

ПРИЛОЖЕНИЯ ЦЕПЕЙ МАРКОВА К БИОЛОГИЧЕСКИМ ЗАДАЧАМ

Вольвачёв Р.Т.

Хорошо известно, что в настоящее время роль математических методов в биологии возрастает. Поэтому задачей математического образования биологов является не только изучение базовых понятий математики, приёмов и методов, но и использование их для решения практических задач.

Поэтому в учебных пособиях (в том числе и изданных кафедрой Общей математики и информатики БГУ, см. например, [1]) приводятся уравнения изменения массы радиоактивного вещества (уравнения роста числа бактерий, прирост объема древесины и т.п.) и решаются задачи на нахождение скорости растворения лекарственных веществ, определение периода полураспада радия и радона и др. задачи, представляющие интерес в биологии.

В настоящей заметке мы рассмотрим понятие цепи Маркова и их простейшие свойства, а также применение Марковских процессов в генетике. Этот материал может быть использован на практических и семинарских занятиях, а также самостоятельных научных исследованиях, для выступлений на конференциях.

ЦЕПИ МАРКОВА

Определение. *Цепью Маркова* называется последовательность испытаний, в каждом из которых система принимает только одно из k состояний u_1, u_2, \dots, u_k полной группы, причём условная вероятность $p_{ij}(s)$ того, что в s -м состоянии система будет находиться в состоянии j , при условии, что после предыдущего $(s-1)$ -го испытания она находится в состоянии i , не зависит от результатов ранее произведенных испытаний.

Мы будем рассматривать только такие цепи Маркова, их условная вероятность $p_{ij}(s)$ – переход состояния i в состояние j – не зависит от номера испытания s . Поэтому вместо $p_{ij}(s)$ будем писать просто p_{ij} . (Такие цепи Маркова называют однородными).

Мы будем рассматривать дискретные цепи Маркова, когда изменение состояний системы происходит в определённые фиксированные моменты времени.

Итак, переходной вероятностью p_{ij} назовём условную вероятность того, что из состояния i в итоге следующего испытания система перейдёт в состояние j . Здесь первый индекс в обозначении p_{ij} указывает номер предшествующего, а второй – номер последующего состояния. Например, p_{11} – вероятность перехода из первого состояния в первое; p_{23} – из второго состояния в третье.

Определение. Пусть число состояний цепи Маркова конечно и равно k : A_1, A_2, \dots, A_k .

Матрицей перехода системы называется матрица, которая содержит все переходные вероятности этой системы:

$$P = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1k} \\ p_{21} & p_{22} & \dots & p_{2k} \\ \dots & \dots & \dots & \dots \\ p_{k1} & p_{k2} & \dots & p_{kk} \end{bmatrix}.$$

Так как эти состояния образуют полную систему событий, то сумма вероятностей каждой строки этой матрицы равна 1,

то есть
$$\sum_{j=1}^k p_{ij} = 1 \quad (i = 1, \dots, k).$$

Такие матрицы называются *стохастическими*.

Пример: Пусть A_1, A_2, A_3 – три состояния, в котором может находиться система.

Матрица перехода $P = \begin{bmatrix} 0,5 & 0,2 & 0,3 \\ 0,4 & 0,5 & 0,1 \\ 0,6 & 0,3 & 0,1 \end{bmatrix}.$

Здесь $p_{11} = 0,5$ вероятность перехода $i=1 \rightarrow j=1$.

$p_{21} = 0,4$ – вероятность перехода $i=2 \rightarrow j=1$.

Определение. Пусть $p_{ij}(n)$ – вероятность того, что в результате n шагов система перейдет из состояния i в состояние j .

Например. $P_{25}(10)$ – вероятность перехода за 10 шагов из второго состояния в пятое.

В частности, при $n=1$ получаем $p_{ij}(1) = p_{ij}$

Задача. Зная переходные вероятности p_{ij} , найти вероятности $p_{ij}(n)$ перехода за n шагов из состояния i в состояние j .

Оказывается, что зная все породные вероятности p_{ij} , т.е. зная матрицу P перехода из состояния i в состояние j за один шаг, можно найти вероятность $p_{ij}(n)$ перехода из состояния i в состояние j за n шагов, $n=2, 3, 4, \dots$

Используя теорему о полной вероятности, получаем:

Теорема. Матрица перехода $P(n)$ из состояния в состояние за n шагов равна:

$P(n) = P^n$. Мы полагаем, что $P(1) = P$.

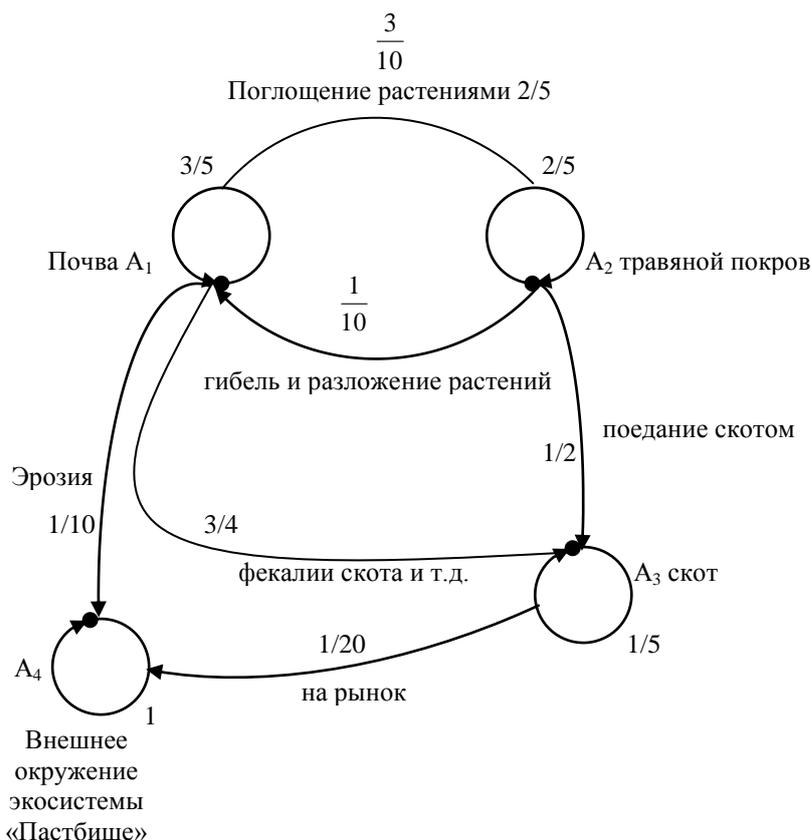
Пример. Пусть $P = \begin{bmatrix} 0,4 & 0,6 \\ 0,3 & 0,7 \end{bmatrix}$. Тогда

$$P(2) = \begin{bmatrix} 0,4 & 0,6 \\ 0,3 & 0,7 \end{bmatrix}^2 = \begin{bmatrix} 0,34 & 0,66 \\ 0,33 & 0,67 \end{bmatrix}.$$

Пример. Рассмотрим простейшую модель прохождения молекулы фосфора через экосистему «Пастбище». Для простоты примем во внимание четыре возможных состояния молекулы.

$$D = \begin{matrix} & A_1 & A_2 & A_3 & A_4 \\ \begin{matrix} A_1 \\ A_2 \\ A_3 \\ A_4 \end{matrix} & \begin{bmatrix} 3/5 & 3/10 & 0 & 1/10 \\ 1/10 & 2/5 & 1/2 & 0 \\ 3/4 & 0 & 1/5 & 1/20 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{matrix}.$$

В начальный момент молекула фосфора находится в почве. Далее, она может быть абсорбирована каким либо растением и перейти в травяной покров, выйти из экосистемы или остаться в почве.



Таким образом, находясь в одном из состояний, молекула может перейти в другие состояния. Однако, если она вышла из экосистемы, она больше в неё не возвращается.

Аналогичные модели можно строить для самых различных пестицидов или других элементов, например радиоактивных.

С помощью матрицы P можно найти среднее число дней до выхода фосфора из системы, при условии, что в начальном состоянии он находился в почве. Оно равно 15,6. Поэтому аналогичные вычисления, сделанные для двух других пестицидов, позволяют получить информацию, необходимую

для определения пестицида с возможно меньшей продолжительностью пребывания в экосистеме. Эта информация может быть использована для выбора альтернативного варианта. Отметим, что проблема использования пестицида ДДТ состоит в чрезвычайно длительном пребывании пестицидов в экосистеме после их внесения. Поэтому применение ДДТ у нас запрещено; значительно раньше запретили за рубежом.

Возвратимся к экосистеме «Пастбище».

$$\text{Имеем } P = \begin{bmatrix} 0,6 & 0,3 & 0 & 0,1 \\ 0,1 & 0,4 & 0,5 & 0 \\ 0,75 & 0 & 0,2 & 0,05 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

$$\text{Тогда } P_2 = P^2 = \begin{bmatrix} 0,390 & 0,300 & 0,150 & 0,160 \\ 0,475 & 0,190 & 0,300 & 0,035 \\ 0,600 & 0,225 & 0,040 & 0,135 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

$$P_3 = P^3 = \begin{bmatrix} 0,3765 & 0,2370 & 0,1800 & 0,2065 \\ 0,5290 & 0,2185 & 0,1550 & 0,0975 \\ 0,4125 & 0,2700 & 0,1205 & 0,1970 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Предположим, что продолжительность шага равна одному дню. Тогда вероятность выхода молекулы из системы (т.е. попадание в A_4) по происшествии трёх дней при условии, что в начале движения она была в почве (состояние A_1) определяется элементом (1,4) матрицы P^3 , т.е. она равна 0,2065.

Задачи. Задана матрица P вероятностей перехода цепи Маркова из состояния $A_i (i=1,2)$ в состояние $A_j (j=1,2)$ за один шаг. Найти матрицу $P(2)$ перехода из состояния A_i в состояние A_j за два шага если:

$$\text{а) } P = \begin{bmatrix} 0,1 & 0,9 \\ 0,2 & 0,8 \end{bmatrix}$$

$$\text{б) } P = \begin{bmatrix} 0,2 & 0,8 \\ 0,3 & 0,7 \end{bmatrix}$$

$$\text{в) } P = \begin{bmatrix} 0,3 & 0,7 \\ 0,4 & 0,6 \end{bmatrix}$$

$$\text{г) } P = \begin{bmatrix} 0,4 & 0,6 \\ 0,5 & 0,5 \end{bmatrix}$$

Возникают и другие важные вопросы при изучении цепей Маркова. Сформулируем некоторые из них.

Определение. Состояние i называется поглощающим, если $p_{ii}=1$.

Вопросы. Найти:

1. Вероятность перехода в поглощающее состояние A_j , при условии, что он начался в непоглощающем состоянии A_i .

2. Среднее число шагов до перехода процесса в некоторое поглощающее состояние при условии, что начальное состояние есть непоглощающее состояние A_i .

3. Среднее время пребывания процесса в непоглощающем состоянии u_i до его перехода до его перехода в некоторое поглощающее состояние при условии, что в начальный момент он находился в непоглощающемся состоянии u_i .

С помощью матрицы вероятностных переходов на эти вопросы можно дать ответы. (см.[1]).

ЗАДАЧА НА ДОМ (для интересующихся)

Всем хороша Земля Оз, но только не своим климатом. Здесь никогда не бывает двух ясных дней подряд. Если сегодня ясно, то завтра с одинаковой вероятностью пойдёт дождь или снег. Если сегодня снег (или дождь), то с вероятностью $\frac{1}{2}$ погода не изменится. Если все же она изменится, то в половине случаев снег заменяется дождем или наоборот, и лишь в половине случаев на следующий день будет ясная погода.

Проверить, что вероятностная матрица перехода имеет вид: (где в качестве состояний погоды примем различные виды погоды Д (Дождь), Я (Ясно), С (Снег)).

$$P = \beta \begin{matrix} & \begin{matrix} \text{Д} & \text{Я} & \text{С} \end{matrix} \\ \begin{matrix} \ddot{A} \\ \tilde{N} \end{matrix} & \begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{bmatrix} \end{matrix}.$$

$$\text{Проверить, что } P(3) = \beta \begin{matrix} & \begin{matrix} \text{Д} & \text{Я} & \text{С} \end{matrix} \\ \begin{matrix} \ddot{A} \\ \tilde{N} \end{matrix} & \begin{bmatrix} 26/64 & 13/64 & 25/64 \\ 13/32 & 6/32 & 13/32 \\ 25/64 & 13/64 & 26/64 \end{bmatrix} \end{matrix}.$$

Отсюда видно, что вероятность хорошей погоды через три дня после дождя равна 13/64, а после ясного дня 6/32.

$$P^2 = \beta \begin{matrix} & \begin{matrix} \text{Д} & \text{Я} & \text{С} \end{matrix} \\ \begin{matrix} \ddot{A} \\ \tilde{N} \end{matrix} & \begin{bmatrix} 0,437 & 0,187 & 0,375 \\ 0,375 & 0,250 & 0,325 \\ 0,375 & 0,187 & 0,437 \end{bmatrix} \end{matrix}.$$

$$P^4 = \beta \begin{matrix} & \begin{matrix} \text{Д} & \text{Я} & \text{С} \end{matrix} \\ \begin{matrix} \ddot{A} \\ \tilde{N} \end{matrix} & \begin{bmatrix} 0,402 & 0,199 & 0,398 \\ 0,398 & 0,203 & 0,398 \\ 0,398 & 0,199 & 0,402 \end{bmatrix} \end{matrix}.$$

$$P^8 = \beta \begin{matrix} & \begin{matrix} \text{Д} & \text{Я} & \text{С} \end{matrix} \\ \begin{matrix} \ddot{A} \\ \tilde{N} \end{matrix} & \begin{bmatrix} 0,400 & 0,200 & 0,400 \\ 0,400 & 0,200 & 0,400 \\ 0,400 & 0,200 & 0,400 \end{bmatrix} \end{matrix}.$$

С точностью до трёх десятичных знаков элементы матрицы P^8 составляют одинаковые строки.

Поэтому вероятность дождя через восемь дней после дождливого дня с большой точностью равна вероятности дождя после ясного дня или через восемь дней после снега.

II. ПРИМЕНЕНИЕ В ГЕНЕТИКЕ

Мы будем базироваться на теории наследственности, восходящей к Грегору Менделю, Монаху ордена св.Августина, опубликовавшему свою работу в 1866г. Проводя свои эксперименты, Мендель работал с растениями гороха. Эти растения имеют стручки двух цветов – зеленого и желтого. Некоторые сорта гороха «чисто» зеленые, другие «чисто» - желтые. К чисто зеленым сортам относят сорта гороха, которые, будучи окрещенными между собой, дают только зеленое потомство; аналогично – чисто желтые сорта. Мендель обнаружил, что каждый раз, когда скрещиваются растения чисто зеленого и чисто желтого сортов, все их потомки обладают лишь зелеными стручками. Если потом скрестить два потомка, то в следующем поколении примерно $\frac{3}{4}$ потомства будут иметь зеленые стручки; а около $\frac{1}{4}$ - желтые стручки.

Эти результаты экспериментов привели Менделя к следующей теории наследственности.

Простейший случай. Признак определяется одной парой чеков, каждый из которой может быть одного из двух типов. Например, G и d . Каждый индивидум может обладать генами в одной из трех возможных комбинаций

GG , dd Gd (с точки зрения генетики это равносильно dG).

Очень часто комбинации GG и Gd внешне неотличимы. В этом случае мы говорим, что ген G доминирует над геном d . В этом случае индивидум с генами GG называется доминантам, а индивидум с генами dd называется рецессивным, а индивидум с генами Gd – гибридом.

При скрещивании двух животных их потомство наследует по одному гену от каждого из родителей и в основу генетики кладётся предположение о том, что эти гены выбираются случайным обзором. Это предположение определяется вероятностью появления потомства каждого гена.

Так, потомство двух доминантных родителей должно быть доминантным, потомство двух рецессивных родителей должно быть

рецессивным, а потомство одного доминантного и одного рецессивного родителя должно быть гибридом.

При скрещивании доминантного животного и гибрида их потомство должно получить ген G от первого родителя, а от второго родителя с вероятностью $\frac{1}{2}$ ген G и с вероятностью $\frac{1}{2}$ ген d . Отсюда вероятности получения либо гибридного потомства, либо доминантного потомства равны. В то же время при скрещивании рецессивного родителя с гибридом вероятность получения рецессивного или гибридного потомства одинаковы. При скрещивании двух гибридов их потомство с вероятностью $\frac{1}{2}$ получает гены от каждого из родителей. Поэтому вероятность получены гены GG равна $\frac{1}{4}$, гены Gd – $\frac{1}{2}$, гены dd – $\frac{1}{4}$.

Рассмотрим процесс непрерывного скрещивания. В качестве одного из родителей в самом начале берём индивидуум с неизвестными генетическими признаками, а в качестве второго мы берём гибрид. Потомство этой пары вновь скрещивают с гибридом и т.д.

Эти состояния будем обозначать D (доминант), P (рецессивный индивид) и G (гибрид).

Матрица P вероятностей перехода имеет вид:

$$P = \begin{matrix} & \begin{matrix} D & G & P \end{matrix} \\ \begin{matrix} D \\ P \\ G \end{matrix} & \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 1/2 & 1/2 \end{bmatrix} \end{matrix}.$$

Дальнейшее математическое изучение показывает, что к какому бы типу не принадлежала первая особь, вероятность появления доминантной особи после большого числа скрещиваний равна приблизительно $\frac{1}{4}$, вероятность появления гибридной особи равна $\frac{1}{2}$ и вероятность появления рецессивной особи равна $\frac{1}{4}$.

Изменим условие предыдущего примера следующим образом: вновь получившееся потомство мы будем скрещивать с доминантным животным. В этом случае матрица вероятностей перехода имеет вид:

$$D = \begin{matrix} & \text{Д} & \text{Г} & \text{Р} \\ \begin{matrix} \ddot{A} \\ \tilde{A} \\ \dot{D} \end{matrix} & \begin{bmatrix} 1 & 0 & 0 \\ 1/2 & 1/2 & 0 \\ 0 & 1 & 0 \end{bmatrix} \end{matrix}.$$

В этом случае математическое дальнейшее изучение показывает, что по истечении достаточно длительного времени мы получим доминантное потомство. Это подтверждается экспериментальными данными.

Литература

1. Высшая математика: Практикум для студентов биологического факультета Н.В. Кепчик - Минск, БГУ, 2010-99с.
2. Мацкевич И.П., Свирид Г.П., Булдык Г.М. Сборник задач и упражнений по высшей математике. Теория вероятностей и математическая статистика. Мн., Выш.школа, 1996.
3. Дж.Кемени, Дж.Снелл, Дж.Томпсон. Введение в конечную математику. М., 1963.
4. Робертс Ф.С. Дискретные математические модели с приложениями к социальным, биологическим и экономическим задачам. М., Наука, 1986.