We work with a background probability space $(\Omega, \Sigma_\Omega, \mathbb{P})$. For a measurable space $(\mathcal{X}, \Sigma_\mathcal{X})$ we denote the set of probability measures on this space $\mathcal{P}(\Sigma_\mathcal{X})$ or simply $\mathcal{P}(\mathcal{X})$ when the σ-algebra is unambiguous. When taking cartesian products $\mathcal{X} \times \mathcal{Y}$ of measurable spaces $(\mathcal{X}, \Sigma_\mathcal{X}), (\mathcal{Y}, \Sigma_\mathcal{Y})$ we always endow such with the product σ-algebra $\Sigma_\mathcal{X} \otimes \Sigma_\mathcal{Y}$, unless otherwise specified. A map $f : \mathcal{X} \to \mathcal{Y}$ is called $\Sigma_\mathcal{X}$-$\Sigma_\mathcal{Y}$ measurable provided $f^{-1}(\Sigma_\mathcal{Y}) \subseteq \Sigma_\mathcal{X}$ and we denote the set of such functions $\mathcal{M}(\Sigma_\mathcal{X}, \Sigma_\mathcal{Y})$. By a random variable X on $(\mathcal{X}, \Sigma_\mathcal{X})$ mean a Σ_Ω-$\Sigma_\mathcal{X}$ measurable map.

Definition 1 (Probability kernel)**.** Let $(\mathcal{X}, \Sigma_\mathcal{X}), (Y, \Sigma_\mathcal{Y})$ be measurable spaces. A function

$$\kappa(\cdot \mid \cdot) : \Sigma_\mathcal{Y} \times \mathcal{X} \to [0, 1]$$

is a $(\mathcal{X}, \Sigma_\mathcal{X})$-**probability kernel** on $(\mathcal{Y}, \Sigma_\mathcal{Y})$ provided

1. $B \mapsto \kappa(B \mid x) \in \mathcal{P}(\Sigma_\mathcal{Y})$ that is $\kappa(\cdot \mid x)$ is a probability measure for any $x \in \mathcal{X}$.
2. $x \mapsto \kappa(B \mid x) \in \mathcal{M}(\Sigma_\mathcal{X}, \Sigma_\mathcal{Y})$ that is $\kappa(B \mid \cdot)$ is $(\Sigma_\mathcal{X}$-$\Sigma_\mathcal{Y})$ measurable for any $B \in \Sigma_\mathcal{Y}$.

When the σ-algebras are unambiguous we shall simply say an $\mathcal{X} \rightsquigarrow \mathcal{Y}$ kernel. For any $x \in \mathcal{X}$ and $f \in \mathcal{L}_1(\kappa(\cdot \mid x))$ we write the integral of f over $\kappa(\cdot \mid x)$ as $\int f(y) \mathrm{d}\kappa(y \mid x)$.

Definition 2 (Dynamic programming model)**.** A general dynamic programming model is determined by

1. $(\mathcal{S}_n, \Sigma_{\mathcal{S}_n})_{n \in \mathbb{N}}$ a measurable space of **states** for each timestep.
2. $(\mathcal{A}_n, \Sigma_{\mathcal{A}_n})_{n \in \mathbb{N}}$ a measurable space of **actions** for each timestep.

for each $n \in \mathbb{N}$ write $\mathcal{H}_n = \mathcal{S}_1 \times \mathcal{A}_1 \times \cdots \times \mathcal{S}_n$, $\mathcal{H}_\infty = \mathcal{S}_1 \times \mathcal{A}_1 \times \dots$, with associated σ-algebras $\Sigma_{\mathcal{H}_n} := \left(\bigotimes_{i=1}^{n-1} (\Sigma_{\mathcal{S}_i} \otimes \Sigma_{\mathcal{A}_i}) \right) \otimes \Sigma_{\mathcal{S}_n}$ and $\Sigma_{\mathcal{H}_\infty} := \bigotimes_{i=1}^{\infty} (\Sigma_{\mathcal{S}_i} \otimes \Sigma_{\mathcal{A}_i})$. These are called the **history** spaces.

3. $(P_n)_{n \in \mathbb{N}}$ a sequence of $\mathcal{H}_n \times \mathcal{A}_n \rightsquigarrow \mathcal{S}_{n+1}$ kernels called the **transition** kernels.
4. $(R_n)_{n \in \mathbb{N}}$ a sequence of $\mathcal{H}_{n+1} \rightsquigarrow \mathbb{R}$ kernels called the **reward** kernels.

For such a model we can define

Definition 3 (Policy)**.** A (randomized) **policy** $\pi = (\pi_n)_{n \in \mathbb{N}}$ is a sequence of $\mathcal{H}_n \rightsquigarrow \mathcal{A}_n$ kernels. The set of all policies we denote $R\Pi$.

We know state some fundamental results on probability kernels

Theorem 1 (Integration of a kernel)**.** Let $\mu \in \mathcal{P}(\mathcal{X})$ and $\kappa : \mathcal{X} \rightsquigarrow \mathcal{Y}$. Then there exists a uniquely determined probability measure $\lambda \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ such that

$$\lambda(A \times B) = \int_A \kappa(B, x) \mathrm{d}\mu(x)$$

We denote this $\lambda = \kappa(\cdot \mid \mu)$.

Proof. We refer to [ref to EH markov, thm. 1.2.1]. □

For an idea how to actually compute integrals with kernel derived measures we here include

Theorem 2 (Extended Tonelli and Fubini)**.** Let $\mu \in \mathcal{P}(\mathcal{X})$, $f \in \mathcal{M}(\Sigma_\mathcal{X} \otimes \Sigma_\mathcal{Y}, \mathbb{B})$ be a measurable function and $\kappa : \mathcal{X} \rightsquigarrow \mathcal{Y}$ be a probability kernel. Then

$$\int |f| \, \mathrm{d}\kappa(\cdot \mid \mu) = \int \int |f| \, \mathrm{d}\kappa(\cdot \mid x) \mathrm{d}\mu(x)$$

Furthermore if this is finite, i.e. $f \in \mathcal{L}_1(\kappa(\cdot, \mu))$ then $A_0 := \left\{ x \in \mathcal{X} \;\middle|\; \int \mathrm{d}\kappa(\cdot \mid x) < \infty \right\} \in \Sigma_\mathcal{X}$ with $\mu(A_0) = 1$,

$$x \mapsto \begin{cases} \int f \mathrm{d}\kappa(\cdot \mid x) & x \in A_0 \\ 0 & x \notin A_0 \end{cases}$$

is $\Sigma_\mathcal{X}$-$\mathbb{B}$ measurable and

$$\int f \mathrm{d}\kappa(\cdot \mid \mu) = \int_{A_0} \int f \mathrm{d}\kappa(\cdot \mid x) \mathrm{d}\mu(x)$$

Proof. We refer to [ref to EH markov, thm. 1.3.2 + 1.3.3] $\square$

Proposition 1 (Composition of kernels)**.** Let $\kappa : \mathcal{X} \rightsquigarrow \mathcal{Y}, \psi : \mathcal{Y} \rightsquigarrow \mathcal{Z}$ be probability kernels. Then

$$(\psi * \kappa)(A \mid x) := \int \kappa(A \mid y) \mathrm{d}\psi(\cdot \mid x), \qquad \forall A \in \Sigma_{\mathcal{Z}}, x \in \mathcal{X}$$

is a $\mathcal{X} \rightsquigarrow \mathcal{Z}$ probability kernel called the composition of κ and ψ. The composition operator $*$ is associative, i.e. if $\phi : \mathcal{Z} \rightsquigarrow \mathcal{W}$ is a third probability kernel then $(\phi * \psi) * \kappa = \phi * (\psi * \kappa)$. The associativity also extends to measures, i.e. $\forall \mu \in \mathcal{X} : (\psi * \kappa)(\mu) = \psi(\kappa(\cdot \mid \mu))$ and this is uniquely determined by ψ, κ and μ.

Proof. The first assertion is a trivial verification of the two conditions in definition 1 and left as an exercise. For the associativity we refer to [todo ref to EH markov, lem. 4.5.4]. $\square$

Proposition 2 (Existence and uniqueness of finite kernel processes)**.** Let $(\mathcal{X}_i, \Sigma_{\mathcal{X}_i})_{i \in \mathbb{N}}$ be a sequence of measurable spaces. For each $i \in \mathbb{N}$ define $\mathcal{X}^{\underline{i}} := \mathcal{X}_1 \times \cdots \times \mathcal{X}_i$ and let $\kappa_i : \mathcal{X}^{\underline{i}} \rightsquigarrow \mathcal{X}_{i+1}$ be a probability kernel. Given a probability measure $\rho_1 \in \mathcal{P}(\mathcal{X}_1)$ there exists for every $n \in \mathbb{N}$ a unique probability measure ρ_n on $\mathcal{X}^{\underline{n}}$ defined by

$$\rho_s := (\kappa_{n-1} * \cdots * \kappa_1)(\rho_1)$$

Proof. This follows simply by induction using proposition 1. $\square$

Proposition 3 (Existence and uniqueness of finite policy generated processes)**.** For every policy $(\pi_n)_{n \in \mathbb{N}}$ and probability measure $\rho_1 \in \mathcal{P}(S_1)$ there exists a unique probability measure $\rho_n \in \mathcal{P}(\mathcal{H}_{n+1})$ for every $n \in \mathbb{N}$ such that $\rho_n = (P_n * \pi_n)(\rho_{n-1})$.

Proof. This is directly from proposition 2 with $\kappa_1 = \pi_1 * P_1, \kappa_2 = \pi_2 * P_2 \ldots$ $\square$

Proposition 3 is not enough to establish existence of a policy generated measure on $(\mathcal{H}_\infty, \Sigma_{\mathcal{H}_\infty})$ which we will need later. This problem was solved by Cassius Ionescu-Tulcea in 1949:

Theorem 3 (Ionescu-Tulcea extension theorem)**.** Let $(\mathcal{X}_1, \Sigma_{\mathcal{X}_1}, \rho_1)$ be a probability space and $(\mathcal{X}_i, \Sigma_{\mathcal{X}_i})_{i \in \mathbb{N}}$ be a sequence of measurable spaces. Define $\mathcal{X}^{\underline{i}} := \mathcal{X}_1 \times \cdots \times \mathcal{X}_i$ and $\mathcal{X}^{\underline{\infty}} := \prod_{i \in \mathbb{N}} \mathcal{X}_i$. For all $i \in \mathbb{N}$ let $(\kappa_i) : \mathcal{X}^{\underline{i}} \rightsquigarrow \mathcal{X}_{i+1}$ be a probability kernel and $\rho_i = (\kappa_{i-1} * \cdots * \kappa_1)(\rho_1)$. Then there exists a unique probability measure $\rho \in \mathcal{P}(\mathcal{X}^{\underline{\infty}})$ such that

$$\rho_i(A) = \rho\left(A \times \prod_{k=i+1}^{\infty} \mathcal{X}_k\right)$$

for all $i \in \mathbb{N}$.

Proof. Todo: what about this. $\square$

Corollary 1. A policy $(\pi_n)_{n \in \mathbb{N}}$ and a probability measure $\rho_1 \in \mathcal{P}(S_1)$ determines a unique probability measure $\rho \in \mathcal{H}_\infty$.