Chapter 5

Lifts and splittings

Date: September 27, 2004

Definition 5.1 *Let $P = (P_x)_{x \in \mathcal{X}}$ be an $\mathcal{X}$-Markov kernel on $(\mathcal{X}, \mathbb{E})$. A* **lift** *of P is a Markov kernel Q on a space $(\mathcal{Z}, \mathbb{G})$ and a measurable map $h : \mathcal{Z} \to \mathcal{X}$ with the following property:*

For every probability measure λ on $\mathcal{X}$ there is a probability measure λ^* on $\mathcal{Z}$ such that whenever $Z_0, Z_1, Z_2, \ldots$ is a time homogenous Markov chain on $\mathcal{Z}$ satisfying

$$\begin{aligned} Z_{n+1} \mid Z_n &\overset{\mathcal{D}}{=} Q \\ Z_0 &\overset{\mathcal{D}}{=} \lambda^* \end{aligned}$$

then the proces $X_0, X_1, \ldots$, given by

$$X_n = h(Z_n) \qquad \text{for } n = 0, 1, \ldots$$

is a time homogenous Markov chain on X, satisfying

$$\begin{aligned} X_{n+1} \mid X_n &\overset{\mathcal{D}}{=} P \\ X_0 &\overset{\mathcal{D}}{=} \lambda \end{aligned}$$

The definition has a somewhat provocative touch to it. One of the basic facts of life in Markov chain theory is that if we apply a function to a Markov chain, the resulting proces will not be Markovian. There is an explicit example of this in example 2.14, and there is nothing perculiar in the example: this is the way things turn out in most examples, if we care to carry out the computations. Markov properties are easily lost.

But of course there may be some settings, where the Markov property is actually preserved, even if it is counterinutitive, and in these situations we may be able to find a lift. This discussion shows that the existence of lifts is a quite delicate matter.

The reader would probably have prefered a simpler formulation of the definition of a lift: why do we not simply require that if $Z_0, Z_1, \dots$ is a time homogenous Markov chain on $\mathcal{Z}$ with one-step transition probability Q, then $h(Z_0), h(Z_1), \dots$ is a Markov chain with one-step transition probability P? In this formulation, no attention is paid to the initial distribution. Unfortunately, the lifts we will be considering, will typically **not** have this property. If the Z-chain is started with an unfortunate initial distribution, it may not project down to a P-Markov chain on $\mathcal{X}$! It is quite easy to see this, once we have the construction, and it is also easy to see that the deviance is of no great importance - but it is there, and of course we need a definition which encompasses the properties of the actual constructions.

Example 5.2 A very trivial example of a lift of P is P itself, where the projection map $h : \mathcal{X} \to \mathcal{X}$ is the identity. It is hard to imagine that any useful conclusions can be obtained on P if we think of P as a lift, but formally it satisfies the definition.

∘

Example 5.3 An equally uninteresting example of a lift of P is a productkernel. If $R = (R_y)_{y \in \mathcal{Y}}$ is a $\mathcal{Y}$-kernel on $(\mathcal{Y}, \mathbb{K})$, we can construct a kernel $Q = (Q_{x,y})$ on $\mathcal{X} \times \mathcal{Y}$ by

$$Q_{x,y} = P_x \otimes R_y \qquad \text{for all } x \in \mathcal{X}, y \in \mathcal{Y} .$$

Running a Q-Markov chain $Z_n = (X_n, Y_n)$ on $\mathcal{X} \times \mathcal{Y}$ means running a P-Markov chain on $\mathcal{X}$ and a R-Markov chain on $\mathcal{Y}$ independent of each other. This particular lift actually has the property that not matter how it is started, the X-chain will be a P-Markov. But on the other hand, it is clear that we do not gain additional insight on the behaviour of P by studying it in combination with an independent chain with its own one-step transition probability.

∘

We have dismissed the lifts in example 5.2 and 5.3 as 'uninteresting'. But what would an interesting lift look like? Imagine for instance that the lift Q has an atom α, and that this atom is Harris recurrent. It is described in chapter 4 how such an atom often leads to strong law of large numbers by regenerative techniques. So based on the atom, we may know that

$$\frac{1}{n}\sum_{k=0}^{n-1} f(Z_k) \to L(f) \qquad \text{for } \ n \to \infty \quad \text{a.e.}$$

for instance for all boundede measurable functions $f : \mathcal{Z} \to \mathbb{R}$. At present, we do not have much to say about the limit $L(f)$, in principle it could depend on the initial distirbution of the Z-chain. But note the following: if $X_0, X_1, \ldots$ is a P-Markov chain on $\mathcal{X}$, we may think of it as the projection of the Z-chain. If $g : \mathcal{X} \to \mathbb{R}$ is measurable and bounded, we have that

$$\frac{1}{n}\sum_{k=0}^{n-1} g(X_k) = \frac{1}{n}\sum_{k=0}^{n-1} g \circ h(Z_n) \to L(g \circ h) \qquad \text{for } \ n \to \infty \quad \text{a.e.}$$

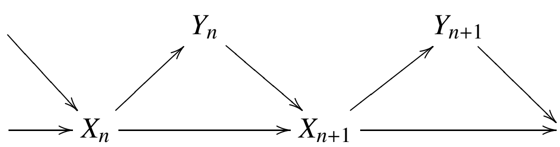
So we get a law of large numbers for the P-chain, simply by knowing that P has a lift with an atom (and a few minor technical details, of course, where we rigorously establish the law of large numbers in the atomic case). The P-chain does not enherit the atom itself, but it does enherit many of the properties that can be deduced from the atom.

This way of arguing is really the backbone of modern Markov chain theory. Results are proved for chains with an atom, using regenerative techniques. And then they are extended to chains that have lifts with an atom. At first sight this program seems rather hopeless: it would seem that the property of possesing a lift with an atom is a quite unsual. But it turns out to be surprisingly common.

Example 5.4 Let $\psi : \mathcal{X} \times \mathcal{Y} \times (0,1) \to \mathcal{X}$ and $\xi : \mathcal{X} \times (0,1) \to \mathcal{Y}$ be measureable maps. Consider the update schemes

$$Y_n = \xi(X_n, V_n)\,, \quad X_{n+1} = \phi(X_n, Y_n, U_{n+1}) \qquad \text{for } n = 0, 1, 2, \ldots$$

Here the error variables $U_1, V_1, U_2, V_2, \ldots$ are independent and thought of as standard uniformly distributed (though the space they have values in and their specific distribution is not relevant for the arguments). The updates run like this:



This scheme is perhaps more elaborate than the schemes we have considered so far. But it is easy to see that it givs rise to two Markov chains. The X-proces is time homogenous Markov, because it is created by the update scheme

$$\Big(x, (u, v)\Big) \mapsto \psi\Big(x, \xi(x, v), u\Big) \tag{5.1}$$

Similarly, the coupled proces (X_n, Y_n) is a time homogenous Markov, as it is created by the update scheme

$$\Big((x, y), (u, v)\Big) \mapsto \Big(\psi(x, y, u),\ \xi(\psi(x, y, u), v)\Big) \tag{5.2}$$

If P is the kernel generated by the X updates scheme (5.1) and Q is the kernel generated by the (X, Y)-update scheme (5.2), then clearly Q is a lift of P. Afterall, the X-proces is a projection of the (X, Y)-proces, and we can start the updates machinery with any distribution of X_0 that suits our fancy.

The challenge is to go the other way: starting with P, find ψ and ξ such that the X update scheme (5.1) generates P. Then the Q resulting from the same ψ and ξ will be a lift of P. If this construction is succesfull, we will say that Q is an **Athreya-Ney lift** of P.

Athrea-Ney lifts provide examples where it is clear that attention has to be paid to the initial distribution of the lifted chain, in order for the projected proces to be P-Markov. If we draw (X_0, Y_0) from an arbitrary distribution on $\mathcal{X} \times \mathcal{Y}$ and supply them to the update procedure, we in fact change the rules for the first update - the update from X_0 to X_1 follows a different procedure than the rest of the X-updates. The resulting X-proces will actually be Markov, as it is given by a time dependend update procedure, and we see that

$$X_{n+1} \mid X \quad \stackrel{\mathcal{D}}{=} \quad P \qquad \text{for all } n = 1, 2, \dots$$

But the first X-update follows a different rule, and hence the conditional distribution of X_1 given X_0 presumably will be different from P.

∘

We will focus on the construction in Athreya-Ney lifts in the case of a socalled **split** of $\mathcal{X}$, which simply means a situation where $\mathcal{Y}$ has exactly two points, say $\mathcal{Y} = \{0, 1\}$. The productspace $\mathcal{X} \times \mathcal{Y}$ is perhaps best visualized as two copies of $\mathcal{X}$, a low copy (corresponding to the $\mathcal{Y}$-point 0) and a high copy (corresponding to the $\mathcal{Y}$-point 1). We think of these two copies obtained by slicing a single $\mathcal{X}$-copy carefully in two, hence the word 'split'.

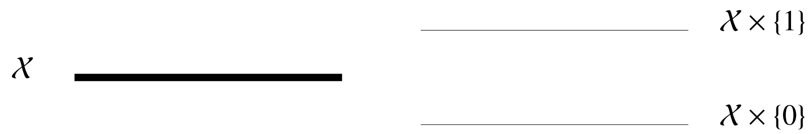


Figure 5.1: *The split construction. The space $\mathcal{X}$ is sliced in two, and replaced by the product set $\mathcal{X} \times \{0, 1\}$.*

In the split scenario, the $(x, y) \mapsto x$ update given by ψ can be thought of as two seperate $x \mapsto x$ updates with update schemes

$$(x, u) \mapsto \psi(x, 0, u) \,, \qquad (x, u) \mapsto \psi(x, 1, u) \,.$$

Suppose that these two update schemes generate $\mathcal{X}$-kernels P' and P''. Suppose furthermore that the $x \mapsto y$ update given by ξ gives rise to the $\mathcal{X}$-kernel $(W_x)_{x\in\mathcal{X}}$ on $\mathcal{Y}$, and let

$$\gamma(x) = W_x(\{1\}) \qquad \text{for all } x \in \mathcal{X}$$

A less formal - and perhaps less confusing - way of saying this is that $\gamma(x)$ is the conditional distribution of placing the (X, Y)-point on the upper $\mathcal{X}$-copy, given that $X = x$.

We claim that in this situation the kernel generated by (5.1) which we denote by $P = (P_x)_{x\in\mathcal{X}}$, is simply given by the formula

$$P_x = \Big(1 - \gamma(x)\Big)\, P'_x + \gamma(x)\, P''_x \qquad \text{for all } x \in \mathcal{X} \,. \tag{5.3}$$

The proof of this fact is most easily understood using the updating formalism. So we consider $(X_n, Y_n)_{n\in\mathbb{N}}$ given by update machinery, and we look for the conditional distribution of X_{n+1} given X_n.

$$\begin{aligned} P(X_n \in A, X_{n+1} \in B) &= P(X_n \in A, Y_n = 0, X_{n+1} \in B) + P(X_n \in A, Y_n = 1, X_{n+1} \in B) \\ &= \int_{A\times\{0\}} P'_x(B)\, d(X_n, Y_n)(P)(x, y) + \int_{A\times\{1\}} P''_x(B)\, d(X_n, Y_n)(P)(x, y) \\ &= \int_A (1 - \gamma(x))\, P'_x(B)\, dX_n(P)(x) + \int_A \gamma(x)\, P''_x(B)\, dX_n(P)(x) \\ &= \int_A (1 - \gamma(x))\, P'_x(B) + \gamma(x)\, P''_x(B)\, dX_n(P)(x) \,. \end{aligned}$$

So (5.3) does qualify as the conditional distribution of X_{n+1} given X_n.

Definition 5.5 *Let $P = (P_x)_{x \in \mathcal{X}}$ be an $\mathcal{X}$-Markov kernel on $(\mathcal{X}, \mathbb{E})$. A set $C \in \mathbb{E}$ is* **1-small** *for P if there is a probability measure ν and a $\delta \in (0,1)$ such that*

$$P_x(B) \geq \delta \nu(B) \qquad \text{for all } x \in \mathcal{X},\ B \in \mathbb{E}\,. \tag{5.4}$$

We usally refer to (5.4) as the **minorization condition**. An atom is of course a 1-small set. And a 1-small set C is close to being an atom in the following sense: we can split $\mathcal{X}$ in two and construct an Athreye-Ney lift using C such that one of the $\mathcal{X}$-copies is an atom for the lift.

Fundamentally, the idea is very simple: construct two new $\mathcal{X}$-kernels P' and P'' by

$$P'_x(B) = \begin{cases} P_x(B) & x \notin C \\ \big(P_x(B) - \delta\nu(B)\big)/(1-\delta) & x \in C \end{cases}$$

and

$$P''_x(B) = \nu(B) \qquad \text{for all } x \in \mathcal{X}, B \in \mathbb{E}\,.$$

Furthermore, define $\gamma : \mathcal{X} \to [0,1]$ by

$$\gamma(x) = \begin{cases} \delta & \text{if } x \in C \\ 0 & \text{if } x \notin C\,. \end{cases}$$

Construct update functions $\phi', \phi'' : \mathcal{X} \times (0,1) \to \mathcal{X}$ corresponding to the two kernels P' and P''. We may assume that ϕ'' does not involve x, as the kernel it is supposed to generate does not vary with x. Define

$$\psi(x,y,u) = \begin{cases} \phi'(x,u) & \text{if } y = 0 \\ \phi''(x,u) & \text{if } y = 1 \end{cases}$$

and

$$\xi(x,u) = \begin{cases} 1 & \text{if } u < \gamma(x) \\ 0 & \text{otherwise} \end{cases}$$

Using these update functions the Athrey-Ney construction in example 5.4 **will** give a lift of P to a Markov kernel Q on $\mathcal{X} \times \{0,1\}$, since

$$(1-\gamma(x))\,P''_x + \gamma(x)\,P''_x = P_x$$

by construction. And we note that $\mathcal{X} \times \{1\}$ is a vissible atom for the (X,Y)-update.

It is of course debatable what the atom really is. It is perhaps more reasonable to say that the atom is $C \times \{1\}$. A subset of an atom is always itself an atom, so this is in a sense trivial. But by the same token $(\mathcal{X} \setminus C) \times \{1\}$ is an atom, and this is a purely irrelvant atom, as the proces will **never** enter that part of the space.