ON THE "PATCH PRODUCT" AND ITS APPLICATION TO CONTINUOUS MARKOV STATES

CARLOS PASQUALI

Contents

1. Abstract

In examining discrete Markov chains, the idea came to me (while contemplating a Manzanillo blue beach and a pair of hungry pelicans fish from it) that although it is indeed interesting that one can jump from a discrete state to another with fixed probability, there is really no reason why the number of states need be discrete. In other words, we could very well have an (uncountable) infinitude of states as represented by the closed interval, say, from 0 to 1, embedded in $\mathbb{R}$, and a probability of jumping to another state represented by a continuous density. The problem of calculating the n-step (this is necessarily discrete) probabilities distribution becomes a mechanical one, not a conceptual one, and this is how the idea of a "patch product" arose in my mind.

2. Definition of a patch

Assume a continuous function on $A \times B = [0,1] \times [0,1] \subset \mathbb{R} \times \mathbb{R}$, say $g(x,y)$. An $x \in A$ represents the state a random variable T will jump *to*, while a $y \in B$ represents the same random variable jumps *from*. Then $g(x,y)$ represents the probability that T will jump *to* state x *from* state y. To remain consistent with the axioms of probability, our function $g(x,y)$ is furthermore special, in that we require that $\int_0^1 g(x,y)dx = 1$, that is, picking any state in B, say y_c, any other state x, including y_c itself, is reachable with probability $p_{y_c,x} = g(x, y_c)$, where the whole sample space is any of the infinitude of states in A, and all individual probabilities therefore sum (continuously) to 1. Thus $g(x, y)$ is not a probability distribution, although $g(x, y_c)$ is; it is a "patch," a continuous collection of probability distributions.

From an unconventional point of view, we will want to think of this "patch" as the continuous version of a discrete (Markov) square matrix. Thinking about it this way greatly facilitates the mechanical operations we will want to describe presently.

3. The patch product

3.1. **Function vector by function vector.** It is very well studied that, for functions in the function linear space $C[0, 1]$ the inner product $\int_0^1 f(x)g(x)dx$ is well-defined and induces a useful distance metric. It can be thought of as the continuous sum of point-wise multiplications of two aligned functions. Note that they must be the same "size", that is, they are continuously defined on [0,1]. In the discrete case, when multiplying two vectors of the same size we do something similar: multiply entry-by-entry and then sum the individual products.

Suppose now that we have defined a probability distribution $j(x)$ on $C = [0, 1] \subset \mathbb{R}$ and another function $k(y)$ on $D = [0, 1] \subset \mathbb{R}$, and let us imagine that $j(x)$ can be thought of to lie horizontally (we could require $\int_0^1 j(x)dx = 1$) where $k(y)$ can be thought of as lying vertically (we do not require the integral across y to equal to 1 in this case, any function will do) on a Cartesian plane. In order to achieve a dotting mechanism, we must make sure that these two functions are aligned. Following the procedure for dotting two vectors, let us transform $j(x)$ by the isometry or isomorphism which rotates the function clockwise 90 degrees, and then shifts the function upwards one unit: $x \rightsquigarrow -y \rightsquigarrow 1 - y$ to obtain $j(x) \rightsquigarrow j(1 - y)$. Having aligned $j(1 - y)$, we can dot it with $k(y)$ with the inner product as defined before[1] by

$$\int_0^1 j(1 - y) \cdot k(y)dy$$

We could have also used the isometry $x \rightsquigarrow y$ so that $j(x) \rightsquigarrow j(y)$, with inner product $\int_0^1 j(y)k(y)dy$, but we wanted to emulate the mechanical procedure as done in usual matrix multiplication.

3.2. **Function vector by patch.** As before define a function $j(t)$ on $[0, 1] \subset \mathbb{R}$ (we can require that the integral across the interval be 1, in keeping with probability axioms), but this time define $k(x, y)$ on $[0, 1] \times [0, 1]$ (again in keeping with probability axioms, we can define that the integral across the domain be 1). In order to express the function vector by patch product analogously to the way we do it with matrix products, we would have to multiply $j(t)$ by each dy strand of

[1] To be thorough, we should show that the integral is an inner product.

the function $k(x,y)$. In order to bypass such difficulty, construct the new function $j(t,w) = j(t)$ also on $[0,1] \times [0,1]$, which consists of an infinitude of functions $j(t)$ along $[0,1]$, that is $j(t) \times [0,1]$. As before, let us apply unto this new function the isometry which sends $t \rightsquigarrow 1-y$, and $w \rightsquigarrow t$, so that $j(t,w) \rightsquigarrow j(1-y,t) = j(1-y)$. Having aligned the patches, we can dot them in the same way we do matrices mechanically by:

$$p(x) = \int_0^1 j(1-y,t) \cdot k(x,y)dy = \int_0^1 j(1-y) \cdot k(x,y)dy$$

Of course the "dimension" of $j(t)$, which in this case is the length of [0,1], matches the vertical length of $k(x,y) \in [0,1] \times [0,1]$.

We have created a function vector $p(x)$ with continuous entries defined on $[0,1]$, analogous to vector times matrix multiplication, which produces a vector.

3.3. **Patch by patch, the patch product.** There is no reason why we cannot define a "patch product," having defined a function vector by function vector "multiplication" and a function vector by patch "multiplication." Create a function $j(t,w)$, and in keeping with probability axioms make it so that having chosen a particular t, when sum across the domain, such continuous sum is equal to one. Next define $k(x,y)$ similarly. As before, apply on $j(t,w)$ the transformation that sends $t \rightsquigarrow 1-y$ and $w \rightsquigarrow t$, so that $j(t,w) \rightsquigarrow j(1-y,t)$. The point of this is of course to obtain a new "square" patch defined by

$$p(x,t) = \int_0^1 j(1-y,t) \cdot k(x,y)dy \in [0,1] \times [0,1]$$

This makes intuitive sense too, for, having chosen a particular t, the function $j(1-y,t) = j(1-y)$, and we have a function vector by patch product.

Having calculated the above, if we want, we can send the patch product $p(x,t)$ to $p(x,y)$ via the isometry or isomorphism that sends $t \rightsquigarrow y$.

3.4. **Patch Trace.** For any function on $[0,1] \times [0,1]$, say $h(x,y)$, we may be interested in the diagonal $y = -x+1$ or $x = 1-y$. Thus, we can either calculate $h(x, 1-x)$ or $h(1-y,y)$. The trace is given by the simple integral

$$tr(h) = \int_0^1 h(x, 1-x)dx$$

or

$$tr(h) = \int_0^1 h(1-y,y)dy$$

If we have multiplied two patches together, the patch product diagonal is given simply by

$$p(x, 1-x) = \int_0^1 j(1-y, 1-x) \cdot k(x, y) dy$$

or

$$p(1-t, t) = \int_0^1 j(1-y, t) \cdot k(1-t, y) dy$$

Thus, the patch product trace is given by

$$tr(p) = \int_0^1 p(x, 1-x) dx = \int_0^1 \int_0^1 j(1-y, 1-x) \cdot k(x, y) dy dx$$

or

$$tr(p) = \int_0^1 p(1-t, t) dt = \int_0^1 \int_0^1 j(1-y, t) \cdot k(1-t, y) dy dt$$

3.5. Continuous-State Closed-Interval Markov Chains.

3.5.1. *Transition probability patch.* Let a patch represent the continuous transition probabilities of a Markov chain, and call it $g(x, y)$. Analogously to the discrete case, we call this patch the "transition probability patch." We take care that each $g(x, y) \geq 0$, and that $\int_0^1 g(x, y) dx = 1$.

3.5.2. *Higher-Order Transition Probabilities and the Chapman-Kolmogorov Equation for patches.* Since we have devised a way to mechanically manipulate a patch the same way we do a matrix, tractability should not be an issue. Patch powers are defined by

$$P^2 = P \star P = p(x, t)_{(2)} = \int_0^1 g(1-y, t) \cdot g(x, y) dy$$

Similarly,

$$P^3 = P \star P^2 = p(x, t)_{(3)} = \int_0^1 g(1-y, t) \cdot p(x, y)_{(2)} dy$$

(here notice we transformed $p(x, t)_{(2)} \rightsquigarrow p(x, y)_{(2)}$ *after* we integrated by y) and in general:

$$P^{n+1} = P \star P^n = p(x, t)_{(n+1)} = \int_0^1 g(1-y, t) \cdot p(x, y)_{(n)} dy$$

Finally, we are left with the job of defining P^0, the identity patch. For this consider the function on $[0, 1] \times [0, 1]$ called $i(t, w)$, which is zero everywhere except along the diagonal $w = -t + 1$ or $t = 1 - w$. That is, $i(1 - w, w) = i(t, 1 - t) = 1$, and 0 otherwise.[2] Then

$$P^0 \star P = \int_0^1 i(1 - y, t) \cdot g(x, y) dy$$

is nonzero only for values where $t = y$, at which point we can do the substitution

$$\int_0^1 (1) \cdot g(x, t) dy = g(x, t) \rightsquigarrow g(x, y) = P.$$

Let $p(x, y)_{(n)}$ represent the n-step transition probability patch, that is, the patch that shows the probability of transitioning from state y to x in n steps. We compute this by taking patch powers and transforming $t \rightsquigarrow y$ at each step.

The patch identity

$$P^{n+m} = P^n P^m$$

can be written in terms of patch n-step transition probability patches, as

$$p(x, t)_{(n+m)} = \int_0^1 p(1 - y, t)_{(n)} \cdot p(x, y)_{(m)} dy \rightsquigarrow p(x, y)_{(n+m)}$$

in direct analogy with the Chapman-Kolmogorov equation. Thus, transition from y to x in $n + m$ steps can be achieved by moving from y to an intermediate in n steps and then moving from that intermediate to x in m more steps.

3.5.3. *The probability distribution of a beginning probability function vector p(0).* We may be interested in knowing how the continuous function vector $p(0)$ (initial-state probability function vector or initial-state distribution) propagates after n discrete transitions. This is simply a function vector multiplied by a patch.

$$p(n) = p(0) P^n$$

Equivalently,

$$p(x)_{(n)} = \int_0^1 p(1 - y)_{(0)} \cdot p(x, y)_{(n)} dy$$

The probability distribution $p(x)$ at step n is completely determined by the one-step transition probability patch P and the initial-state probability function vector (initial-state distribution) $p(0)$.

[2]Consider constructing the function piece-wise, as $i(t, w) = t + w$, which is equal to 1 along the diagonal of interest, and define it to be zero everywhere else.