University of Oslo / Department of Economics / NCF 2009–02–10

# ECON5160: On the Feb. 06 lecture

This note will touch and review a few essential concepts from TK ch. III and IV, although its main purposes are to

1. correct an error given in the lecture
2. supplement with a couple of pieces of information not given thoroughly in the lecture
3. explain how the theory of irreducible processes can be applied to equivalence classes of communicating states separately, and how to deal with transition between equivalence classes.

## 1 Erroneous definition of aperiodicity

A correct definition is that a state $i$ is aperiodic if

$$\mathsf{Pr}[X_n = i | X_0 = i] > 0 \quad \text{for all } n \text{ large enough.}$$

(The book uses a slightly different definition though, but this is equivalent: The $\Rightarrow$ part is given in TK p. 239 property 2, and the $\Leftarrow$ part follows from this property by picking $N$ and $N + 1$ and applying the definition of periodicity.)

## 2 A few bits and pieces

- For a few of the properties, they are said to hold for the Markov chain itself if they hold for all states. (E.g. $X$ is «aperiodic» if each state is.) Also, they are said to hold for communications equivalence classes if they hold for each state in the class.
- The *regularity* property was used, but was it properly defined? It is the property that for some $n$ large enough, $\mathbf{P}^n$ has only positive entries.
- Problem IV 3.2. I should have mentioned the result – a finite state aperiodic irreducible Markov chain is regular and recurrent. The proof is left as an exercise – it is maybe a bit abstract though.
- If the state space is finite, there has to be some recurrent state.
- Furthermore, null recurrence is only possible when the state space is infinite.
- For a positive recurrent irreducible Markov chain, there is a *unique* stationary distribution.
- From linear algebra: a matrix $\mathbf{\Pi}$ with all rows equal (to, say, $\boldsymbol{\pi}$) can be written as $\mathbf{1}\boldsymbol{\pi}$ where $\mathbf{1}$ is a *column* vector.

## 3 Reducible Markov chains

We will give an example with a process $X$ living on a finite state space decomposable into three communicating classes, a generalization of the example given in TK p. 262, where

- each class is a «communication» equivalence class,
- classes $C_1$ and $C_3$ are recurrent but will each trap the process, and
- class $C_2$ is transient.

Then the transition matrix $\mathbf{P}$ for $X$ can be written as

$$\mathbf{P} = \begin{bmatrix} \mathbf{A} & \mathbf{0} & \mathbf{0} \\ \mathbf{B} & \mathbf{C} & \mathbf{D} \\ \mathbf{0} & \mathbf{0} & \mathbf{E} \end{bmatrix} \quad \begin{matrix} \} \text{ class } C_1 \\ \} \text{ class } C_2 \\ \} \text{ class } C_3 \end{matrix} \tag{1}$$

where all the boldfaced symbols are matrices. Then, in ten steps:

I) The representation (1) is assumed chosen so that the matrix dimensions match – e.g. $\mathbf{A}$, $\mathbf{B}$ and the lower-left $\mathbf{0}$ have the same width (= number of columns).

II) We assume that $\mathbf{A}$ and $\mathbf{E}$ have no null-columns (any such can be incorporated in the «$\mathbf{0}$»). Of course they have no null-rows either.

III) For class $C_1$ to trap the process, $\mathbf{A}$ (now assumed to contain no null-columns!) must be at least as high as wide; if not, it means that some state outside $C_1$ is accessible from some state in $C_1$.
Same argument goes for $\mathbf{E}$.

IV) For class $C_1$ to be an equivalence class (in the sense of communication), then under the «trap» assumption it must be at least as wide as high – if not, there is some state $i$ in $C_1$ inaccessible from the other states in $C_1$, since by the trap assumption, $X$ cannot reach $i$ by exiting $C_1$.
Same argument goes for $\mathbf{E}$.

V) For this three-class case we may therefore assume $\mathbf{A}$ and $\mathbf{E}$ to be square. Then $\mathbf{C}$ is also square, and describes the one-step transition probabilities from states in $C_2$ to states in $C_2$.

VI) Observe that the matrix powers $\mathbf{P}^n$ of $\mathbf{P}$ also has the structure given in (1), with zeroes as indicated:

$$\mathbf{P}^n = \begin{bmatrix} \mathbf{A}_n & \mathbf{0} & \mathbf{0} \\ \mathbf{B}_n & \mathbf{C}_n & \mathbf{D}_n \\ \mathbf{0} & \mathbf{0} & \mathbf{E}_n \end{bmatrix} \quad \begin{matrix} \} \; C_1 \\ \} \; C_2 \\ \} \; C_3 \end{matrix} \tag{2}$$

where $\mathbf{A}_n = \mathbf{A}^n$ (matrix power), $\mathbf{C}_n = \mathbf{C}^n$ and $\mathbf{E}_n = \mathbf{E}^n$ (but not similarly for $\mathbf{B}_n$ and $\mathbf{D}_n$).

VII) From the first time hitting $C_1$, the Markov property ensures that $X$ will have the same probability distributions as a process $Y$ which has $\mathbf{A}$ as transition matrix. Similar goes for $\mathbf{E}$ as the transition matrix for a process $Z$ (from the time $X$ hits $C_3$). Then the properties of $Y$ and $Z$ may be studied standalone, and the main issue concerning the reducibility property is «how to deal with transient states» (i.e. $C_2$).

VIII) Let $T$ be the first exit time from $C_2$. Let $\mathbf{U}$ be the matrix of probabilities

$$u_{ij} = \mathsf{Pr}[X_T = j | X_0 = i] \quad \text{for } i \in C_2,\ j \in C_1,$$

and let similarly $\mathbf{V}$ have entries indexed over $j \in C_3$

$$v_{ij} = \mathsf{Pr}[X_T = j | X_0 = i] \quad \text{for } i \in C_2,\ j \in C_3.$$

By first-step analysis,

$$\begin{aligned} \mathbf{U} &= \mathbf{B} + \mathbf{CU} \qquad \text{so that} \qquad \mathbf{U} = (\mathbf{I} - \mathbf{C})^{-1}\mathbf{B} \\ \mathbf{V} &= \mathbf{D} + \mathbf{CV} \qquad \text{so that} \qquad \mathbf{V} = (\mathbf{I} - \mathbf{C})^{-1}\mathbf{D} \end{aligned}$$

(the transience of $C_2$ will grant invertibility.) If we merely want the probabilities $u_i$ that $X$ will hit (somewhere in) $C_1$, that will be the «sum over $j$ for each $i$», namely

$$\mathbf{u} = \mathbf{U1} = (\mathbf{I} - \mathbf{C})^{-1}\mathbf{B1} \qquad \text{and similarly} \qquad \mathbf{v} = \mathbf{V1} = (\mathbf{I} - \mathbf{C})^{-1}\mathbf{D1},$$

where $\mathbf{1}$ is the column vector of ones.[1]

IX) By recurrence, we can find unique stationary distributions $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{A}$ for $Y$ and $\boldsymbol{\zeta} = \boldsymbol{\zeta}\mathbf{E}$ for $Z$. Then the time averages (i.e. mean occupation ratios) converge, and to[2]

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m<n} \mathbf{P}^m = \begin{bmatrix} \mathbf{1}\boldsymbol{\pi} & \mathbf{0} & \mathbf{0} \\ \mathbf{u}\boldsymbol{\pi} & \mathbf{0} & \mathbf{v}\boldsymbol{\zeta} \\ \mathbf{0} & \mathbf{0} & \mathbf{1}\boldsymbol{\zeta} \end{bmatrix}. \tag{3}$$

X) The limiting distributions may or may not exist, depending on the convergence or not of $\lim_n \mathbf{A}^n$ and $\lim_n \mathbf{E}^n$. If they converge, the corresponding entry is given as in (3). In the TK example, $\mathbf{A}$ is regular so $C_1$ (i.e. $Y$) is aperiodic and $\boldsymbol{\pi}$ is a limiting distribution for $Y$, but $C_3$ is periodic. Then $\lim_n \mathbf{A}^n$ converges and its limit $\mathbf{\Pi}$ is $= \mathbf{1}\boldsymbol{\pi}$, and furthermore $\mathbf{u}\boldsymbol{\pi} = \mathbf{U1}\boldsymbol{\pi} = \mathbf{U\Pi}$. However, $\lim_n \mathbf{E}^n$ diverges. Then formula in TK top of p. 263 corresponds to

$$\mathbf{P}^n \text{ «tends to» } \begin{bmatrix} \mathbf{1}\boldsymbol{\pi} & \mathbf{0} & \mathbf{0} \\ \mathbf{u}\boldsymbol{\pi} & \mathbf{0} & \star \\ \mathbf{0} & \mathbf{0} & \star \end{bmatrix} = \begin{bmatrix} \mathbf{\Pi} & \mathbf{0} & \mathbf{0} \\ \mathbf{U\Pi} & \mathbf{0} & \star \\ \mathbf{0} & \mathbf{0} & \star \end{bmatrix} \qquad \text{(«}\star\text{» for divergent submatrices)} \tag{4}$$

[1] The corresponding relation on TK p. 262 are $u_2 = \frac{1}{3}(1) + 0(1) + \ldots$ and $u_3 = \frac{1}{6}(1) + \frac{1}{6}(1) + \ldots$ – observe from the next line therein that the constant coefficient is the sum of the «(1)» coefficients for each row.

[2] The corresponding formula in TK is middle p. 263. Note that the «center» matrix is null since transience of $C_2$ implies $\lim_n \mathbf{C}^n = \mathbf{0}$.