Lecture 11. Markov chains and Pólya's theorem

Martin Klazar

December 15, 2020

Markov chains, again

Here is some literature on Markov chains.
• P. Billingsley, *Probability and Measure*, J. Wiley, 1995 (Chapter 8).
• A. Rényi, *Teorie pravděpodobnosti* (Probability theory), Academia, 1972 (Kapitola VIII.8).
• W. Feller, *An Introduction to Probability Theory and its Applications. Volume I*, J. Wiley, 1957 (Chapters 15 and 16).

• *Alfréd Rényi (1921–1970)* was a Hungarian mathematician. In 1959 he created, together with P. Erdős, random graphs.

• *William Feller (Vilibald Srećko Feller) (1906–1970)* was a Croatian-American probabilist who wrote the canonical two-volume textbook on probability theory. The volumes treat, respectively, discrete and continuous probability.

Recall that $\mathbb{N} = \{1, 2, \dots\}$ and $\mathbb{N}_0 = \{0, 1, \dots\}$. Let $P = (\Omega, \Sigma, \Pr)$ be a probability space. We give again definition of Markov chains; be warned that in the literature one can encounter imprecise formulations. Another, and lesser, problem is that some (especially more practically oriented) texts blissfully ignore the question of their existence.

Definition (Markov chain). Suppose that $S \neq \emptyset$ is an at most countable set (often $S = [n]_0$ for $n \in \mathbb{N}$ or $S = \mathbb{N}_0$) and $p_{i,j} \geq 0$, $i, j \in S$, are given real constants such that $\sum_j p_{i,j} = 1$ for every $i \in S$. Then a sequence $X_0, X_1, X_2, \dots$ of discrete random variables on P, where $X_t \colon \Omega \to S$, is a *Markov chain (with transition probabilities*

$p_{i,j}$) if for every $t \in \mathbb{N}$ and every $i, j, a_0, \ldots, a_{t-2} \in S$ one has that

$$\begin{aligned} p_{i,j} &= \Pr(X_t = j \,|\, X_{t-1} = i \wedge X_{t-2} = a_{t-2} \wedge \cdots \wedge X_0 = a_0) \\ &= \Pr(X_t = j \,|\, X_{t-1} = i) \end{aligned}$$

whenever the initial conditional probability is defined, i.e. whenever $\Pr(X_{t-1} = i \wedge X_{t-2} = a_{t-2} \wedge \cdots \wedge X_0 = a_0) > 0$. □

The variable t is discrete *time*, S is the *set of states* (so the values of any X_t are the *states*), *finite* Markov chains range in a finite set S, the $p_{i,j}$ are the *transition probabilities*, and the $\alpha_i := \Pr(X_0 = i)$ are the *initial probabilities*, or the *initial distribution* (note that $\sum_i \alpha_i = 1$). The transition probabilities are independent of time.

As a short detour we mention here that for $a, b \in \mathbb{R}$ and continuous real random variables X and Y one defines the conditional probabilities $\Pr(X < a \,|\, Y = b)$ by means of the *Radon–Nikodym theorem* (Rényi, Kap. V; Billingsley, Chap. 32–34).

• *Johann Radon (1887–1956)* was an Austrian mathematician who was born in Děčín (Tetschen), like the lecturer. There is a commemorative plaquette on a house on the main square (commemorating JR, not me). The Radon transform has application in tomography.

• *Otto M. Nikodym (1887–1974)* was a Polish mathematician who lived since 1948 in the USA.

From now on $S = [n]_0$ for some $n \in \mathbb{N}$ or $S = \mathbb{N}_0$, if it is not said else. A finite or infinite real matrix $(p_{i,j}) = (p_{i,j})_{i,j \in S}$, $p_{i,j} \geq 0$, is *stochastic* if $\sum_j p_{i,j} = 1$ for every i. The following existence theorem shows that Markov chains is a reasonable concept.

Theorem (Billingsley, Theorem 8.1). *For any stochastic matrix* $(p_{i,j})$ *and any real numbers* $\alpha_i \geq 0$ *with* $\sum_i \alpha_i = 1$ *there exist a probability space and random variables* $X_0, X_1, \ldots$ *on it such that the* X_t *form a Markov chain with initial probabilities* α_i *and transition probabilities* $p_{i,j}$.

As for the proof, for a countable set of states S one needs Lebesgue measure on the unit interval, see Billingsley's book. For finite S he suggests in Problem 8.1 in his book to construct the required probability space on the set $\Omega = S^{\mathbb{N}_0}$.

We present two examples of Markov chains.

Example 1 (Ehrenfest model). This model is due to T. and P. Ehrenfest in 1907 and is also called the *dog-flea model.* Two dogs D_1 and D_2 stand close each to the other and there are $a \in \mathbb{N}$ fleas on them. At each time $t \in \mathbb{N}_0$ a randomly selected flea jumps on the other dog. Let X_t be the number of fleas on the dog D_1 at time t. Thus defined Markov chain has transition probabilities, for $j \in \mathbb{N}_0$,

$$\begin{aligned} p_{j,j-1} &= \Pr(X_t = j-1 \,|\, X_{t-1} = j) = \frac{j}{a}\ (j > 0)\ , \\ p_{j,j+1} &= \Pr(X_t = j+1 \,|\, X_{t-1} = j) = \frac{a-j}{a} \end{aligned}$$

and $p_{j,k} = 0$ in any other case. The original interpretation is that there are in total $a \in \mathbb{N}$ molecules in two neighboring containers, and at each time $t \in \mathbb{N}_0$ a randomly selected molecule moves to the other container. The Markov chain X_t then describes the evolution of the number of molecules in the first (and in the second) container. Obviously, there should be tendency to equalization of the numbers of molecules (fleas) in both containers (on both dogs). I return to this model in the next lecture. It was important in the development of thermodynamics and statistical physics as a reply to objections to so called H-theorem of L. Boltzmann. $\square$

• *Tatiana Ehrenfest (Kiev, 1876–Leiden, 1964)*, née *Afanasjeva*, was a Russian–Dutch mathematician and physicist working in statistical mechanics. Since 1904 she was the wife of P. Ehrenfest.
• *Paul Ehrenfest (1880–1933)* was an Austrian–Dutch physicist.
• *Ludwig Boltzmann (1844–1906)* was an Austrian philosopher and physicist who founded statistical physics and defined entropy. Sadly, both men (PE and LB) ended their lives by suicide.

Example 2 (random walk on $\mathbb{Z}^k$). Random variables $X_0, X_1, \ldots$ of this Markov chain range in the set $S := \mathbb{Z}^k$, $k \in \mathbb{N}$, of lattice points (points with integral coordinates) in the Euclidean space $\mathbb{R}^k$. Two points $a, b \in \mathbb{Z}^k$ are *neighbors* if they have (Euclidean) distance 1.

Each point $a \in \mathbb{Z}^k$ has exactly $2k$ neighbors. The transition probabilities are

$$p_{a,b} = \Pr(X_t = b \,|\, X_{t-1} = a) = \begin{cases} 1/2k & \ldots \quad a \text{ and } b \text{ are neighbors} \ , \\ 0 & \ldots \quad \text{else} \end{cases}$$

— one moves from a point to any of its neighbors with the same probability. One can imagine excitation of an atom in a crystal moving randomly around, or a drunkard wandering aimlessly in the net of streets and avenues in New York, etc. $\square$

Pólya's theorem

Pólya's theorem describes long-time behavior of the previous random walk on $\mathbb{Z}^k$. For the proof we need the next simple lemma.

Lemma (the 1st Borel–Cantelli lemma). *Suppose that* $(\Omega, \Sigma, \Pr)$ *is a probability space and* $A_n \in \Sigma$*,* $n \in \mathbb{N}$*, are events in it such that* $\sum_n \Pr(A_n) < +\infty$*. Then*

$$\Pr(\limsup A_n) = 0 \ ,$$

where

$$\limsup A_n := \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n \in \Sigma$$

is the event that infinitely many of the events A_n *occur.*

Proof. For every $m \in \mathbb{N}$ we clearly have that $\limsup A_n \subset \bigcup_{n \geq m} A_n$. Thus, by the union bound,

$$\Pr(\limsup A_n) \leq \sum_{n \geq m} \Pr(A_n) \ .$$

Tails of any convergent series go to 0, thus $\Pr(\limsup A_n) = 0$. $\square$

• *Émile Borel (1871–1956)* was a French measure theorist, probabilist and politician. In 1925 he served as the minister of marine under the premier Paul Painlevé who was also a mathematician.
• *Francesco P. Cantelli (1875–1966)* was an Italian mathematician, born in Palermo, who started his career in astronomy and celestial mechanics.

We start the general part of the proof of Pólya's theorem. I follow Billingsley's book, thus notation and wording is often his. We fix $S = \mathbb{N}_0$ and a Markov chain $X_0, X_1, \dots$ with all initial probabilities $\alpha_i > 0$. We denote by P_i probabilities conditional on $X_0 = i \in \mathbb{N}_0$: for any event A,

$$P_i(A) := \Pr(A \mid X_0 = i) .$$

Thus, as we know from an exercise in the previous lecture,

$$P_i(X_t = i_t,\ 1 \le t \le n) = p_{i,i_1}\, p_{i_1,i_2}\, \dots\, p_{i_{n-1}i_n} .$$

Therefore for any $i, i_k, j_k \in \mathbb{N}_0$ we have that

$$\begin{aligned} &P_i(X_1 = i_1,\ \dots,\ X_m = i_m,\ X_{m+1} = j_1,\ \dots,\ X_{m+n} = j_n) \\ &= P_i(X_1 = i_1,\ \dots,\ X_m = i_m) \cdot P_{i_m}(X_1 = j_1,\ \dots,\ X_n = j_n) . \end{aligned}$$

Suppose that I is a set (finite or infinite) of m-long sequences of states, J is a set of n-long sequences of states, and every sequence in I ends in j. Adding both sides of the previous equation for $(i_1, \dots, i_m)$ ranging over I and $(j_1, \dots, j_n)$ ranging over J gives

$$\begin{aligned} &P_i((X_1,\ \dots,\ X_m) \in I,\ (X_{m+1},\ \dots,\ X_{m+n}) \in J) \\ &= P_i((X_1,\ \dots,\ X_m) \in I) \cdot P_j((X_1,\ \dots,\ X_n) \in J) . \end{aligned} \tag{1}$$

Here it is essential that each sequence in I ends in j.

Let

$$f_{i,j}^{(n)} := P_i(X_1 \ne j,\ \dots,\ X_{n-1} \ne j,\ X_n = j)$$

be the probability of a first visit to j at time n when we start in i, and let

$$f_{i,j} := P_i\big(\textstyle\bigcup_{n=1}^{\infty} \text{the event that } X_n = j\big) = \displaystyle\sum_{n=1}^{\infty} f_{i,j}^{(n)}$$

be the probability of an eventual visit.

Definition. A state i is called *persistent* (or *recurrent*) if the Markov chain starting at i is certain sometime to return to i: $f_{i,i} = 1$. The state is called *transient* in the opposite case: $f_{i,i} < 1$. □

Suppose that $n_1, \dots, n_k$ are integers satisfying $1 \le n_1 < \cdots < n_k$ and consider the event that the chain visits j at times $n_1, \dots, n_k$ but not in between. This event is determined by the conditions that $X_1 \neq j$, ..., $X_{n_1-1} \neq j$, $X_{n_1} = j$, $X_{n_1+1} \neq j$, ..., $X_{n_2-1} \neq j$, $X_{n_2} = j$, ..., $X_{n_{k-1}+1} \neq j$, ..., $X_{n_k-1} \neq j$, and $X_{n_k} = j$. Repeated application of (1) shows that the P_i-probability of this event is

$$F := f_{i,j}^{(n_1)}\, f_{j,j}^{(n_2-n_1)}\, f_{j,j}^{(n_3-n_2)}\, \dots\, f_{j,j}^{(n_k-n_{k-1})}\ .$$

We add this over the k-tuples $n_1, \dots, n_k$ by the nested summation

$$\sum_{n_1=1}^{\infty} \sum_{\substack{n_2 \\ n_2>n_1}} \sum_{\substack{n_3 \\ n_3>n_2}} \cdots \sum_{\substack{n_k \\ n_k>n_{k-1}}} F\ ,$$

and by the above definition of $f_{i,j}$ get that the P_i-probability of $X_n = j$ for at least k different values of n is $f_{i,j} f_{j,j}^{k-1}$ (we always consider the first k visits and these are unique). Letting $k \to \infty$ therefore gives the formula

$$P_i(X_n = j \text{ i.o.}) = \begin{cases} 0 & \text{if } f_{j,j} < 1\ , \\ f_{i,j} & \text{if } f_{j,j} = 1 \end{cases}$$

where "i.o." abbreviates "infinitely often".

In more details (Billingsley is laconic), the event A that $X_0 = i$ and $X_n = j$ i.o. is for any k contained in the event A_k that $X_0 = i$ and $X_n = j$ for at least k different values of n. This gives the first case of the formula. In the second case we use that the events A_k are nested, $A_1 \supset A_2 \supset \dots$, have the same probability $\Pr(A_k) = f_{i,j}$, and A is their intersection. From

$$A = \bigcap_{k=1}^{\infty} A_k = A_1 \setminus \bigcup_{k=1}^{\infty} (A_k \setminus A_{k+1})$$

we get that indeed

$$\Pr(A) = \Pr(A_1) - \sum_{k=1}^{\infty} (\Pr(A_k) - \Pr(A_{k+1})) = f_{i,j} - \sum_{k=1}^{\infty} 0 = f_{i,j}\ .$$

Setting $i = j$ in the formula gives

$$P_i(X_n = i \text{ i.o.}) = \begin{cases} 0 & \text{if } f_{i,i} < 1\ , \\ 1 & \text{if } f_{i,i} = 1\ . \end{cases} \tag{2}$$

For $n \in \mathbb{N}_0$ and states i, j we denote by $p_{i,j}^{(n)}$ the probability of transition from i to j in n steps,

$$p_{i,j}^{(n)} := P_i(X_n = j) .$$

Thus $p_{i,j}^{(1)} = p_{i,j}$, $p_{i,j}^{(0)} = 0$ for $i \neq j$ and $p_{i,i}^{(0)} = 1$.

Theorem (Billingsley, Theorem 8.2). *The above defined transient and persistent (recurrent) states are characterized by the following conditions.*

1. *Transience of a state i is equivalent to $P_i(X_n = i \text{ i.o.}) = 0$ and to $\sum_n p_{i,i}^{(n)} < +\infty$.*
2. *Persistence (recurrence) of a state i is equivalent to $P_i(X_n = i \text{ i.o.}) = 1$ and to $\sum_n p_{i,i}^{(n)} = +\infty$.*

Proof. By the first Borel–Cantelli lemma, $\sum_n p_{i,i}^{(n)} < +\infty$ implies $P_i(X_n = i \text{ i.o.}) = 0$, which by (2) in turn implies the transience $f_{i,i} < 1$. The entire theorem will be proved if it is shown that $f_{i,i} < 1$ implies $\sum_n p_{i,i}^{(n)} < +\infty$.

We look at the first passages through a state j. By (1) (used on the third line),

$$\begin{aligned}
p_{i,j}^{(n)} &= P_i(X_n = j) \\
&= \sum_{s=0}^{n-1} P_i(X_1 \neq j,\, \ldots,\, X_{n-s-1} \neq j,\, X_{n-s} = j,\, X_n = j) \\
&= \sum_{s=0}^{n-1} P_i(X_1 \neq j,\, \ldots,\, X_{n-s-1} \neq j,\, X_{n-s} = j) P_j(X_s = j) \\
&= \sum_{s=0}^{n-1} f_{i,j}^{(n-s)} p_{j,j}^{(s)} .
\end{aligned}$$

Therefore, by changing order of summation in the next finite double sum,

$$\sum_{t=1}^{n} p_{i,i}^{(t)} = \sum_{t=1}^{n} \sum_{s=0}^{t-1} f_{i,i}^{(t-s)} p_{i,i}^{(s)} = \sum_{s=0}^{n-1} p_{i,i}^{(s)} \sum_{t=s+1}^{n} f_{i,i}^{(t-s)} \leq \sum_{s=0}^{n} p_{i,i}^{(s)} f_{i,i} .$$

Rearranging the obtained inequality and using that $p_{i,i}^{(0)} = 1$, we get that $(1 - f_{i,i}) \sum_{t=1}^{n} p_{i,i}^{(t)} \le f_{i,i}$. If $f_{i,i} < 1$, we get for every $n \in \mathbb{N}$ the bound

$$\sum_{t=1}^{n} p_{i,i}^{(t)} \le \frac{f_{i,i}}{1 - f_{i,i}} \ .$$

Thus the series $\sum_n p_{i,i}^{(n)}$ converges. $\square$

We start the specific part of the proof of Pólya's theorem. In the following I use the asymptotic notation $\ll$ and $\gg$ as synonymous to the $O(\cdots)$ notation; in physics it often has the $o(\cdots)$ meaning.

Lemma. *The following three results hold.*

1. *If $\alpha_1, \ldots, \alpha_k \ge 0$ are real numbers with $\sum_{i=1}^{k} \alpha_i = 1$ then*

$$\sum_{i=1}^{k} \alpha_i^2 \le \max_{1 \le i \le k} \alpha_i \ .$$

2. *If $a > b \ge 0$ are integers with $a \ge b + 2$ then*

$$a! \cdot b! > (a-1)! \cdot (b+1)! \ .$$

3. *Let $m \in \mathbb{N}$. Then for $n = 3m$, $n = 3m+1$ and $n = 3m+2$ we have, respectively,*

$$\frac{n!}{m!^3} \ll \frac{3^n}{n}, \quad \frac{n!}{(m+1)! \cdot m!^2} \ll \frac{3^n}{n} \quad \text{and} \quad \frac{n!}{(m+1)!^2 \cdot m!} \ll \frac{3^n}{n} \ .$$

4. *For every $n \in \mathbb{N}_0$,*

$$\sum_{u=0}^{n} \binom{n}{u} \binom{n}{n-u} = \binom{2n}{n} \ .$$

Proof. Do it as (easy) exercises. In part 3 use the Stirling (asymptotic) formula

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \quad \text{as} \quad n \to \infty \ .$$

$\square$

Theorem (G. Pólya, 1921). *In the above Example 2 of the random walk on $\mathbb{Z}^k$, all states (points in $\mathbb{Z}^k$) are persistent (recurrent) if $k = 1$ or $k = 2$, and all are transient if $k \geq 3$.*

Proof. (After Billingsley, case $k = 3$ modified.) The probability $p_{a,a}^{(n)}$ (of a return to the point a in n steps) is the same for all a; we set $a_n^{(k)} := p_{a,a}^{(n)}$. Clearly, $a_{2n+1}^{(k)} = 0$ because transition to a neighbor of a point $a \in \mathbb{Z}^k$ flips the parity of the sum of coordinates. We only consider cases $k = 1, 2$ and 3, the case $k \geq 4$ is similar to (but in notation more complicated than) the case $k = 3$.

Let $k = 1$: we are on the line, in $\mathbb{Z}$. We have that

$$a_{2n}^{(1)} = \binom{2n}{n} \frac{1}{2^{2n}} = \frac{(2n)!}{n!^2} \cdot \frac{1}{2^{2n}} \; .$$

Plugging in the Stirling formula we get that $a_{2n}^{(1)} \sim (\pi n)^{-1/2}$. So $\sum_n a_n^{(1)} = +\infty$ and all states are persistent by the previous theorem.

Let $k = 2$: we are in the plane, in $\mathbb{Z}^2$. Now a return to the starting point in $2n$ steps means equal numbers of steps east and west as well as equal numbers north and south:

$$\begin{aligned} a_{2n}^{(2)} &= \sum_{u=0}^{n} \frac{(2n)!}{u!^2 (n-u)!^2} \cdot \frac{1}{4^{2n}} = \frac{1}{4^{2n}} \binom{2n}{n} \sum_{u=0}^{n} \binom{n}{u} \binom{n}{n-u} \\ &= \frac{1}{4^{2n}} \binom{2n}{n}^2 \sim \frac{1}{\pi n} \quad \text{(4 of the Lemma and the Stirling f.)} \; . \end{aligned}$$

Again, $\sum_n a_n^{(2)} = +\infty$ and every state is persistent.

Let $k = 3$: we are in the space, in $\mathbb{Z}^3$. Now we get $(u, v \in \mathbb{N}_0)$

$$\begin{aligned} a_{2n}^{(3)} &= \frac{1}{6^{2n}} \sum_{u+v \leq n} \frac{(2n)!}{u!^2 v!^2 (n-u-v)!^2} \\ &= \binom{2n}{n} 4^{-n} \sum_{u+v \leq n} \left[\frac{1}{3^n} \binom{n}{u,\, v,\, n-u-v} \right]^2 . \end{aligned}$$

The numbers in the $[\ldots]$s sum up to 1 because $3^n = (1+1+1)^n = \sum_{u+v \leq n} \binom{n}{u,v,n-u-v}$ by the multinomial theorem. By parts 1 and 2 of the Lemma we get that $(x, y, z \in \mathbb{N}_0)$

$$\sum_{\cdots}[\cdots]^2 \le \max_{x+y+z=n} \frac{1}{3^n}\binom{n}{x,\, y,\, z} = \frac{1}{3^n}\binom{n}{x_0,\ y_0,\ z_0}$$

where $(m \in \mathbb{N})$ (x_0, y_0, z_0) equals (m, m, m) if $n = 3m$, $(m+1, m, m)$ if $n = 3m+1$, and $(m+1, m+1, m)$ if $n = 3m+2$. By part 3 of the Lemma,

$$\binom{n}{x_0,\ y_0,\ z_0} \ll \frac{3^n}{n}\ .$$

Since, as we know, $\binom{2n}{n} \cdot 4^{-n} \sim cn^{-1/2}$ for a constant $c > 0$, we get the bound

$$a_{2n}^{(3)} \ll n^{-1/2}n^{-1} = n^{-3/2}\ .$$

Thus $\sum_n a_n^{(3)} < +\infty$ and by the previous theorem all states (points) are transient. $\square$

This theorem was proved (not exactly in this way, of course) in the article G. Pólya, Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt im Straßennetz, *Math. Ann.* **84** (1921), 149–160.

• *George (György) Pólya (1887–1985)* was a Hungarian–American mathematician who worked mainly in complex analysis, but also — as we have seen — in probability theory and in combinatorics (Pólya's enumeration method). He is known for his book *How to solve it* discussing heuristics for solving mathematical problems. The book was published in Czech translation by MATFYZPRESS.

Thank you!

(final version of January 13, 2021)