Lecture 26. Time-homogeneous Markov chains.

Now let us restrict ourselves to a narrower class of Markov chains.

A Markov chain is called (time-)homogeneous if the transition matrix is the same at every step: $P_1 = P_2 = ... = P_k = ... = P = (p_{xy})_{x,\, y \in X}$. All examples up to now were those of homogeneous chains; and from now on we will be considering only homogeneous chains.

It is clear that for homogeneous Markov chains formula (25.26) becomes just

$$P^{n\,m} = P^{m-n} \tag{26.1}$$

(the powers of the same square matrix), and this matrix depends only on the number of steps $k = m - n$ between the time moments n and m. Let me introduce the notation for the transition probability from x to y in k steps: $p_{xy}^{(k)} = p_{xy}^{n,\, n+k}$ (the $(x,\, y)$-th entry of the k-th power P^k of the one-step transition matrix; all notations with superscripts without parentheses are already in use). It is clear that $p_{xy}^{(0)} = \delta_{xy}$, $p_{xy}^{(1)} = p_{xy}$.

Equality (25.24) becomes, for homogeneous chains,

$$p^{(k+l)} = \sum_{z \in X} p_{xz}^{(k)} \cdot p_{zy}^{(l)}; \tag{26.2}$$

in particular,

$$p_{xy}^{(k+1)} = \sum_{z \in X} p_{xz}^{(k)} \cdot p_{zy} = \sum_{z \in X} p_{xz} \cdot p_{zy}^{(k)}. \tag{26.3}$$

Sequences of random variables are one class of mathematical models of ‘real-world’ processes. The very simplest model of this kind is that of a sequence of independent random variables. But sometimes we cannot disregard dependence between the values of a ‘real-world’ process at different times; the simplest model taking this into account is that of Markov chains. In the ‘real-world’ processes, whose mathematical models are sequences of random variables, sometimes the dependence between phenomena divided by some time interval decreases as this time interval becomes larger, and finally vanishes. E. g., in a text written in a natural language, if a letter is a vowel, its neighbor is most probably a consonant, and vice versa; a dependence through a space of one letter still remains, but it is clear that the fact that the 50-th letter of a text is a consonant practically does not have any influence on whether the 2008-th letter is a vowel or a consonant.

The theorem that follows is the simplest result within our mathematical theory that is a reflection of this empirical observation.

Theorem 26.1. *Let* ξ_0, ξ_1, ..., ξ_n, ... *be a homogeneous finite* (which means that the space X is finite) *Markov chain. Let the transition matrix* P *have a column with all entries positive*: *for some* $y_0 \in X$ *for all* $x \in X$ *we have* $p_{xy_0} > 0$.

Then the transition probabilities $p_{xy}^{(k)}$ *have limits as* $k \to \infty$, *and these limits do not depend on the starting point* x: *for all* $x, y \in X$

$$\lim_{k\to\infty} p_{xy}^{(k)} = p_y. \tag{26.4}$$

Proof. There exists a purely algebraic proof of this theorem using the Jordan form of the matrix P; I will give another proof, using inequalities. Let us denote by ε a positive number that is not greater than all p_{xy_0}.

Let us introduce the notations:

$$m_y^{(k)} = \min_{x\in X} p_{xy}^{(k)}, \qquad M_y^{(k)} = \max_{x\in X} p_{xy}^{(k)}. \tag{26.5}$$

We know that

$$p_{xy}^{(k)} = \sum_z p_{xz}\, p_{zy}^{(k-1)}. \tag{26.6}$$

It follows from $m_y^{(k-1)} \le p_{zy}^{(k-1)} \le M_y^{(k-1)}$, $p_{xz} \ge 0$, $\sum_z p_{xz} = 1$ that

$$m_y^{(k-1)} = \sum_z p_{xz}\, m_y^{(k-1)} \le \sum_z p_{xz}\, p_{zy}^{(k-1)} = p_{xy}^{(k)} \le \sum_z p_{xz}\, M_y^{(k-1)} = M_y^{(k-1)}; \tag{26.7}$$

that is, $m_y^{(k-1)} \le p_{xy}^{(k)} \le M_y^{(k-1)}$, whence

$$m_y^{(k-1)} \le m_y^{(k)} = \min_x p_{xy}^{(k)} \le M_y^{(k)} = \max_x p_{xy}^{(k)} \le M_y^{(k-1)}. \tag{26.8}$$

So we see that for every $y \in X$ the sequence $m_y^{(k)}$ is non-decreasing, and $M_y^{(k)}$ non-increasing. This means that the limits $\lim_{k\to\infty} m_y^{(k)}$, $\lim_{k\to\infty} M_y^{(k)}$ exist. To establish that the limits $\lim_{k\to\infty} p_{xy}^{(k)}$ exist and do not depend on x, it is enough to check that $\lim_{k\to\infty}[M_y^{(k)} - m_y^{(k)}] = 0$ (then the limits of $m_y^{(k)}$ and of $M_y^{(k)}$ coincide, and to the same limit converge also the transition probabilities $p_{xy}^{(k)}$ since they are between these bounds).

Let us estimate the difference $M_y^{(k)} - m_y^{(k)}$ through $M_y^{(k-1)} - m_y^{(k-1)}$. Let ε be a positive number that is not greater than all p_{xy_0}. Let the minimum $m_y^{(k)}$ be attained at $x = x_1$, the maximum $M_y^{(k)}$ at $x = x_2$:

$$m_y^{(k)} = p_{x_1 y}^{(k)}, \qquad M_y^{(k)} = p_{x_2 y}^{(k)}. \tag{26.9}$$

Then we have:

$$\begin{aligned} M_y^{(k)} - m_y^{(k)} &= p_{x_2 y}^{(k)} - p_{x_1 y}^{(k)} \\ &= \sum_z p_{x_2 z}\, p_{zy}^{(k-1)} - \sum_z p_{x_1 z}\, p_{zy}^{(k-1)} = \sum_z (p_{x_2 z} - p_{x_1 z})\, p_{zy}^{(k-1)}. \end{aligned} \tag{26.10}$$

Let us divide this sum into two: with positive $p_{x_2z} - p_{x_1z}$ and that with non-positive. Making use of the fact that $p_{zy}^{k-1} \le M_y^{(k-1)}$ in the first sum, and the fact that $p_{zy}^{k-1} \ge m_y^{(k-1)}$ in the second one, we obtain:

$$M_y^{(k)} - m_y^{(k)} \le \sum_{z:\, p_{x_2z}-p_{x_1z}>0} (p_{x_2z} - p_{x_1z}) \cdot M_y^{(k-1)} + \sum_{z:\, p_{x_2z}-p_{x_1z}\le 0} (p_{x_2z} - p_{x_1z}) \cdot m_y^{(k-1)}. \quad (26.11)$$

The sums that are multiplied by $M_y^{(k-1)}$ and by $m_y^{(k-1)}$ are the same in absolute value and opposite in sign (because together they add up to $1 - 1 = 0$). One of them contains the summand with $z = y_0$; this sum is not greater than $\sum_z p_{x_2z} - p_{x_1y_0} \le 1 - \varepsilon,$ and is not smaller than $-1 + \varepsilon$ (of course the other sum, that which does not contain the term with $z = y_0$, is also within these bounds). Therefore

$$M_y^{(k)} - m_y^{(k)} \le (M_y^{(k-1)} - m_y^{(k-1)}) \cdot (1 - \varepsilon). \quad (26.12)$$

We have for $k = 0$: $M_y^{(0)} - m_y^{(0)} = 1 - 0 = 1$, from which by induction

$$M_y^{(k)} - m_y^{(k)} \le (1 - \varepsilon)^k. \quad (26.13)$$

So the difference $M_y^{(k)} - m_y^{(k)}$ converges to 0 as $k \to \infty$, and this exponentially fast; it follows from this that $p_{xy}^{(k)} \to p_y$ as $k \to \infty$ (exponentially fast).

Theorem 26.2. *Let ξ_0, ξ_1, ..., ξ_n, ... be a homogeneous finite Markov chain. Let there exits a natural k_0 such that the transition matrix in k steps P^{k_0} has a column with all entries positive.*

Then the transition probabilities $p_{xy}^{(k)}$ have limits as $k \to \infty$, and these limits do not depend on the starting point x.

Proof. Instead of using (26.6), we use

$$p_{xy}^{(k)} = \sum_z p_{xz}^{(k_0)} p_{zy}^{(k-k_0)}; \quad (26.14)$$

we obtain, for $k \ge k_0$:

$$M_y^{(k)} - m_y^{(k)} \le (M_y^{(k-k_0)} - m_y^{(k-k_0)}) \cdot (1 - \varepsilon), \quad (26.15)$$

and for $k = r \cdot k_0 + s$, $0 \le s < k_0$

$$M_y^{(k)} - m_y^{(k)} \le (1-\varepsilon)^r \cdot (M_y^{(s)} - m_y^{(s)}) \le (1-\varepsilon)^r \le (1-\varepsilon)^{(k-k_0)/k_0} \to 0 \quad (k \to \infty). \quad (26.16)$$

Let us give the following definition: a homogeneous Markov chain is called *ergodic* if its transition probabilities in k steps converge as $k \to \infty$ to a *probability distribution* that does not depend on the initial point x: $\lim_{k\to\infty} p_{xy}^{(k)} = p_y$, $\sum_y p_y = 1$.

Theorem 26.2′. *Under the conditions of* Theorem 26.2 *the Markov chain under consideration is ergodic.*

All sums being finite ones, the fact that the sum of p_y is equal to 1 follows from the fact that the pre-limit sums are equal to 1.

46 Give an example of a time-homogeneous Markov chain for which the limit $\lim_{k\to\infty} p_{xy}^{(k)}$ does not exist.

47 Give an example of a time-homogeneous Markov chain for which the limit $\lim_{k\to\infty} p_{xy}^{(k)}$ exists, but depends on x.

48 Is the Markov chain with the transition matrix (23 – 24.24) ergodic?

49* For finite homogeneous Markov chains, is the condition of Theorem 26.2 *necessary and sufficient* for the chain to be ergodic?

In fact, there are necessary and sufficient conditions for ergodicity of time-homogeneous Markov chains, not only finite (i. e., in a finite space X) as we have considered, but also for infinite countable – such as, say, random walks considered in Lectures 23 – 24; but we have no time for that now.

What follows was not in Lecture 26; but I include it in the lecture note.

Theorem 26.3. *For an ergodic Markov chain* (finite, or infinite) *with arbitrary initial probabilities* q_x, $x \in X$,

$$\lim_{n\to\infty} P\{\xi_n = y\} = p_y, \tag{26.17}$$

where $p_y = \lim_{n\to\infty} p_{xy}^{(n)}$.

Proof. By the total probability formula,

$$P\{\xi_n = y\} = \sum_{x\in X} P\{\xi_0 = x\} \cdot P\{\xi_n = y|\xi_0 = x\} = \sum_x q_x \cdot p_{xy}^{(n)}. \tag{26.18}$$

So we have:

$$\lim_{n\to\infty} P\{\xi_n = y\} = \lim_{n\to\infty} \sum_x q_x \cdot p_{xy}^{(n)} = \sum_x \lim_{n\to\infty} q_x \cdot p_{xy}^{(n)} = \sum_x q_x \cdot p_y = p_y \cdot \sum_x q_x = p_y. \tag{26.19}$$

Why can we conclude that the limit of the sum is equal to the sum of the limits for this *infinite* sum? An infinite sum is an integral (with respect to the counting measure #); and we can apply our usual theorems about integrals. Here the integrands (the summands) $q_x \cdot p_{xy}^{(n)}$ for every n are dominated by an integrable function, namely, $q_x \cdot p_{xy}^{(n)} \le q_x$, which is integrable: $\sum_{x\in X} q_x = 1 < \infty$.

All of this was deduced from the equalities (26.3); only I can never remember which of these two equalities to use. So I write both; and see which of the two equalities is clearly useless – so you are to use the remaining one.

So under the conditions of Theorems 26.1 – 26.2 the transition probabilities $p_{xy}^{(k)}$ have limits as $k \to \infty$, and these limits don't depend on x; but *how to find* these limits? Again the equalities (26.3) help, and again I cannot remember which one of the two.

Again we restrict ourselves to finite chains (i. e., $X = \{x^1, ..., x^m\}$).

Let us take the limits of its both parts of the equalities $p_{xy}^{(k)} = \sum_z p_{xz} \cdot p_{zy}^{(k-1)}$ as k (and also $k-1$) goes to ∞:

$$\lim_{k\to\infty} p_{xy}^{(k)} = p_y = \sum_z p_{xz} \cdot \lim_{k\to\infty} p^{(k-1)} p_{zy} = \sum_z p_{xz} \cdot p_y = p_y. \qquad (26.20)$$

So the only thing that we get from this is that $p_y = p_y$ for every $y \in X$; but we knew it before. Let us try the second equality in (26.3):

$$p_{xy}^{(k)} = \sum_z p_{xz}^{(k-1)} \cdot p_{zy}, \qquad (26.21)$$

$$\lim_{k\to\infty} p_{xy}^{(k)} = p_y = \sum_z \lim_{k\to\infty} p^{(k-1)} p_{xz} \cdot p_{zy} = \sum_z p_z \cdot p_{zy}, \qquad y \in X. \qquad (26.22)$$

So we get a system of m linear equations $p_y = \sum_z p_z \cdot p_{zy}$ (one for each $y \in X$) with m unknowns p_y, $y \in X$. Does this system have a unique solution, allowing us to find the limiting probabilities p_y, $y \in X$? It turns out that no, and this is a very good thing.

Indeed, let us rewrite our system $p_y = \sum_z p_z \cdot p_{zy}$, $y \in X$, with all unknowns in the left-hand side; it is:

$$\sum_z (p_{zy} - \delta_{zy}) \cdot p_z = \mathbf{0}, \qquad y \in X: \qquad (26.23)$$

in the y-th equation, all coefficients by the unknowns p_z, $z \in X$, are equal to p_{zy}, except for $z = y$, which coefficient is equal to $p_{yy} - 1$. The matrix of this system is obtained from the (transposed) transition matrix by subtracting ones on the diagonal: the vector-matrix form of the equations (26.23) is:

$$\boldsymbol{p} \cdot (P - I) = \boldsymbol{p}, \qquad (26.24)$$

where $\boldsymbol{p}$ is the row vector of the unknowns p_y, $y \in X$ (and I is the identity matrix, with ones on the diagonal and zeros outside it).

If the solution of equation (26.24) (system (26.23) or, which is the same, (26.22)) were unique, it would be the *trivial* solution: $\boldsymbol{p} = \mathbf{0}$, $p_y = 0$, $y \in X$ ($\mathbf{0}$ is always a solution of a linear system with zero right-hand side); and $\boldsymbol{p} = \mathbf{0}$ definitely is *not* a probability distribution.

But luckily the solution is *not* unique: the matrix $P - I$ is singular.

How can we see it?

A finite square matrix $A = P - I$ is singular if and only if there exists a nonzero row vector $\boldsymbol{p}$ such that $\boldsymbol{p} \cdot A = \mathbf{0}$; and also if and only if a nonzero column vector $\boldsymbol{f}$ exists such that $A \cdot \boldsymbol{f} = \mathbf{0}$ (please excuse me for denoting with the same symbol $\mathbf{0}$ both the row and the column zero vectors).

In other words, there exists a non-zero left eigenvector $\boldsymbol{p}$ of the matrix P corresponding to the eigenvalue 1: $\boldsymbol{p} \cdot P = \boldsymbol{p}$ ($\boldsymbol{p}$ a row vector), if and only if there exists a non-zero *right* eigenvector $\boldsymbol{f}$ of the same matrix corresponding to the same eigenvalue ($\boldsymbol{f}$ a column vector).

And the vector equation

$$P \cdot \boldsymbol{f} = \boldsymbol{f}, \tag{26.25}$$

which means, in vector components f_y,

$$\sum_y p_{xy} \cdot f_y = f_x, \qquad x \in X, \tag{26.26}$$

definitely has a nonzero solution; namely, the column vector $\mathbf{1}$ with all components equal to 1: we have

$$\sum_y p_{xy} \cdot 1 = 1, \qquad x \in X, \tag{26.27}$$

$$P \cdot \mathbf{1} = \mathbf{1}. \tag{26.28}$$

So the rank of the matrix $P - I$ is less than its order m, and there *is* a nonzero solution $\boldsymbol{p} = (p_y)_{y \in X}$ of the system $\sum_z p_z \cdot p_{zy} = p_y$.

It turns out (I am *not* giving the proof) that for a finite ergodic chain the rank of the matrix $P - I$ is equal to $m - 1$; so if we add one equation to this system with coefficients that are linearly independent with the columns of this matrix, the solution becomes unique. As this additional equation we take the obvious one: $\sum_y p_y = 1$, or $\boldsymbol{p} \cdot \mathbf{1} = 1$.

So the limiting probabilities are found as the unique solution of the following system of $m + 1$ equations with m unknowns:

$$\begin{cases} \sum_z p_z \cdot p_{zy} = p_y, & y \in X, \\ \sum_y p_y = 1. \end{cases} \tag{26.29}$$

50 For the Markov chain with the transition matrix (23 – 24.24), find the solution $\boldsymbol{p} = (p_0, ..., p_4)$ of the system $\boldsymbol{p} \cdot P = \boldsymbol{p}$, $\sum_{y=0}^{4} p_y = 1$.

51 Let us change the last example in Lectures 23 – 24 taking the remainder after dividing by 6. Find the corresponding transition matrix P. Find the solution $\boldsymbol{p}$ of the system $\boldsymbol{p} \cdot P = \boldsymbol{p}$, $\boldsymbol{p} \cdot \mathbf{1} = 1$. Is $\lim_{k \to \infty} p_{xy}^{(k)} = p_y$, $x, y = 0, 1, ..., 5$?

52* For a finite ergodic Markov chain, is the solution of the system $\boldsymbol{p} \cdot P = \boldsymbol{p}$, $\boldsymbol{p} \cdot \mathbf{1} = 1$ necessarily unique?