AN INTRODUCTION TO MARKOV CHAINS FOR INTERESTED HIGH SCHOOL STUDENTS

S. HALVERSCHEID AND P. SIBBERTSEN

Abstract. Markov Chains are introduced by only assuming some knowledge of the notion of probability. The modelling of a situation in a context of biology gives the opportunity to students to approach various concepts of probability theory themselves.

1. Modelling prompting stochastic thinking

At the beginning of a course introducing the basic concepts of probability theory one is tempted as a teacher to introduce axioms and several models for stochastic experiments. This approach is classical and well tried.
Often, the modelling is rather mentioned in order to legitimatize parts of a theory than developed from a topic as a thread. Some of the classical models (e.g. urns or gambling models) are presented in such a close way to the theory that it seems almost artificial to call them models. These may also have doubtful consequences for the students' intuition as Henze points out ([H]).
In this survey, we suggest a "medias in res"-approach for introducing Markov chains to mathematically interested high school students. Students of that level are able to work out models for probability theory themselves. Of course, this is not meant to replace entirely the classical way the basics of stochastics are introduced. We aim at encouraging the students to find new models for several concepts in probability theory within a Markov chain problem coming from biology. This topic is chosen both as a challenge making further assumptions and explanations in the modelling necessary and as an example deserving some interest in its own right. We have not used elementary gambling models deliberately for the reasons mentioned above. The transfer to those is straightforward however and seems appropriate depending on the students' background.
The guideline for them is given here by questions which were given to students in written or oral form. This concept has been carried out in 4 groups of 15 national winners (coming from the Netherlands, Poland, Hungary, Czech Republik and Germany) of the mathematical Kangaroo competition and twice with groups of the *Saturday University* at the Universities of Bochum and Dortmund, Germany. The reactions of the students, which were examined with the help of questionnaires, are illustrated in the last section.

2. Bacteria cultures

In order to explain what Markov chains are about we consider first a simple model for the growth of a bacteria culture. Later, we present a more realistic, but still idealized model which investigates the fluctuation of gene frequency in a population under the influence of selection.

Assumptions:
Our bacteria culture in a liquid fertilizer should be modelled under the following rules:

(1) The liquid fertilizer does only provide enough energy for one bacteria cell to start the duplication process in a time interval. The remaining bacteria are assumed not to change during that time.
(2) Duplication often fails: On the long run, one expects among the cells starting the duplication process roughly as many cells to double successfully as to die during this process.
(3) The overall number of bacteria is limited (to N, say).

Question 1. Considering the process of duplication of a bacteria cell we want to attribute probabilities to the case that the cell doubles successfully and to the case that the cell dies. What does assumption (2) mean in terms of percentages of the likelihood for duplication? What is the possible range of values of a probability?

Question 2. Under all these assumptions: Which questions about the model do you regard worthwhile examining?

Temporary Assumption: We take now even a very small number of bacteria which can exist, at most: $N = 4$. This is not realistic at all at this stage, but we hope it helps to illustrate what Markov chains are about.
We want to discuss whether it is possible that the bacteria die out or that they reach the maximal number possible $N = 4$ when the duplication process is started again and again as long as none of these cases is reached.

Question 3. Give an example of a chain of such processes such that the bacteria never die out but do not reach the maximal number 4 neither.

The number of bacteria in our population can be 0; 1; 2; 3 or 4. We call these the *states*. The state at the beginning, i. e. the number of bacteria at the beginning, is called *initial state*. The sequences of states in the order of their occurrence are called *path* or *trajectory*. If it is possible to come to a state j from a state i the state j is called *attainable* from i. In our example the states 2 or 3 are attainable from 1 but not from 0, because the state 0 means that the bacteria died out.

Question 4. Give one possible path that the bacteria die out after exactly five generations. Give a path that bacteria die out after exactly four generations.

Beginning with the initial state 1 we can go up to the final state 4 for example with the path given in figure 1.

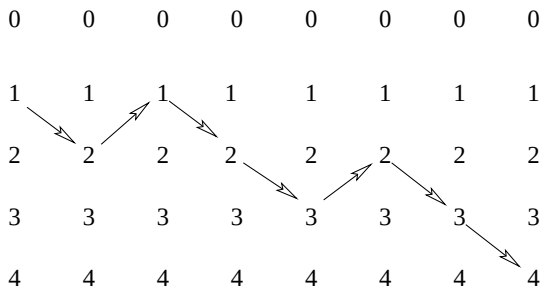

FIGURE 1. Path of states for the Markov chain

Here, the series ends in state four after seven generations because the maximal number of bacteria is reached and no duplication process can be started anymore. The whole fluctuations up to the fifth generation can be displayed in the tree diagram in figure 2.

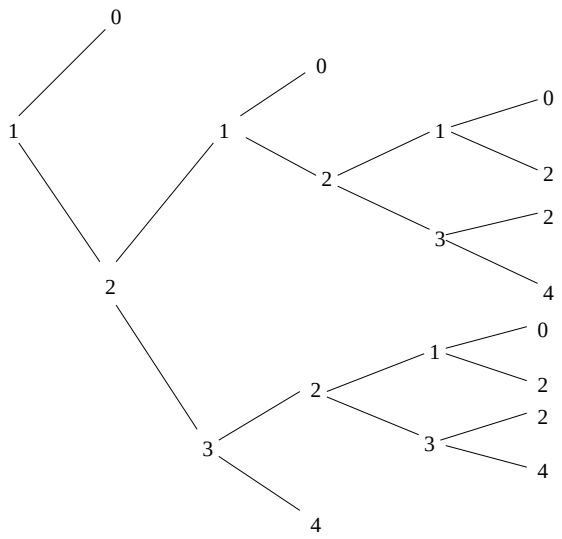

FIGURE 2. Fluctuation of the number of bacteria

The probabilities for the several transitions between the states are called *transition probabilities*.

Question 5. What are the transition probabilities in our example?

This diagram is quite confusing as soon as a bigger number of generations must be displayed. But there are two properties of transition probabilities which allow us to handle this in a much clearer way.

(1) The probability of the transition does not depend on the history of the bacteria culture up to this point. (This property is referred to as *homogeneity*.)

(2) The transition probability from state i to state j does not depend on the states before i. (This is called *Markov property*.)

In order to make clear what these properties mean for our model let us consider the following questions:

Question 6. (1) What does it mean in respect of the above properties if the probability of a bacteria to die increases by 1% in each year of its life.

(2) Now let us assume that the energy for a bacteria to duplicate is provided by some food in the liquid and we have a fixed amount of food at the beginning of our experiment and no further food will be added during the experiment. If every bacteria cell now needs some amount of food in order to survive whether it duplicates or not, what does this means for the above properties?

Whenever one of the states 0 or 4 is reached no further fluctuations will be possible because the bacteria died out or do not have enough energy to start the duplication process. These states are called *absorbing*.
The transition probabilities together with an initial state describe a *Markov Chain*. A more formal way of describing the transitions from one state to another are *transition matrices* as they are called. These are actually a very useful tool when describing more complex Markov chains. Let us have a closer look at our previous tree diagram.

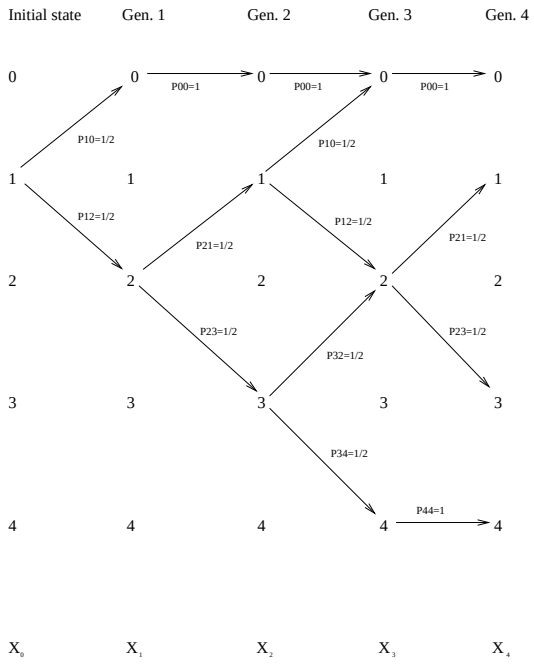

FIGURE 3. States with corresponding transition probabilities

Let M be the set of possible situations within a stochastic model. Each fluctuation - i. e. each element in M - is now described by (infinitely many) functions X_n, $n \in \mathbb{N}$, which give the state of the system in generation n. In other words: If one knows the number of bacteria at every step, one knows everything about the history of the bacteria.
The range of each of these random variables contains all possible states and thus is $\{0; 1; 2; 3; 4\}$. Let p_{ij} denote the probability to pass from state i to state j in the next single generation, where $i, j \in M$.

Question 7. It is certain that one passes from one state i to any other state $j \in M$. How is this expressed in terms of the transition probabilities p_{ij}?

Question 8. The transition probabilities p_{ij} are set up with respect to the next generation of successors. Try to work out the probabilities to pass from state i to state j in exactly two generations.

Question 9. Challenge: Try to work out the probabilities to pass from state i to state j in exactly k generations. (It is meant here to understand what sums determine the various probabilities not to write these down explicitly. Start with $k = 2, 3, 4$ to see a certain pattern.)

Instead of the tree diagram we have a more compact way of gathering this information by using a matrix containing all transition probabilities. Since we have five possible states the matrix consists of five rows and five columns. The transition matrix looks like this:

$$\begin{pmatrix} p_{00} & p_{01} & p_{02} & p_{03} & p_{04} \\ p_{10} & p_{11} & p_{12} & p_{13} & p_{14} \\ p_{20} & p_{21} & p_{22} & p_{23} & p_{24} \\ p_{30} & p_{31} & p_{32} & p_{33} & p_{34} \\ p_{40} & p_{41} & p_{42} & p_{43} & p_{44} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

This matrix gives the probability to go in one step from state i to state j. The probability to do this in two steps we obtain by multiplying the matrix P by itself.

Question 10. What does this mean? Give formulae for the entries of the matrix in two steps with the help of your answer to question 9.

In the same way we obtain the probability to go in n steps from state i to state j by considering the n-fold product of P with itself, P^n. This is why the transition matrix contains all necessary information to compute the transition probabilities for every number of steps.

Question 11. Let us consider the following questions for the fluctuations:

(1) What is the probability for reaching 0 or 4 (in an arbitrary number of steps)?

Hint: If it is too hard to compute P^n, try to find out the probability with which 0 or 4 are reached in step $1, 2, 3, \ldots$

(2) How many generations does it take in average until the bacteria dies out?

There are, of course, various ways to change the situation slightly. We just list a couple of them:

- What happens if the maximal number of bacteria is increased?
- What is the probability that the bacteria die out if the number of bacteria is unlimited?
- Try to vary the probability that a cell which started the duplication process dies!
- Introduce more options for the destiny of a cell in the duplication process: e. g. the baby cell could die with a certain probability but the mother cell survives.

3. A genetic model

For a more realistic example of Markov chains we consider an idealized genetic model introduced by S. Wright (cf. [K-T]). The basic idea of the model is to investigate the fluctuation of gene frequency in a population under the influence of selection. Let us in our model disregard effects as mutation or selective forces for simplicity. The population size is assumed to be fixed throughout every step in our model. Our population shall contain $2N$ individuals either having a type-a genetic characteristic or a type-A characteristic. The next generation is determined by $2N$ independent binomial trials. This means that one individual of the parent generation is randomly chosen and the successor is from the same genetic type.
Let us assume that the population contains j type-a individuals. All the other individuals are of type-A.

Question 12. Which probabilistic concept drives this model? What is the probability that a randomly chosen individual of the parents generation is of type-a or of type-A respectively?

The question arising in this genetic model is whether one or the other genetic type can die out at some time and if what is the probability for this. To answer this question we need an appropriate model for this problem.

Question 13. Argue why it makes sense to use Markov chains for modelling this problem.

Let us assume for simplicity that our population contains four individuals that is we have $N = 2$.

Question 14. (1) What are the possible states of the Markov chain?

(2) Are there absorbing states in this model? If yes, which are the absorbing states?

The next step to fully describe the Markov chain are the transition probabilities.

Question 15. Give the transition matrix in this model.

Question 16. Will always one genetic type die out? Or more exactly: Do we observe with a probability of 1 either of the states 0 or 4?
Hint: Try this again as in the previous section by first computing the matrix P^n.

4. The evaluation of the method by students

Teaching mathematically talented and interested students leads to special difficulties for the didactic reconstruction. Although all members of such groups are bright, a wide range of abilities and knowledge can be expected.
Since students of that kind are often very demanding – especially if they come to a course voluntarily at some cost of money and leisure time – the group is often very heterogeneous concerning their age, their mathematical background and their abilities.
In four groups in the International Kangaroo Maths Camp in September 2002 at Münster, Germany, and during the *Saturday University* at Bochum and Dortmund University, Germany, we asked 102 students after the course to evaluate their background knowledge in probability theory.
The answers in school marks from 1 (very good) to 5 (poor) were:

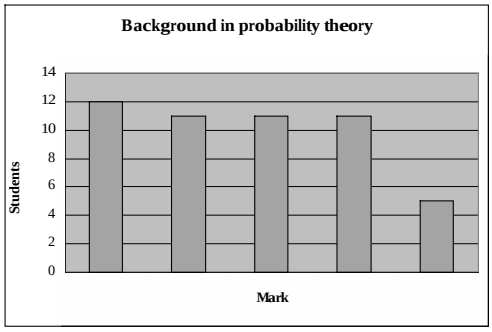

Figure 4. Self-assessment of statistical background

The attitudes towards Mathematics and its applications proved also to be quite diverse. This is also not surprising because there are very diverse reasons for taking part in Maths competitions. We asked the participants to rate whether they are much more interested in mathematics than in its applications. The mark 1 stands for a very affirmative answer and the mark 5 for the opposite.
The reactions on the course were unanimously positive. In the following tabular, some of the answers of the international students are listed. More than numbers,

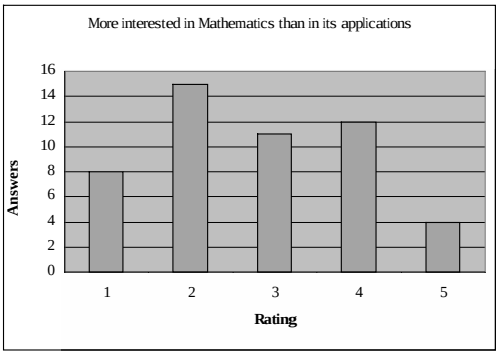

FIGURE 5. Appreciations of mathematical applications

we think that Mark's opinion describes best the course. The dutch student said, he rated the course on Markov chains best because "we had to do a lot ourselves." However, the actual numbers rating the course again from 1 = very good to 5 = poor are:

Table I *Evaluation of the course*

	1	2	3	4	5
The course was well structured	37	49	15	1	0
The examples helped me to understand the course better	40	41	14	6	1
Theory and applications were presented in a unified way	26	42	28	3	0
The participants were given the opportunity to discuss	38	37	21	4	1
There were sufficiently many exercises for the participants	21	49	26	5	0

We see clearly that exercises are most important for the students. Also our approach to present the theory of Markov chains by using applications close to real life situations have been rated well. Almost 60% of the students said that the course motivated them very much to get a deeper insight into the theory of Markov chains.

The approach sketched above shows the following advantages when considering the methodology of how probability theory is used here:

- Only a very limited amount of knowledge in probability theory needs to be assumed. Axioms and theory do not play an important role but can be developed naturally in this context (see, e. g., [R]).

- In particular, it is not necessary to discuss probabilities with certain conditions as well as a thorough discussion of the notion of stochastic independence.
- The exercises aim at encouraging students to construct other problems themselves. Although the exercises are given a clear aim and a restricted area of discussion, they do not have the character of finishing the subject. In the courses where this approach was tried it rather appeared to stimulate further studies.

With these properties, it offers many opportunities when chosen as the starting point of a course on probability theory. The following features of such a course could be linked to this course for instance:

- Various combinational questions and (discrete) distributions.
- The notion of independence of random variables.
- Conditional probabilities.
- Time series.
- Stopping time.
- Measure theory (In the basic example, e. g., there is an infinite path, which is possible, but which occurs to a probability of 0. Some of our students discovered this as a phenomenon themselves.)
- Statistics with data analysis - e. g. with records in a casino.

We also had good experiences with the following constructivist's approach following this course: Students should carry out the modelling for a certain situation, e. g. the expectancy of life of a goat in the desert if a lion is around. Having seen different models in this course, it seemed to us that students were confident to find good models and to impose appropriate conditions on the situation to work these models out.

Acknowledgements

The financial support of the Robert-Bosch foundation and the Ruhr-University Bochum with the price of innovation in teaching is gratefully acknowledged. The second author was also supported by the DFG under SFB 475. The authors thanks Katja Ickstadt for helpful discussions.

References

[Hen] Henry, M. (edt). (2001). *Autour de la modélisation en probabilités* Paris: Presses Universitaires Franc-Comtoises.

[H] Henze, N. (2000). Stochastic model building between the mathematics of games of chance and real application-orientation - a critical overview.
In: Hirscher, H. (edt.) (2000). *Modellbildung, Computer und Mathematikunterricht.* Hildesheim: Franzbecker. P. 49-57

[K-T] Karlin, S.; Taylor, H. (1975). *A first course in stochastic processes.* New York: Academic Press

[R] Resnick, S. L. (1994). *Adventures in stochastic processes.* Boston: Birkhäuser.

(S.H.) Westfälische Wilhelms-Universität, Institut für Didaktik der Mathematik, Einsteinstr. 62, D-48149 Mnster, Germany
E-mail address: stefan@halverscheid.net

(P.S.) Universität Dortmund, Fachbereich Statistik, Vogelpothsweg 87, D-44227 Dortmund, Germany
E-mail address: sibberts@statistik.uni-dortmund.de