# MATH3200: APPLIED LINEAR ALGEBRA
# SELF-STUDY AND PRACTICE MODULE 12: ESTIMATING TRANSITION PROBABILITIES

WINFRIED JUST, OHIO UNIVERSITY

We will use the terminology and notation of Conversations 7 and 8.

In Module 11 we considered a sample space $\{S, R, O\}$ with elementary outcomes $S$ (sunshine), $R$ (rain), and $O$ (dry, but overcast). We then estimated probabilities of these elementary outcomes in a hypothetical location based on a sequence of 56 given observations. These probabilities represent *long-range averages.* However, as you learned in Conversation 8, *if* we observe a rainy day today, then the probability of "sunny day tomorrow," the *transition probability* from a rainy day today to a sunny day tomorrow, of may not be the same as the long-range average probability of observing a sunny day. These transition probabilities are needed in the construction of Markov chain models for weather forecasting, and in this module you will learn how to estimate them from data.

Let us illustrate the procedure first with an example as in the conversations, where we have only two elementary outcomes that will also be the states in our Markov chain:

- State 1: Elementary outcome $S$, "sunny day."
- State 2: Elementary outcome $R$, "rainy day."

Then we need to determine the transition probabilities $p_{ij}$ between these states that were discussed in Conversations 7 and 8. They have the following meanings:

- $p_{11}$ be the probability that if day $t$ is a sunny day, then day $t+1$ will also be sunny.
- $p_{12}$ be the probability that if day $t$ is a sunny day, then day $t+1$ will be rainy.
- $p_{21}$ be the probability that if day $t$ is a rainy day, then day $t+1$ will be sunny.
- $p_{22}$ be the probability that if day $t$ is a rainy day, then day $t+1$ will also be rainy.

Consider the following sequence of hypothetical observations:

$SSRSSSRRSRSSSSSSSSSRRRRSRSSSSRSSSSRSSSRSSRRSSRRRRSSSSRRSS$

Let us start with estimating the transition probability $p_{21}$ that if day $t$ is a rainy day, then day $t+1$ will be sunny. In the above sequence, we have 20 rainy days. Of those, 11 are followed by a sunny day. So we would estimate $p_{21} = \frac{11}{20}$.

Similarly, 9 of these 20 rainy days are followed by another rainy day, so we estimate $p_{21} = \frac{9}{20}$. Note that $p_{21} + p_{22} = 1$. This is exactly as it should be, since every rainy day must be followed either by a sunny day or another rainy day. The matrix $\mathbf{P}$ of transition probabilities must be a *stochastic matrix,* a matrix whose entries are probabilities and whose rows all sum up to 1.

Now let us estimate $p_{11}$ and $p_{12}$. The given sequence of observations contains 36 sunny days. Of those, 24 are followed by another sunny day, 11 are followed by a rainy day, and for one of them we don't have a record of what happened on the next day. Thus we can only use the first 35 of these 36 sunny days in our estimates of transition probabilities. We get $p_{11} = \frac{24}{35}$ and $p_{12} = \frac{11}{35}$. Again we have $p_{11} + p_{12} = 1$, as required.

We have found the following matrix of estimated transition probabilities:

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = \begin{bmatrix} 24/35 & 11/35 \\ 11/20 & 9/20 \end{bmatrix}$$

Notice that this is a stochastic matrix.

**Question 12.1:** Which of the following $2 \times 2$ matrices are stochastic? For each matrix that you identified as not being stochastic, briefly explain why it is not.

$$\mathbf{P}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \mathbf{P}_2 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad \mathbf{P}_3 = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \quad \mathbf{P}_4 = \begin{bmatrix} 0.5 & 0.6 \\ 0.5 & 0.4 \end{bmatrix} \quad \mathbf{P}_5 = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$$

$$\mathbf{P}_6 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \quad \mathbf{P}_7 = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \quad \mathbf{P}_8 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \mathbf{P}_9 = \begin{bmatrix} 0 & 1 \\ 0.5 & 0.5 \end{bmatrix} \quad \mathbf{P}_{10} = \begin{bmatrix} 0.4 & 0.6 \\ 0.5 & 0.5 \end{bmatrix}$$

Now consider a Markov chain with the following three states:

- State 1: Elementary outcome $S$, "sunny day."
- State 2: Elementary outcome $R$, "rainy day."
- State 3: Elementary outcome $O$, "dry, but overcast day."

Suppose on 56 consecutive days we recorded the following observations about the weather:
$SSROOSRRORSOSSSSOSRRRRSROOOOROOSSROOORSORRSSRRRRSOOORROS$

**Question 12.2:** Find the matrix of transition probabilities for this Markov chain that would represent the best estimates from the given data. Do not convert your answers into decimal fractions.