Markov Processes as Function Transformers

Part II: Functorial View of Expectation Values

Prakash Panangaden[1]

[1]School of Computer Science
McGill University

CIRM Marseille 20 - 24 February 2012

Outline

Approximation via Averaging

1. Approximation of Markov processes should be based on "averaging".
2. Averages are computed by expectation values.
3. Beautiful functorial presentation of expectation values d'après Vincent Danos.
4. Make bisimulation and approximation live in the same universe

Duality is the Key

$$
\begin{tikzcd}
\mathcal{M}^{\ll p}(X) \arrow[r, leftrightarrow, "\sim"] \arrow[d, leftrightarrow, dotted] & L_1^+(X,p) \arrow[r, leftrightarrow, "\sim"] \arrow[d, leftrightarrow, dotted] & L_\infty^{+,*}(X,p) \arrow[d, leftrightarrow, dotted] \\
\mathcal{M}^p_{\mathsf{UB}} \arrow[r, leftrightarrow, "\sim"] & L_\infty^+(X,p) \arrow[r, leftrightarrow, "\sim"] & L_1^{+,*}(X,p)
\end{tikzcd} \tag{1}
$$

where the vertical arrows represent dualities and the horizontal arrows represent isomorphisms.

Pairing function

There is a map from the product of the cones $L_\infty^+(X,p)$ and $L_1^+(X,p)$ to $\mathbb{R}^+$ defined as follows:

$$\forall f \in L_\infty^+(X,p), g \in L_1^+(X,p) \quad \langle f,\ g\rangle = \int fg\mathrm{d}p.$$

Some notation

1. Given (X, Σ, p) and (Y, Λ) and a measurable function $f : X \longrightarrow Y$ we obtain a measure q on Y by $q(B) = p(f^{-1}(B))$. This is written $M_f(p)$ and is called the *image measure* of p under f.
2. We say that a measure ν is **absolutely continuous** with respect to another measure μ if for any measurable set A, $\mu(A) = 0$ implies that $\nu(A) = 0$. We write $\nu \ll \mu$.

The Radon-Nikodym Theorem

The Radon-Nikodym theorem is a central result in measure theory allowing one to define a "derivative" of a measure with respect to another measure.

Radon-Nikodym

If $\nu \ll \mu$, where ν, μ are finite measures on a measurable space (X, Σ) there is a positive measurable function h on X such that for every measurable set B

$$\nu(B) = \int_B h \, \mathrm{d}\mu.$$

The function h is defined uniquely up to a set of μ-measure 0. The function h is called the Radon-Nikodym derivative of ν with respect to μ; we denote it by $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$. Since ν is finite, $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} \in L_1^+(X, \mu)$.

Notation for Radon-Nikodym

1. Given an (almost-everywhere) positive function $f \in L_1(X, p)$, we let $f \cdot p$ be the measure which has density f with respect to p.
2. Two identities that we get from the Radon-Nikodym theorem are:
   - given $q \ll p$, we have $\frac{\mathrm{d}q}{\mathrm{d}p} \cdot p = q$.
   - given $f \in L_1^+(X, p)$, $\frac{\mathrm{d}f \cdot p}{\mathrm{d}p} = f$
3. These two identities just say that the operations $(-) \cdot p$ and $\frac{\mathrm{d}(-)}{\mathrm{d}p}$ are inverses of each other as maps between $L_1^+(X, p)$ and $\mathcal{M}^{\ll p}(X)$ the space of finite measures on X that are absolutely continuous with respect to p.

Expectation and conditional expectation

1. The expectation $\mathbb{E}_p(f)$ of a measurable function f is the average computed by $\int f \mathrm{d}p$ and therefore it is just a number.
2. The *conditional* expectation is not a mere number but a random variable.
3. It is meant to measure the expected value in the presence of additional information.
4. The additional information takes the form of a sub-σ algebra, say Λ, of Σ. The experimenter knows, for every $B \in \Lambda$, whether the outcome is in B or not.
5. Now she can recompute the expectation values given this information.

Formalizing conditional expectation

- It is an immediate consequence of the Radon-Nikodym theorem that such conditional expectations exist.

Kolmogorov

Let (X, Σ, p) be a measure space with p a finite measure, f be in $L_1(X, \Sigma, p)$ and Λ be a sub-σ-algebra of Σ, then there exists a $g \in L_1(X, \Lambda, p)$ such that for all $B \in \Lambda$

$$\int_B f \mathrm{d}p = \int_B g \mathrm{d}p.$$

- This function g is usually denoted by $\mathbb{E}(f|\Lambda)$.
- We clearly have $f \cdot p \ll p$ so the required g is simply $\frac{\mathrm{d}f \cdot p}{\mathrm{d}p|_\Lambda}$, where $p\mid_\Lambda$ is the restriction of p to the sub-σ-algebra Λ.

Properties of conditional expectation

1. The point of requiring Λ-measurability is that it "smooths out" variations that are too rapid to show up in Λ.
2. The conditional expectation is *linear*, *increasing* with respect to the pointwise order.

4. It is defined uniquely p-almost everywhere.

Where the action happens

- We define two categories $\mathbf{Rad}_\infty$ and $\mathbf{Rad}_1$ that will be needed for the functorial definition of conditional expectation.
- This will allow for L_∞ and L_1 versions of the theory.
- Going between these versions by duality will be very useful.

The "infinity" category

$\mathbf{Rad}_\infty$

The category $\mathbf{Rad}_\infty$ has as objects probability spaces, and as arrows $\alpha : (X, p) \longrightarrow (Y, q)$, measurable maps such that $M_\alpha(p) \leq Kq$ for some real number K.

The reason for choosing the name $\mathbf{Rad}_\infty$ is that $\alpha \in \mathbf{Rad}_\infty$ maps to $d/dqM_\alpha(p) \in L^+_\infty(Y, q)$.

The "one" category

$\mathbf{Rad}_1$

The category $\mathbf{Rad}_1$ has as objects probability spaces and as arrows $\alpha : (X, p) \longrightarrow (Y, q)$, measurable maps such that $M_\alpha(p) \ll q$.

1. The reason for choosing the name $\mathbf{Rad}_1$ is that $\alpha \in \mathbf{Rad}_1$ maps to $d/dqM_\alpha(p) \in L_1^+(Y, q)$.
2. The fact that the category $\mathbf{Rad}_\infty$ embeds in $\mathbf{Rad}_1$ reflects the fact that L_∞^+ embeds in L_1^+.

Pairing function revisited

Recall the isomorphism between $L_\infty^+(X,p)$ and $L_1^{+,*}(X,p)$ mediated by the pairing function:

$$f \in L_\infty^+(X,p) \mapsto \lambda g : L_1^+(X,p).\langle f,\ g\rangle = \int fg\mathrm{d}p.$$

Precomposition

1. Now, precomposition with α in $\mathbf{Rad}_\infty$ gives a map $P_1(\alpha)$ from $L_1^+(Y,q)$ to $L_1^+(X,p)$.
2. Dually, given $\alpha \in \mathbf{Rad}_1 : (X,p) \longrightarrow (Y,q)$ and $g \in L_\infty^+(Y,q)$ we have that $P_\infty(\alpha)(g) \in L_\infty^+(X,p)$.
3. Thus the subscripts on the two precomposition functors describe the *target* categories.
4. Using the $*$-functor we get a map $(P_1(\alpha))^*$ from $L_1^{+,*}(X,p)$ to $L_1^{+,*}(Y,q)$ in the first case and
5. dually we get $(P_\infty(\alpha))^*$ from $L_\infty^{+,*}(X,p)$ to $L_\infty^{+,*}(Y,q)$.

Expectation value functor

- The **functor** $\mathbb{E}_\infty(\cdot)$ is a functor from $\mathbf{Rad}_\infty$ to ω**CC** which, on objects, maps (X, p) to $L^+_\infty(X, p)$ and on maps is given as follows:
- Given $\alpha : (X, p) \longrightarrow (Y, q)$ in $\mathbf{Rad}_\infty$ the action of the functor is to produce the map $\mathbb{E}_\infty(\alpha) : L^+_\infty(X, p) \longrightarrow L^+_\infty(Y, q)$ obtained by composing $(P_1(\alpha))^*$ with the isomorphisms between $L^{+,*}_1$ and L^+_∞

$$\begin{tikzcd}
L^{+,*}_1(X, p) \arrow[d, "(P_1(\alpha))^*"'] & L^+_\infty(X, p) \arrow[l, dotted] \arrow[d, "\mathbb{E}_\infty(\alpha)"] \\
L^{+,*}_1(Y, q) \arrow[r, dotted] & L^+_\infty(Y, q)
\end{tikzcd}$$

Consequences

1. It is an immediate consequence of the definitions that for any $f \in L_\infty^+(X,p)$ and $g \in L_1(Y,q)$

$$\langle \mathbb{E}_\infty(\alpha)(f), g \rangle_Y = \langle f, P_1(\alpha)(g) \rangle_X.$$

$$\begin{tikzcd}
\lambda h : L_1^+(X,p).\langle f,\ h\rangle \arrow[d, mapsto] & f \arrow[l, mapsto] \arrow[d, dotted, mapsto] \\
\lambda g : L_1^+(Y,q).\langle f,\ g \circ \alpha\rangle \arrow[r, mapsto] & \mathbb{E}_\infty(\alpha)(f)
\end{tikzcd}$$

2. One can informally view this functor as a "left adjoint" in view of this proposition.
3. Note that since we started with α in $\mathbf{Rad}_\infty$ we get the expectation value as a map between the L_∞^+ cones.

The other expectation value functor

The **functor** $\mathbb{E}_{\mathbf{1}}(\cdot)$ is a functor from $\mathbf{Rad}_1$ to ω**CC** which maps the object (X,p) to $L_1^+(X,p)$ and on maps is given as follows:
Given $\alpha : (X,p) \longrightarrow (Y,q)$ in $\mathbf{Rad}_1$ the action of the functor is to produce the map $\mathbb{E}_1(\alpha) : L_1^+(X,p) \longrightarrow L_1^+(Y,q)$ obtained by composing $(P_\infty(\alpha))^*$ with the isomorphisms between $L_\infty^{+,*}$ and L_1^+ as shown in the diagram below

$$\begin{tikzcd} L_\infty^{+,*}(X,p) \arrow[d, "(P_\infty(\alpha))^*"'] & L_1^+(X,p) \arrow[l, dotted] \arrow[d, "\mathbb{E}_1(\alpha)"] \\ L_\infty^{+,*}(Y,q) \arrow[r, dotted] & L_1^+(Y,q) \end{tikzcd}$$

Another "adjoint"

Once again we have an "adjointness" statement; this time it is a right adjoint.

Right adjoint

Given $f \in L_\infty^+(Y,q)$ and $g \in L_1^+(X,p)$ we have

$$\langle f, \mathbb{E}_1(\alpha)(g)\rangle_Y = \langle P_\infty(\alpha)(f), g\rangle_X.$$

Relating the two expectation value functors

Given $\alpha \in \mathbf{Rad}_\infty[(X,p),(Y,q)]$ we have

$$(a)\ \ \mathbb{E}_1(\alpha)(f \circ \alpha) = \mathbb{E}_\infty(\alpha)(\mathbf{1}_X)f, \qquad \text{for } f \in L_1^+(Y,q) \text{ and}$$
$$(b)\ \mathbb{E}_\infty(\alpha)(f \circ \alpha) = \mathbb{E}_1(\alpha)(\mathbf{1}_X)f, \qquad \text{for } f \in L_\infty^+(Y,q).$$

Markov kernels as linear maps

1. Given τ a Markov kernel from (X, Σ) to (Y, Λ), we define $T_\tau : \mathcal{L}^+(Y) \longrightarrow \mathcal{L}^+(X)$, for $f \in \mathcal{L}^+(Y)$, $x \in X$, as $T_\tau(f)(x) = \int_Y f(z)\tau(x, dz)$.
2. This map is well-defined, linear and ω-continuous.
3. If we write $\mathbf{1}_B$ for the indicator function of the measurable set B we have that $T_\tau(\mathbf{1}_B)(x) = \tau(x, B)$.
4. It encodes all the transition probability information

From linear maps to markov kernels

1. Conversely, any ω-continuous morphism L with $L(\mathbf{1}_Y) \leq \mathbf{1}_X$ can be cast as a Markov kernel by reversing the process on the last slide.
2. The interpretation of L is that $L(\mathbf{1}_B)$ is a measurable function on X such that $L(\mathbf{1}_B)(x)$ is the probability of jumping from x to B.

Backwards

1. We can also define an operator on $\mathcal{M}(X)$ by using τ the other way.
2. We define $\bar{T}_\tau : \mathcal{M}(X) \longrightarrow \mathcal{M}(Y)$, for $\mu \in \mathcal{M}(X)$ and $B \in \Lambda$, as $\bar{T}_\tau(\mu)(B) = \int_X \tau(x, B)\,\mathrm{d}\mu(x)$.
3. It is easy to show that this map is linear and ω-continuous.

What do they mean?

1. The operator $\bar{T}_\tau$ transforms measures "forwards in time"; if μ is a measure on X representing the current state of the system, $\bar{T}_\tau(\mu)$ is the resulting measure on Y after a transition through τ.
2. The operator T_τ may be interpreted as a likelihood transformer which propagates information "backwards", just as we expect from predicate transformers.
3. $T_\tau(f)(x)$ is just the expected value of f after one τ-step given that one is at x.

Labelled abstract Markov processes

The definition

An **abstract Markov kernel** from (X, Σ, p) to (Y, Λ, q) is an ω-continuous linear map $\tau : L^+_\infty(Y) \longrightarrow L^+_\infty(X)$ with $\|\tau\| \leq 1$.

LAMPS

A **labelled abstract Markov process** on a probability space (X, Σ, p) with a set of labels (or actions) $\mathcal{A}$ is a family of abstract Markov kernels $\tau_a : L^+_\infty(X, p) \longrightarrow L^+_\infty(X, p)$ indexed by elements a of $\mathcal{A}$.

The approximation map

The expectation value functors project a probability space onto another one with a possibly coarser σ-algebra.
Given an AMP on (X,p) and a map $\alpha : (X,p) \longrightarrow (Y,q)$ in $\mathbf{Rad}_\infty$, we have the following approximation scheme:

Approximation scheme

$$\begin{tikzcd}
L^+_\infty(X,p) \arrow[r, "\tau_a"] & L^+_\infty(X,p) \arrow[d, "\mathbb{E}_\infty(\alpha)"'] \\
L^+_\infty(Y,q) \arrow[u, "P_\infty(\alpha)"] \arrow[r, dotted, "\alpha(\tau_a)"] & L^+_\infty(Y,q)
\end{tikzcd}$$

A special case

- Take (X, Σ) and (X, Λ) with $\lambda \subset \Sigma$ and use the measurable function $id : (X, \Sigma) \to (X, \Lambda)$ as α.

Coarsening the σ-algebra

$$\begin{tikzcd}
L^+_\infty(X, \Sigma, p) \arrow[r, "\tau_a"] & L^+_\infty(X, \Sigma, p) \arrow[d, "\mathbb{E}_\infty(\alpha)"'] \\
L^+_\infty(X, \Lambda, p) \arrow[u, "P_\infty(\alpha)"] \arrow[r, dotted, "id(\tau_a)"] & L^+_\infty(X, \Lambda, p)
\end{tikzcd}$$

- Thus $id(\tau_a)$ is the approximation of τ_a obtained by averaging over the sets of the coarser σ-algebra Λ.
- We now have the machinery to consider approximating along arbitrary maps α.

The tower law

- The special case on the previous slide can also be done for the L_1 situation, we get the map $\mathbb{E}_1(id) : L_1^+(X, \Sigma, p) \longrightarrow L_1^+(X, \Lambda, p)$.
- This is exactly the map that is written as $\mathbb{E}(\cdot||\Lambda)$ in probability theory books.
- The tower law is written $\mathbb{E}[\mathbb{E}[X||\Lambda_2]||\Lambda_1] = \mathbb{E}[X||\Lambda_1]$ where $\Lambda_1 \subset \Lambda_2$ and is given a half-page proof.
- But this is just functoriality!