

A Formalization of the Ionescu-Tulcea Theorem in mathlib

Author : Etienne Marion

ENS de Lyon, 46, Allée d'Italie, 69007 Lyon, France
etienne.marion@ens-lyon.fr

Abstract

1 Introduction

Being able to talk about the joint distribution of an infinite family of random variables is crucial in probability theory. For example, one often requires a family of independent random variables. The existence of such a family relies on the existence of an infinite product measure. Indeed, given $(\Omega_i, \mathcal{F}_i, \mu_i)_{i \in \iota}$ a family of probability spaces, the existence of the product measure $\bigotimes_{i \in \iota} \mu_i$ yields a new probability space $(\prod_{i \in \iota} \Omega_i, \bigotimes_{i \in \iota} \mathcal{F}_i, \bigotimes_{i \in \iota} \mu_i)$, and the projections $X_i : \prod_{j \in \iota} \Omega_j \rightarrow \Omega_i$ give the desired family. For another example, consider discrete-time Markov chains. Given a measurable space (E, \mathcal{A}) and a Markov kernel $\kappa : E \rightarrow E$, one might want to build a sequence $(X_n)_{n \in \mathbb{N}}$ of random variables with values in E such that the conditional distribution of X_{n+1} given X_0, \dots, X_n is $\kappa(X_n, \cdot)$. Such objects are fundamental in probability theory: families of independent variables allow to build more complicated objects, such as Brownian motion, while discrete-time Markov chains form a huge class of stochastic processes which contains random walks for instance. It so happens that those objects always exist without any restrictions on the spaces we consider. This is a direct consequence of the Ionescu-Tulcea theorem.

The goal of this contribution is to provide a formalization of the proof of the aforementioned theorem using the proof assistant Lean and the associated library `mathlib`. It heavily relies on previous work done in [cite] which formalizes the Kolmogorov extension theorem. While this specific theorem is not used directly in the formalization presented here, we use the notion of measurable cylinders, projective families or additive contents which were introduced to prove the Kolmogorov extension theorem.

2 Statement of the theorem

The Ionescu-Tulcea theorem is about the existence of a certain Markov kernel. We therefore start with a reminder about those.

Definition 2.1. Let (X, \mathcal{A}) and (Y, \mathcal{B}) be two measurable spaces. A *Markov kernel* from X to Y is a map $\kappa : X \times \mathcal{B} \rightarrow [0, 1]$ such that:

- for any $x \in X$, $\kappa(x, \cdot)$ is a probability measure;
- for any $B \in \mathcal{B}$, $\kappa(\cdot, B)$ is measurable.

One can therefore consider a Markov kernel as a measurable map which to any point in X associates a probability distribution over the space Y .

To ease the writing, let us give some notations.

Notation. If I is a set of indices and $(X_i, \mathcal{A}_i)_{i \in I}$ a family of measurable spaces, then:

- we write X for $\prod_{i \in I} X_i$ and \mathcal{A} for $\bigotimes_{i \in I} \mathcal{A}_i$;
- if $J \subseteq I$, we write X^J for $\prod_{i \in J} X_i$ and \mathcal{A}^J for $\bigotimes_{i \in J} \mathcal{A}_i$;
- if $I = \mathbb{N}$ and $n \in \mathbb{N}$ we write $X^{\geq n}$ for $X^{\{k \in \mathbb{N} | k \geq n\}}$ and $\mathcal{A}^{\geq n}$ for $\mathcal{A}^{\{k \in \mathbb{N} | k \geq n\}}$, with similar notations for $\leq, >, <$;
- if $J \subseteq I$, we write π_J for the canonical map from X to X^J ;
- if $K \subseteq J \subseteq I$, we write $\pi_{J \rightarrow K}$ for the canonical map from X^J to X^K .

Let us now state the theorem.

Theorem 2.2 (Ionescu-Tulcea). *Let $(X_n, \mathcal{A}_n)_{n \in \mathbb{N}}$ be a family of measurable spaces. Let $(\kappa_n)_{n \in \mathbb{N}}$ be a family of Markov kernels such that for any n , κ_n is a kernel from $X_0 \times \dots \times X_n$ to X_{n+1} . Then there exists a unique Markov kernel $\eta : X_0 \rightarrow \prod_{n \geq 1} X_n$ such that for any $n \geq 1$,*

$$\pi_{n*} \eta = \kappa_0 \otimes \dots \otimes \kappa_{n-1}.$$

By taking each kernel to be constant we easily get the product of a countable family of probability measures, which can then be used to build the product of an arbitrary family of measures. Likewise if we take $X_n = X$ for all n and $\kappa_n = \kappa$ a kernel from $X \rightarrow X$ we get a Markov kernel from X to $X^{\mathbb{N}^+}$ which, if we compose it with a measure μ at the beginning, gives the measure associated to a Markov chain with initial distribution μ and transition kernel κ .

We will now give a sketch of the proof of the theorem to highlight the elements which are required for the formalization. The idea is to use the Carathéodory extension theorem which allows to extend a measure once we have defined it on a restricted family of measurable sets for which it is to give the expression of the measure. In what follows we consider $(X_n, \mathcal{A}_n)_{n \in \mathbb{N}}$ a family of measurable spaces and we set $\pi_n := \pi_{[1, n]}$. we also consider of Markov kernels $(\kappa_n : X^{\leq n} \rightarrow_* X_{n+1})_{n \in \mathbb{N}}$ and we set $\eta_{k, n} := \kappa_k \otimes \dots \otimes \kappa_{n-1} : X^{\leq k} \rightarrow_* X^{[k+1, n]}$ and $\eta_n := \eta_{0, n}$.

Definition 2.3. Sets of the form $A \times X^{>n}$, where $n \geq 1$ and $A \in \mathcal{A}^{\leq n}$ are called cylinders.

By writing $A \times X^{>n} = \pi_n^{-1}(A)$ it is easy to check that the set of cylinders is closed by union and complement, and thus it is an algebra of sets which generates the product σ -algebra. We now have to define a measure over cylinders which we will then extend to the whole σ -algebra. We also do this for cylinders in $X^{\geq n}$ for $n \geq 1$ as we will need it for the extension afterwards.

Definition 2.4. Let $x_0 \in X_0$. If $A := A' \times X^{>n} \subseteq X^{\geq 1}$ is a cylinder we set

$$P_0(x_0, A) := \eta_n(x_0, A').$$

This map is well defined because if $n \leq m$ then

$$\begin{aligned}
\eta_m(x_0, A' \times X^{\llbracket n+1, m \rrbracket}) &= \eta_n \otimes \eta_{n+1, m}(x_0, A' \times X^{\llbracket n+1, m \rrbracket}) \\
&= \int_{A'} \eta_{n+1, m}(x_0, \dots, x_n, X^{\llbracket n+1, m \rrbracket}) \eta_n(x_0, dx_1 \dots dx_n) \\
&= \int \mathbb{1}_{A'} \eta_n(x_0, dx_1 \dots dx_n) \\
&= \eta_n(x_0, A').
\end{aligned}$$

For $k \geq 1$ and $A := A' \times X^{>n} \subseteq X^{>k}$, we define likewise

$$P_k(x_0, \dots, x_k, A) := \eta_{k, n}(x_0, \dots, x_k, A').$$

Lemma 2.5. *For any $k \in \mathbb{N}$ and any $(x_0, \dots, x_k) \in X^{\leq k}$, the map $P_k(x_0, \dots, x_k, \cdot)$ is an additive content over the cylinders. Namely if A, B are two disjoint cylinders then $P_0(x_0, \dots, x_k, A \cup B) = P_0(x_0, \dots, x_k, A) + P_0(x_0, \dots, x_k, B)$.*

Proof. We only give the proof in the case $k = 0$ as the other cases are identical. Let $A := A' \times X^{>m}$ and $B := B' \times X^{>n}$ be two disjoint cylinders and assume without loss of generality that $m \leq n$. We then set $A'' := A' \times X^{\llbracket m+1, n \rrbracket}$. As $A \cap B = \emptyset$ this implies $A'' \cap B' = \emptyset$. Moreover, $A \cup B = (A'' \cup B') \times X^{>n}$. Consequently :

$$P_0(x_0, A \cup B) = \eta_n(x_0, A'' \cup B') = \eta_n(x_0, A'') + \eta_n(x_0, B') = P_0(x_0, A) + P_0(x_0, B). \quad \square$$

To apply Carathéodory theorem it remains to prove that $P_0(x_0, \cdot)$ is σ -additive over cylinders. Because it is an additive content and cylinders are an algebra of sets it is enough to show the following result:

Lemma 2.6. *Let $x_0 \in X_0$. If $(A_n)_{n \in \mathbb{N}}$ is a non-increasing sequence of cylinders with empty intersection then $P_0(x_0, A_n) \xrightarrow{n \rightarrow \infty} 0$.*

Proof. For any n we write $A_n := B_n \times X^{>a_n}$. In this proof if $A \subseteq X^{\geq 1}$ et $(x_1, \dots, x_n) \in X^{\llbracket 1, n \rrbracket}$ we will denote by $A(x_1, \dots, x_n) := \{y \in X^{>n} : (x_1, \dots, x_n, y) \in A\}$ the associated section of A .

We use contraposition. Assume there exists $\epsilon > 0$ such that for any $n \in \mathbb{N}$, $P_0(x_0, A_n) \geq \epsilon$. We wish to prove that the intersection of A_n is nonempty. We define the following sequence of functions:

$$\begin{aligned}
f_n : X_1 &\rightarrow [0, 1] \\
x_1 &\mapsto P_1(x_0, x_1, A_n(x_1)).
\end{aligned}$$

It allows us to write

$$\begin{aligned}
P_0(x_0, A_n) &= \eta_{a_n}(x_0, B_n) \\
&= \kappa_0 \otimes \eta_{1, a_n}(B_n) \\
&= \int_{X_1} \eta_{1, a_n}(x_0, x_1, B_n(x_1)) \kappa_0(x_0, dx_1) \\
&= \int_{X_1} P_1(x_0, x_1, A_n(x_1)) \kappa_0(x_0, dx_1) \\
&= \int_{X_1} f_n(x_1) \kappa_0(x_0, dx_1).
\end{aligned}$$

Because (A_n) is non-increasing it is also the case of (f_n) . We set $g := \lim_{n \rightarrow \infty} f_n$. By dominated convergence theorem we get $\epsilon \leq P_0(x_0, A_n) \xrightarrow{n \rightarrow \infty} \int_{X_1} g(x_1) \kappa_0(x_0, dx_1)$.

As $\kappa_0(x_0, \cdot)$ is a probability measure there exists $x_1 \in X_1$ such that $g(x_1) \geq \epsilon$, and so for any $n \in \mathbb{N}$, $f_n(x_1) \geq \epsilon$ which implies that $P_1(x_0, x_1, A_n(x_1)) \geq \epsilon$.

We can thus repeat the same step replacing $P_0(x_0, \cdot)$ by $P_1(x_0, x_1, \cdot)$ and A_n by $A_n(x_1)$. We get a sequence $x := (x_k)_{k \geq 1}$ such that for any $n \in \mathbb{N}$ and $k \geq 1$ we have $P_k(x_0, \dots, x_k, A_n(x_1, \dots, x_k)) \geq \epsilon$. It remains to show that $x \in \bigcap_{n \in \mathbb{N}} A_n$.

Let $n \in \mathbb{N}$. As $A_n = B_n \times X^{>a_n}$ we in particular have that $A_n(x_1, \dots, x_{a_n})$ is either $X^{>a_n}$ or empty according to whether (x_1, \dots, x_{a_n}) belongs to B_n or not. But as $P_{a_n}(x_0, \dots, x_{a_n}, A_n(x_1, \dots, x_{a_n})) \geq \epsilon > 0$ we have $A_n(x_1, \dots, x_{a_n}) = X^{>a_n}$, which proves that $x \in A_n$. \square

This concludes the main argument of the proof. The measurability is easily provided by applying the π - λ theorem. We will now focus on those aspects of the proof which proved difficult to formalize and how they were overcome.

3 Dealing with composition of kernels

One of the key tools for the Ionescu-Tulcea theorem is to be able to speak about the composition of Markov kernels. Indeed that is necessary even to be able to state the theorem. However formalizing this in a way which allows us to smoothly write products proved to be quite challenging.

One common rule in formalization is that if we use "..." in a proof it will likely be hard to formalize. In our case the statement of Theorem 2.2 itself makes use of these when we write the composition of kernels, and they hide a real problem.

Transition kernels and their composition are already defined in Lean. However we need to take the composition of a finite family of kernels which does not exist in Mathlib, and here is why. By hand, if we have three Markov kernels $\kappa : X \rightarrow_* Y$, $\eta : X \times Y \rightarrow_* Z$ and $\xi : X \times Y \times Z \rightarrow_* T$ we can define $\kappa \otimes \eta \otimes \xi := (\kappa \otimes \eta) \otimes \xi$. Then we easily get $\kappa \otimes \eta \otimes \xi = \kappa \otimes (\eta \otimes \xi)$. But here we forget a detail: $X \times Y \times Z$ does not mean anything a priori. We can write $(X \times Y) \times Z$ or $X \times (Y \times Z)$ but those are two different types. In paper maths they are canonically isomorphic and we often act as though they were the same but in Lean this isomorphism has to be made explicit.

Coming back to kernels I can write that ξ is a kernel from $X \times (Y \times Z)$ to T . Then $(\kappa \otimes \eta) \otimes \xi$ is a kernel from X to $(Y \times Z) \times T$. But in that case $\kappa \otimes (\eta \otimes \xi)$ is not even well defined because ξ should start from $(X \times Y) \times Z$. If we were to add an isomorphism to define this object anyway the equality $(\kappa \otimes \eta) \otimes \xi = \kappa \otimes (\eta \otimes \xi)$ would still make no sense because on the left we have a kernel from X to $(Y \times Z) \times T$ while on the right it is a kernel from X to $Y \times (Z \times T)$, so they do not have the same type.

A similar issue arises when formalizing Theorem 2.2: if we have two kernels $\kappa : X_0 \rightarrow_* X^{[1,n]}$ and $\eta : X^{\leq n} \rightarrow_* X^{[n+1,m]}$, their composition does not make sense, we need isomorphisms. First to see η as a kernel from $X_0 \times X^{[1,n]}$ to $X^{[n+1,m]}$ to get $\kappa \otimes \eta : X_0 \rightarrow_* X^{[1,n]} \times X^{[n+1,m]}$, then to see $\kappa \otimes \eta$ as a kernel from X_0 to $X^{[1,m]}$. To solve this problem we need to make these isomorphisms explicit and then hide them inside a black box to be able to use it without worrying about the implementation.

This work was initiated by Degenne. He introduced a composition of kernels indexed by integer intervals which makes the equivalence invisible.

```
def compProdNat {i j k : ℕ} (κ : Kernel ((l : Iic i) → X l))
  ((l : Ioc i j) → X l))
  (η : Kernel ((l : Iic j) → X l) ((l : Ioc j k) → X l)) :
  Kernel ((l : Iic i) → X l) ((l : Ioc i k) → X l)
```

This takes two kernels $\kappa : X^{\leq i} \rightarrow_* X^{\llbracket i+1, j \rrbracket}$ and $\eta : X^{\leq j} \rightarrow_* X^{\llbracket j+1, k \rrbracket}$ and returns a kernel from $X^{\leq i}$ to $X^{\llbracket i+1, k \rrbracket}$. Degenne then proves that this composition is associative, and uses it to define

```
def kerNat ( $\kappa : (k : \mathbb{N}) \rightarrow \text{Kernel } ((l : \text{Iic } k) \rightarrow X \ 1) (X \ (k + 1))$ )
  ( $i \ j : \mathbb{N}$ ) :
  Kernel (( $l : \text{Iic } i$ )  $\rightarrow X \ 1$ ) (( $l : \text{Ioc } i \ j$ )  $\rightarrow X \ 1$ ).
```

This takes as an argument a family of kernels $(\kappa_k)_{k \in \mathbb{N}}$ with $\kappa_k : X^{\leq k} \rightarrow X_{k+1}$ just as in Theorem 2.2. Then given two integers i and j it returns a kernel `kerNat κ i j` which is the composition

$$\kappa_i \otimes \kappa_{i+1} \otimes \cdots \otimes \kappa_{j-1}.$$