Lecture 1.

I started on the theme of Markov processes in the probability theory course MATH 755 taught in the fall of 2008. Since most of my new students did not take this course, I'll remind the necessary facts and refer them to the Fall 2008 lecture notes.

I started with the case of *discrete Markov chains*: that of Markov processes with both the time and the space discrete (**2008.23** − **24** − **2008.26**).

Let X be a countable set (*countable* including *finite*). A sequence $\xi_0, \xi_1, \xi_2, ..., \xi_n, ...$ of random variables taking values in the space X (which we'll call *the phase space*) is called a Markov chain if for all $n, x_0, x_1, ..., x_n, x_{n+1} \in X$

$$P\{\xi_{n+1} = x_{n+1} | \xi_0 = x_0, \xi_1 = x_1, ..., \xi_{n-1} = x_{n-1}, \xi_n = x_n\} = P\{\xi_{n+1} = x_{n+1} | \xi_n = x_n\}, \tag{1.1}$$

provided that the probability of the condition is positive:

$$P\{\xi_0 = x_0, \xi_1 = x_1, ..., \xi_{n-1} = x_{n-1}, \xi_n = x_n\} > 0. \tag{1.2}$$

If the probability of the condition is equal to 0, we require *nothing*. (See formula (2008.23 − 24.15).)

This definition expresses the general Markov idea of the *future* (what happens at time $n+1$, in the present case) depends on the *past* (what happens up to time n) only through the *present* (the value ξ_n of our stochastic process – or rather *sequence* – at time n).

We can introduce the matrix $P_k = (p^k_{xy})_{x, y \in X}$ (called *the matrix of transition probabilities of the Markov chain at the k-th step*, taking $p^k_{xy} = P\{\xi_k = y | \xi_{k-1} = x\}$; we can rewrite the right-hand side of (1.1) as $p^{n+1}_{x_n\, x_{n+1}}$. Each transition matrix P_k is a *stochastic matrix*, i. e. it has nonnegative entries, and the sum $\sum_y p^k_{xy}$ over every row is equal to 1.

It can be proved starting from the definition (1.1) that for an arbitrary $m \ge n$

$$P\{\xi_m = y | (\xi_0, \xi_1, ..., \xi_{n-1}) \in C, \xi_n = x\} = P\{\xi_m = y | \xi_n = x\} = p^{nm}_{x_n\, y}, \tag{1.3}$$

where $P^{nm} = (p^{nm}_{xy})_{x, y \in X}$ is another stochastic matrix (the transition matrix from time n to time m; see formula (2008.25.20)). This is another manifestation of the same Markov idea: the *future* is now what is happening at time m, where m is not necessarily equal to $n+1$.

It turns out that the matrix P^{nm} is a product of the P_k matrices we started with:

$$P^{nm} = \prod_{k=n+1}^{m} P_k = P_{n+1} \cdot P_{n+2} \cdot ... \cdot P_m \tag{1.4}$$

(formula (2008.25.26)); and from this we get quite naturally for $n \le m \le r$

$$P^{nr} = P^{nm} \cdot P^{mr}, \tag{1.5}$$

which means that

$$p_{xy}^{nr} = \sum_{z \in X} p_{xz}^{nm} \cdot p_{zy}^{mr} \tag{1.6}$$

(formulas (2008.25.23), (2008.25.24))

We considered mostly the case of time-homogeneous Markov chains for which the transition matrix P_k is the same at every step: $P_k = P$. In this case the formula (1.4) becomes $P^{nm} = P^{m-n}$ (the powers of the same matrix), and the transition probabilities p_{xy}^{nm} depend only on the difference $m - n$: $p_{xy}^{nm} = p_{xy}^{(m-n)}$.

Now let us go to general Markov processes, for which time and space are not necessarily discrete (lecture notes 2008.32 and the following).

First of all, while we can characterize probability distributions on discrete (countable) spaces with probabilities of the random variable being in one-point sets (what they call the *probability mass function*), we cannot do it inn the uncountable case, and we have to consider probabilities of the random variable being in a *set* as a function of this set (a *distribution* being *a measure*).

Secondly, of we consider a stochastic process – a family ξ_t of random variables depending on a parameter t taking values in some subset T of the real axis, interpreted as *time*, – we cannot hope to build our definitions by analogy to formula (1.1): this formula has to do with what happens with our Markov chain at *the first time* $n+1$ *that is greater than* n; and if T is, for example, the right half-line $[0, \infty)$, there is *no such thing* as *the smallest time that is greater than the given* t. So we'll be imitating formula (1.3) rather than (1.1).

The third difficulty is that the typical situation in the case of uncountable space of time is that all probabilities similar to (1.2),

$$P\{\xi_s = x_s,\, s \le t\} \tag{1.7}$$

are equal to 0; and in this case we cannot use the classical conditional probabilities.

However, we have the concept of conditional probabilities (and conditional expectations) *with respect to a* σ*-algebra* $\mathcal{A} \subseteq \mathcal{F}$, or *with respect to* one, or more, or infinitely many *random variables* (see Lecture Notes **2008.29**, **2008.30**, and also the one-page **2008.Existence**).

So here is the definition of a Markov process: Let ξ_t, $t \in T \subseteq \mathbb{R}^1$, be a stochastic process (a family of random variables depending on the time parameter t) taking values in a measurable space $(X, \mathcal{X})$ (a *measurable space* is a couple consisting of a set (space) X and a σ-algebra $\mathcal{X}$ of its subsets; a random variable with values in $(X, \mathcal{X})$ is a function ξ on our sample space Ω with values in X, such that for every set $C \in \mathcal{X}$ its inverse image $\xi^{-1}(C) = \{\omega\colon\, \xi(\omega) \in C\}$ belongs to our fundamental σ-algebra$\mathcal{F}$ in the sample space – i. e. is an *event*). We say that ξ_t, $t \in T$, is a Markov process if for every t, $u \in T$, $t \le u$, and every set $C \in \mathcal{X}$ almost surely

$$P\{\xi_u \in C \| \xi_s,\, s \le t\} = P\{\xi_u \in C \| \xi_t\}. \tag{1.8}$$

Let me remind you that the conditional probability with respect to a random variable or to a family of random variables is defined as the conditional probability with respect to the σ-algebra in the sample space generated by these random variables:

$$P\{\xi_u \in C \| \xi_s,\, s \le t\} = P\{\xi_u \in C \| \sigma(\xi_s,\, s \le t)\}, \tag{1.9}$$

$\sigma(\xi_s,\, s \le t)$ being the smallest σ-algebra containing all events $\{\xi_s \in C\}$, $s \le t$, $C \in \mathcal{X}$ (see (2008.29.12)). Note that the σ-algebra generated by *one* random variable ξ_t consists just of all events of the form $\{\xi_t \in C\}$:

$$\sigma(\xi_t) = \big\{\{\xi_t \in C\}\colon\, C \in \mathcal{X}\big\}. \tag{1.10}$$

In the equality (1.3) we had a third term, p^{nm}_{xy}. Can we introduce a third term in our definition (1.8)?

In the case of a discrete space X and $T = \{0,\, 1,\, 2,\, ...,\, n,\, ...\}$ the right-hand side here should be $\sum_{y \in C} p^{tu}_{\xi_t\, y}$ (t and u being integers); what in the general case?

Unfortunately conditional probabilities with respect to σ-algebras are defined not in a unique way, but only in an *almost-unique*. We could, starting with the definition (1.8), try to choose one version of these conditional probabilities so that it has certain good properties; but better we go another way.

We'll call a function $P(t,\, x,\, u,\, C)$ of four arguments t, x, u, C, $t,\, u \in T$, $t \le u$, $x \in X$, $C \in \mathcal{X}$, a *Markov transition function* if

1) for any $t \le u$ and $C \in \mathcal{X}$ the function $P(t,\, \bullet,\, u,\, C)$ (that is, the function $P(t,\, x,\, u,\, C)$ as the function of its second argument, for fixed t, u, and C) is $\mathcal{X}$-measurable;

2) for any $t \le u$ and $x \in X$ the function $P(t,\, x,\, u,\, \bullet)$ is a measure on the σ-algebra $\mathcal{X}$, and $P(t,\, x,\, u,\, X) = 1$;

3) for $u = t$ the measure $P(t,\, x\, t,\, \bullet)$ is concentrated at the point x: $P(t,\, x\, t,\, C) = \delta_x(C) = 1$ for $C \ni x$, and $= 0$ for $C \not\ni x$ (this is the analogue of the transition matrix P^{nn} for a discrete Markov chain being the identity matrix: $P^{nn} = I$, $p^{nn}_{xy} = \delta_{xy}$, the Kronecker symbol);

4) for $s \le t \le u$ the Chapman – Kolmogorov equation is satisfied: for every $C \in \mathcal{X}$

$$P(s,\, x,\, u,\, C) = \int_X P(s,\, x,\, t,\, dy)\; P(t,\, y,\, u,\, C) \tag{1.11}$$

(this is the analogue of equalities (1.5), (1.6)).

We say that ξ_t, $t \in T$, is a Markov process *with the transition function* $P(t,\, x,\, u,\, C)$ if for all $t \le u$, $x \in X$, $C \in \mathcal{X}$ almost surely

$$P\{\xi_u \in C \| \xi_s,\, s \le t\} = P\{\xi_u \in C \| \xi_t\} = P(t,\, \xi_t,\, u,\, C). \tag{1.12}$$

We are going to consider only Markov processes *having a transition function.*