A Kolmogorov zero to one law for Markov Categories

Eigil Fjeldgren Rischel

1 Introduction

This is a treatment of Kolmogorov's zero to one law in the setting of Markov categories. Throughout, J is a possibly infinite set and $F \subset J$ is a finite subset.

2 Definitions and preliminaries

Suppose $\mathcal{C}$ is a Markov category. To make sense of infinite families of random variables, we will want to have a notion of infinite tensor product. Of course, there is no reason such a thing should exist - for instance, it would be very strange for $\mathsf{FinStoch}$ to have infinite tensor products. But in cases where the subcategory $\mathcal{C}_{det}$ of deterministic maps has infinite products as well as finite ones, it seems natural to ask whether this is compatible with the structure of $\mathcal{C}$ in a suitable sense. We introduce the following notion

Definition 2.1. Let $\{X_i\}_{i \in J}$ be a collection of objects in $\mathcal{C}$. Then we may form the diagram

$$F \mapsto \bigotimes_{i \in F} X_i$$

indexed over the poset of finite subsets $F \subset J$, ordered by reverse inclusion, where the structure maps are induced by $\mathrm{del}_{X_i} : X_i \to I$ and the coherence isomorphisms.

We define $\bigotimes_{i \in J} X_i$ to be the limit of this (cofiltered) diagram in $\mathcal{C}$, if it exists.

Remark 2.2. If J is finite, this coincides with the normal tensor product indexed over J.

Remark 2.3. It is easily verified that

$$\bigotimes_{i \in J} X_i \otimes \bigotimes_{i' \in J'} X_{i'} \cong \bigotimes_{i \in J \sqcup J'} X_i$$

i.e tensor products of infinite tensor products can be written as single infinite tensor products in the obvious way.

Probably the similar theorem about infinite tensor products of infinite tensor products is true as well, although I have not verified it.

Remark 2.4. Supposing $\bigotimes_{i\in J} X_i$ exists, the structure morphisms $\pi_j : \bigotimes_{i\in J} X_i \to X_j$ are deterministic and exhibit it as the product in $\mathcal{C}_{det}$ of $\{X_i\}$.

Remark 2.5. If we take as input certain maps $I \to X_i$ as well, so that we have an infinite system of probability distributions, we can also try to define the infinite tensor product as a colimit, which will then be a space with a probability distribution, corresponding to an infinite system of random variables. It would also be possible to formulate the zero-to-one law in this setting, but I have not explored this.

To actually use this idea to reproduce the usual Kolmogorov zero-one law, we need to do a bit of work:

Example 2.6. Let $\{(X_i, \Sigma_i)\}_{i\in J}$ be a set of measurable spaces in Stoch. Then consider their product $X = \prod_i X_i$ with the σ-algebra generated by the projections $\pi_i : X \to X_i$, which we denote Σ. Let us write in general $X_F = \prod_{i\in F} X_F$, Σ_F for the corresponding σ-algebra, and $\pi_F : X \to X_F$ for the projection.

Let $\{p_F : A \times \Sigma_F \to [0,1]\}_{F\subset J}$ be a collection of Markov kernels $A \to \prod_{i\in F} X_i$ for each F, compatible in the suitable sense. Consider a point $a \in A$ and a generator $\pi_F^{-1}(B)$ for Σ, where $B \in \Sigma_F$. Then we define $p(a, \pi_F^{-1}(B)) = p_F(a, B)$. Note that the collection of such $\pi_F^{-1}(B)$ is a ring of sets on X which generates Σ - hence by Caratheodory extension, there is a unique probability measure $p(a, -)$ on X extending it. This is how we define $p(a, A)$ for general subsets $A \subset X$.

Now consider the class of subsets $A \in \Sigma$ such that $p(-, A) : A \to X$ is measurable. This class is a σ-algebra, which contains $\pi_F^{-1}(B)$, since $p(-, \pi_F^{-1}(B)) = p_F(-, B)$, which is measurable by assumption. Since these generate Σ, the map is always measurable. Hence p is a Markov kernel.

This argument also shows that this Markov kernel is determined uniquely by the fact that $p(a, \pi_F^{-1}(B)) = p_F(a, B) = \pi_F \circ p$. This proves that (X, Σ) is the limit of (X_F, Σ_F), so that it is the tensor product $\bigotimes_{i\in J} X_i$.

We will refer to the structure maps $\bigotimes_{i\in J} X_i \to \bigotimes_{i\in F} X_i$ as *finite marginalizations.*

This notion of infinite tensor product describes a notion of infinite collection of random variable where all dependence must be witnessed on some finite subset of the variables.

We can codify this in the following definition

Definition 2.7. Let $p : A \to \bigotimes_{i\in J} X_i$ be a map in $\mathcal{C}$. We say it exhibits the independence of $\{X_i\}$ given A if each finite marginalization $A \to \bigotimes_{i\in F} X_i$ exhibits the independence of $\{X_i\}_{i\in F}$ given A.

3 Theorem

Theorem 3.1. Suppose we are given $p : A \to T \otimes \bigotimes_{i\in J} X_i$ satisfying the following conditions:

1. T is a deterministic function of $\bigotimes_{i \in J} X_i$.
2. For each finite subset $F \subset J$, the marginal $A \to T \otimes \bigotimes_{i \in F} X_i$ displays the independence of T and $\bigotimes_{i \in F} X_i$.

Then the marginal $A \to T$ is deterministic as well.

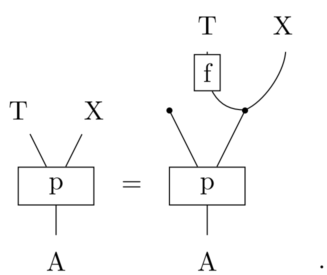
This theorem is a consequence of two lemmas.

Lemma 3.2 (The infinite independence lemma)**.** Suppose $p : A \to \bigotimes X_i$ exhibits the independence of $\{X_i\}$. Then for each j, the map $A \to X_j \otimes \bigotimes_{i \neq j} X_i$ exhibits the independence $X_j \perp \otimes_{i \neq j} X_i \mid\mid A$.

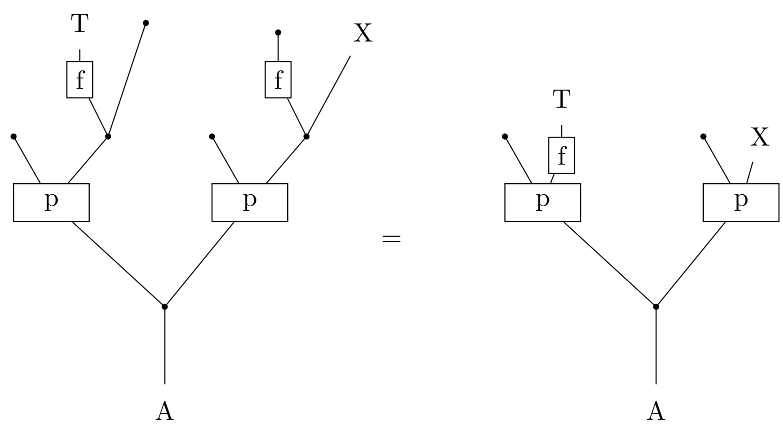
Sketch of proof. To prove this, we must compare two maps $A \to X_j \otimes \bigotimes_{i \neq j} X_i \cong \bigotimes_i X_i$. To show that these maps are equal, it suffices to show that all finite marginalizations of them are equal. But this is an immediate consequence of the assumption. □

Lemma 3.3 (The determinism lemma)**.** Suppose $p : A \to T \otimes X$ is such that T is a deterministic function of X, and p exhibits the independence $T \perp X \mid\mid A$. Then the marginal $A \to T$ is deterministic.

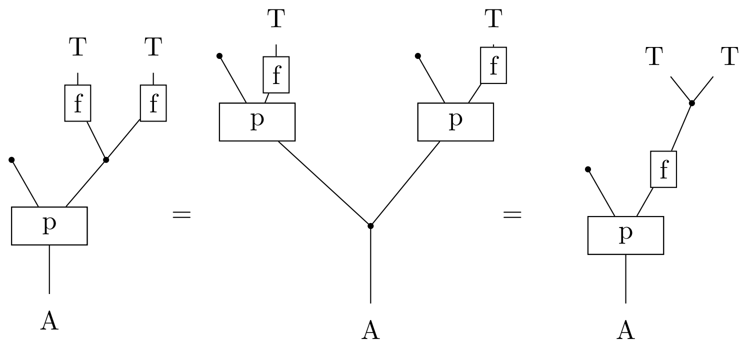
Proof. The proof is a string diagram chase: First, the assumption that T is a deterministic function of X means precisely that we can find f deterministic so that.



Now we can use the independence of X and T, and the axioms, to rewrite this as

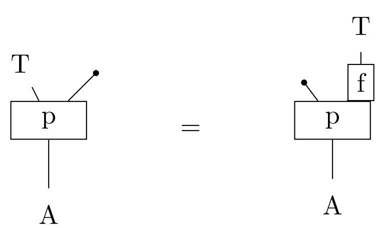


Applying the map $1_T \otimes f$ to the second and fourth diagram now gives an equality



Where we have also applied the determinism of f. This is precisely the statement that $A \to T \otimes X \to X \xrightarrow{f} T$ is deterministic.

Lastly, we use the equality



To see that the marginal $A \to T \otimes X \to T$ is deterministic as desired.

$\square$

Proof of the theorem. It's clear from the definition that $A \to T \otimes \bigotimes X_i$ exhibits the independence of $\{T, X_1, \dots\}$. By the infinite independence lemma, this means that it also exhibits the independence $T \perp \bigotimes X_i \mid\mid A$. Now the determinism lemma implies exactly what we want, that the marginal $A \to T$ is deterministic. $\square$

Corollary 3.4 (Kolmogorov zero to one law)**.** Suppose Ω is a measure space with a probability measure P, $\{f_i : \Omega \to X_i\}$ is a collection of independent random variables, and $T \subset \Omega$ is a subset in the σ-algebra generated by the f_i, such that 1_T is independent of any finite subset of the f_i. Then $P(T)$ is 0 or 1.

Proof. Consider the composite $I \xrightarrow{P} \Omega \xrightarrow{1_T, \{f_i\}} \{0,1\} \otimes \bigotimes_i X_i$. Then independence of the f_i means precisely that this map exhibits the independence of $\{X_i\}$ given I in the above sense. Moreover, $T(\omega)$ is determined by the values of $f_i(\omega)$, so T factors as a map $\prod_i X_i \to \{0,1\}$, and this map is measurable. Hence we can apply the theorem, and conclude that the map $I \xrightarrow{P} \Omega \xrightarrow{T} \{0,1\}$ is deterministic - but this means it's just a constant map, which precisely means that T is true or false with probability 1. □

We can also prove a version of the Hewitt-Savage theorem

Definition 3.5. Let $\alpha : J \to J$ be any map. Then it induces a map

$$]\hat{\alpha} : \bigotimes_{i \in J} X_i \to \bigotimes_{i \in J} X_{\alpha(i)},$$

which we may suggestively write as $(x_i)_{i \in J} \mapsto (x_{\alpha(i)})_{i \in J}$.

It is clear that this map is always deterministic. Moreover, if α is a bijection, this map is an isomorphism.

Theorem 3.6 (Hewitt-Savage for Markov categories)**.** Let $\mathcal{C}$ be a causal Markov category with infinite tensor products. Let $A \to \mathcal{C}$ be a map in $\mathcal{C}$, J an infinite set, and let $f : \bigotimes_{i \in J} X \to T$ be a deterministic map. Suppose for each finite permutation $\sigma : J \to J$, we have $f \circ \hat{\sigma} = f$ - in other words, f is independent of finite permutations of the inputs. Then the composite $A \to \otimes_J A \to \otimes_J X \to T$ is deterministic.

Again, we factor part of the proof into a lemma.

Lemma 3.7. Let $\mathcal{C}$ be a causal markov category, and let $f, g : B \to X$ be maps, f deterministic. Let $p : A \to B$ be any map. Suppose the composites

$$A \to B \to B \otimes B \xrightarrow{f \otimes g} X \otimes X$$

and

$$A \to B \to B \otimes B \xrightarrow{f \otimes f} X \otimes X$$

agree. Then $f = g$ p-almost everywhere.

Sketch of sketch of proof. Consider the map $A \xrightarrow{p} B \to B \otimes B \xrightarrow{f \otimes g} X \otimes X \xrightarrow{copy_{X \otimes X}} X \otimes X \otimes X \otimes X$. Whether we marginalize on the first or second factor of X, we get equal morphisms, because we may replace the first part of this with $A \to B \xrightarrow{f} B \to X \otimes X \otimes X \otimes X$ by using the assumptions and determinism.

Then applying causality, with $f = p$, $g = (f \otimes g) \circ copy_B$, $h_1 = (del_X \otimes 1_X)$, $h_2 = (1_X \otimes del_X)$. Marginalizing the resulting (equal) diagrams, one of them is

$$A \to B \to B \otimes B \xrightarrow{f, 1_B} B \otimes B,$$

one of them is

$$A \to B \to B \otimes B \xrightarrow{g, 1_B} B \otimes B,$$

and this is precisely what must hold for them to agree p-a.e. □

Proof of theorem. Consider the map

$$e : A \to \otimes_J A \overset{\otimes p}{\to} (\otimes_J X) \to (\otimes_J X) \otimes (\otimes_J X) \to T \otimes (\otimes_J X)$$

Denote by e_σ the composite of e with $1_T \otimes \hat{\sigma}$ Note that if σ is a finite permutation, $e_\sigma = e$. To see this, note that we may add a $\hat{\sigma}$ before the f as well, by assumption. Now we can use the determinism of the $\hat{\sigma}$s, and the clear fact that

$$A \to \otimes A \to \otimes B \overset{\hat{\sigma}}{\to} \otimes B = A \to \otimes A \to \otimes B$$

Now let σ instead be an injection. Then for each finite subset $F \subset J$, we can find a finite permutation of J, σ_0, such that $\sigma|_F = \sigma_0|_F$. Then on this marginal, $e_\sigma = e_{\sigma_0} = e$, hence $e_\sigma = e$ in general.

Let $J = J_0 \sqcup J_1$ be a decomposition of J into two disjoint subsets with the cardinality of J (here we use J infinite). Let $\tau_0, \tau_1 : J \to J$ be injections with image J_0, J_1 respectively.

Now consider the map $a_\sigma : A \to T \otimes T$ given by composing e_σ with t. By the above, they all agree. By the lemma, we see that the maps $\otimes_J X \overset{\hat{\sigma}}{\to} \otimes_J X \to T$ are all $\otimes_J p$ a.e. equal. Applying this to τ_0, τ_1, we get that the map $A \to T$ must be deterministic. $\square$

In the last step, the idea is that precomposing with $\hat{\tau_0}$ and $\hat{\tau_1}$ is essentially picking out two disjoint subsets and using them to determine T. Since all the variables are independent, these two are clearly independent, but also equal (by the preceding argument). Hence they are deterministic, but they also agree with the original map.

4 Speculation

It would good to have a way of using this theorem without assuming that the category in question has these infinite tensor products. We can always make sense of the notion of an infinite independent collection of random variables (just a collection of maps $I \to X_i$). Of course to talk about an outcome or a function which depends on the outcome of all the X_i, we need a map $\prod_i X_i \to T$ of some type. But since this map is deterministic, we may want to talk about it without admitting it to our category of Markov kernels.

As an example, the probability theory of finite or countable sets is much simpler than the probability theory of infinite sets, since the whole edifice of σ-algebras can be done away with. This is a reason to prefer $\mathsf{FinStoch}$ over the full Stoch. In this case, it would be desirable to be able to apply this theorem, given a collection of finite variables $I \to X_i$ and a function $\prod_i X_i \to Y$ (which may be required to satisfy some condition like continuity in the product topology), without begin required to invent the larger category Stoch.

For instance, maybe there is a Markov category with objects the profinite sets, which can be obtained from $\mathsf{FinStoch}$ by some formal construction (maybe even simply as the pro-category), which admits infinite tensor products, and where one can profitably interpret this theorem.