begin{aligned} \kappa([\mathcal{X}|U], [\mathcal{Z}|U]) &= \kappa\left(\int_U \overrightarrow{\mathcal{X}|u} dp(u), \int_U \overrightarrow{\mathcal{Z}|u} dp(u)\right) \\ &\leq \int_U \kappa(\overrightarrow{\mathcal{X}|u}, \overrightarrow{\mathcal{Z}|u}) dp(u) = \int_U \kappa(\mathcal{X}|u, \mathcal{Z}|u) dp(u) \\ &\leq \int_U \mathbf{k}(\mathcal{X}|u, \mathcal{Z}|u) dp(u) \leq \int_U \text{kd}(f|u) dp(u) \\ &\leq \sum_{i \in \mathbf{G}} \int_U \text{kd}(f_i|u) dp(u) = \sum_{i \in \mathbf{G}} \text{kd}(f_i) = \text{kd}(f). \end{aligned}$$

An analogous calculation shows that

$$\kappa([\mathcal{Z}|V], [\mathcal{Y}|V]) \leq \text{kd}(g).$$

To bound the second summand, we will use Corollary 2:

$$\begin{aligned} \kappa([\mathcal{Z}|U], [\mathcal{Z}|W]) &= \kappa\left(\int_U \overrightarrow{\mathcal{Z}|u} dp(u), \int_W \overrightarrow{\mathcal{Z}|w} dp(w)\right) \\ &= \kappa\left(\int_U \overrightarrow{\mathcal{Z}|u} dp(u), \int_U \int_{W|u} \overrightarrow{\mathcal{Z}|w} dp(w|u) dp(u)\right) \\ &\leq \int_U \kappa(\overrightarrow{\mathcal{Z}|u}, \int_{W|u} \overrightarrow{\mathcal{Z}|w} dp(w|u)) dp(u). \end{aligned}$$

We will now use Corollary 2 with $\mathcal{A} = \mathcal{Z}|u$ and $E = W|u$ to estimate the integrand. Then,

$$\begin{aligned} \kappa([\mathcal{Z}|U], [\mathcal{Z}|W]) &= \int_U \kappa\left(\overrightarrow{\mathcal{Z}|u}, \int_{W|u} \overrightarrow{\mathcal{Z}|w} dp(w|u)\right) dp(u) \\ &\leq 2\llbracket \mathbf{G} \rrbracket \cdot \int_U \text{Ent}(W|u) dp(u) \\ &\leq 2\llbracket \mathbf{G} \rrbracket \cdot \text{Ent}(W|U) \leq 2\llbracket \mathbf{G} \rrbracket \cdot \text{kd}(f). \end{aligned}$$

Similarly,

$$\kappa([\mathcal{Z}|W], [\mathcal{Z}|V]) \leq 2\llbracket \mathbf{G} \rrbracket \cdot \text{kd}(g).$$

Combining the estimates, we obtain

$$\kappa([\mathcal{X}|U], [\mathcal{Y}|V]) \leq (2\llbracket \mathbf{G} \rrbracket + 1) \cdot (\text{kd}(f) + \text{kd}(g)) = (2\llbracket \mathbf{G} \rrbracket + 1) \cdot \mathbf{k}(\mathcal{X}, \mathcal{Y}).$$

□

3.5. Tropical Conditioning

Let $[\mathcal{X}]$ be a tropical \mathbf{G} -diagram and $[U] = [X_\iota]$ for some $\iota \in \mathbf{G}$. Choose a representative $(\mathcal{X}(n))_{n \in \mathbb{N}_0}$ and denote $u(n) := X_\iota(n)$. We define now a conditioned diagram $[\mathcal{X}|U]$ by the following limit:

$$[\mathcal{X}|U] := \lim_{n \rightarrow \infty} \frac{1}{n} [\mathcal{X}(n)|U(n)].$$

Proposition 1 guarantees that the limit exists and is independent of the choice of representative. For a fixed $\iota \in \mathbf{G}$, the conditioning is a linear Lipschitz map of

$$[\cdot | \cdot] : \mathbf{Prob}[\mathbf{G}] \rightarrow \mathbf{Prob}[\mathbf{G}].$$

Author Contributions: Investigation, R.M. and J.W.P. All authors have read and agreed to the published version of the manuscript.

Funding: Open Access funding was provided by the Max Planck Society.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Matveev, R.; Portegies, J.W. Asymptotic dependency structure of multiple signals. *Inf. Geom.* **2018**, *1*, 237–285. [[CrossRef](#)]
2. Matveev, R.; Portegies, J.W. Tropical diagrams of probability spaces. *Inf. Geom.* **2020**, *3*, 61–88. [[CrossRef](#)]
3. Ay, N.; Bertschinger, N.; Der, R.; Güttler, F.; Olbrich, E. Predictive information and explorative behavior of autonomous robots. *Eur. Phys. J. B* **2008**, *63*, 329–339. [[CrossRef](#)]
4. Bertschinger, N.; Rauh, J.; Olbrich, E.; Jost, J.; Ay, N. Quantifying unique information. *Entropy* **2014**, *16*, 2161–2183. [[CrossRef](#)]
5. Friston, K. The free-energy principle: A rough guide to the brain? *Trends Cogn. Sci.* **2009**, *13*, 293–301. [[CrossRef](#)] [[PubMed](#)]
6. Kingma, D.P.; Welling, M. Auto-encoding variational Bayes. *arXiv* **2013**, arXiv:1312.6114.
7. Steudel, B.; Ay, N. Information-theoretic inference of common ancestors. *Entropy* **2015**, *17*, 2304–2327. [[CrossRef](#)]
8. Dijk, S.G.V.; Polani, D. Informational constraints-driven organization in goal-directed behavior. *Adv. Complex Syst.* **2013**, *16*, 1350016. [[CrossRef](#)]
9. Matveev, R.; Portegies, J.W. Tropical probability theory and an application to the entropic cone. *Kybernetika* **2020**, *56*, 1133–1153. [[CrossRef](#)]
10. Kovačević, M.; Stanojević, I.; Šenk, V. On the hardness of entropy minimization and related problems. In Proceedings of the 2012 IEEE Information Theory Workshop, Lausanne, Switzerland, 3–7 September 2012; pp. 512–516.
11. Vidyasagar, M. A metric between probability distributions on finite sets of different cardinalities and applications to order reduction. *IEEE Trans. Autom. Control* **2012**, *57*, 2464–2477. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.