UNIVERZITA KARLOVA

MATEMATICKO-FYZIKÁLNÍ FAKULTA

Obecná
matematika

Genetika

Vypracoval: **Marek Šešulka**

Ročník: **první**

Předmět: **Matematické aplikace**

Rok: **2016/2017, Letní semestr**

Poslední úprava: **15.9.2017**

Obsah

Úvod

V oblasti genetiky se nachází spousta zajímavých matematických procesů, jenž se dají najít u již elementárních záležitostech například jakou barvu očí bude mít potomek a s jakou pravděpodobností. Matematickými procesy v genetice se zabývá speciální algebra, která se nazývá Genetická algebra (Genetic algebra). Ta se poté dělí na další části, podle toho, kterou částí se zabývá.

Na spoustu genetických situaci v oblasti genetiky se dá použit tzv. Markovův řetězec (Markov chain)

Markovův řetězec

Jedná se o způsob, jak popsat diskrétní náhodný proces, po který platí, že následující akce záleží pouze na aktuální situaci, a nikoliv na předchozích (tzn. daný systém splňuje Markovovskou vlastnost), proto můžeme proces znázornit stavovým diagramem.

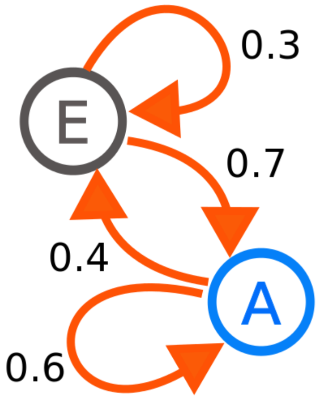
Tento stavový diagram říká:

Je-li proces v A pak s pravděpodobností 0,6 zůstane v bodě A a s pravděpodobností 0,4 přejde do bodů E

Je-li proces v bodě E tak s pravděpodobností 0,3 zůstane v bodě E a s pravděpodobností 0,7 přejde do bodů A.

(Dvou stavový Markovův řetězec)

Díky tomu, že nám stačí pouze aktuální stav řadí se tento způsob k bezpaměťovým.



Obrázek 1- Příklad Stavového diagramu
Obrázek je převzat z [1]

Formální definice

<u>Definice (Markovův řetězec)</u>

Buďte $X_0, X_1, X_2, \ldots, X_n, \ldots$ celočíselné náhodné veličiny. Systém $\{X_n\}_{n=0}^{\infty}$ je **Markovův řetězec**, jestliže platí:

$$P\left[X_{n+1} = j / X_n = j_n \wedge X_{n-1} = j_{n-1} \wedge \ldots \wedge X_1 = j_1 \wedge X_0 = j_0\right] = P\left[X_{n+1} = j / X_n = j_n\right]$$

pro každé $n \in \mathbf{N}_0;\ \ j_0, j_1, \ldots, j_n, j \in \mathbf{N}_0;$

<u>Poznámka</u>

Celočíselné náhodné veličiny, jsou náhodné veličiny, jenž nabývají pouze hodnot z množiny celých čísel.

<u>Definice (Matice pravděpodobností přechodu, vektor absolutních pravděpodobností)</u>

V Markovově řetězci se pravděpodobnosti $p_i(n)$ nazývají **absolutní pravděpodobnosti**, pravděpodobnosti $p_{ij}(n;m)$ se nazývají **pravděpodobnosti přechodu**.

Dále definujeme **matici pravděpodobností přechodu** takto: $\mathbf{P}(n;m)=\left(p_{ij}(n;m)\right)_{i=0,1,\dots}^{j=0,1,\dots}$,

vektor absolutních pravděpodobností takto: $\mathbf{p}^T(n)=\left(p_i(n)\right)_{i=0,1,\dots}$.

<u>Poznámka:</u>

Matice pravděpodobností přechodu může mít též nekonečně mnoho řádků i sloupců. Vektor absolutních pravděpodobností může mít nekonečně mnoho složek.

Rozlišujeme Markovovy řetězce s konečně mnoha stavy a nekonečně mnoha stavy.

Řádkový součet prvků v matici pravděpodobností přechodu musí dát 1, tj. $\sum_j p_{ij} = 1;$

<u>Definice (Homogenní Markovův řetězec)</u>

Platí-li $p_{ij}(n;m)=p_{ij}(n+l;m+l)\ \ \forall n,m,l\in\mathbf{N}\ \ \forall i,j\in\mathbf{N}_0$, nazývá se **Markovův řetězec homogenní**. (Znamená to, že pravděpodobnosti přechodu nezávisí na čase.)

Všechny tyto definice jsou převzaté z [4].

<u>Poznámka</u>

Řekneme, že Markovův řetězec je nehomogenní, pokud není homogenní. (Jednotlivé pravděpodobnosti závisí na čase.)

V našem případe si vystačíme pouze s homogenním Markovovým řetězcem.

Malý příklad

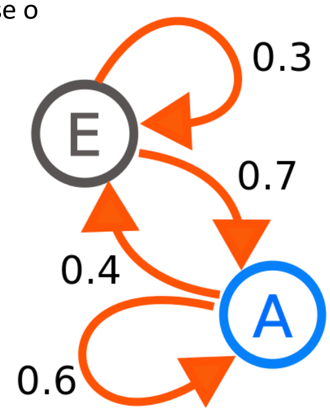
Vraťme se k předchozímu stavovému diagramu Markovova řetězce. Jedná se o konečný systém. Tudíž matice přechodu bude mít konečně mnoho řádků i sloupců a vektory absolutních pravděpodobností budu mít konečný počet složek. Konkrétně:

Stav A označme stavem 1

Stav E označme stavem 2

Poté vektory absolutních pravděpodobností budou mít dvě složky a to:

$$\mathbf{p}^T(1)=((0{,}6),(0{,}4))$$
$$\mathbf{p}^T(2)=((0{,}7),(0{,}3))$$



Obrázek 2- Příklad Stavového diagramu
Obrázek je převzat z [1]

Z těchto dvou vektorů snadno složíme matici přechodů **P,** jež bude mít rozměry 2x2 a její řádky budou tvořit transponované vektory absolutních pravděpodobností. Tedy:

P= $\begin{pmatrix} 0{,}6 & 0{,}4 \\ 0{,}7 & 0{,}3 \end{pmatrix}$

Povšimneme si, že součet složek jednotlivých absolutních vektorů se rovná 1 a tudíž také součet složek v jednotlivých řádcích matice přechodu se rovná 1. Takové to matici se říká stochastická matice. Pokud navíc platí, že i součet složek ve všech sloupcích se rovná 1 pak ji nazveme dvojně stochastická matice.

Homogenní Markovův řetězec

Pokud je Markovův řetězec homogenní pak nechť $\mathbf{x_0}$ je počáteční stav daného jevu a nechť P je matice pravděpodobnostního přechodu. Potom platí:

V čase 1 je systém ve stavu $\mathbf{x_1}$ a ten je definován vztahem: $\mathbf{x_1 = P x_0}$

Expresně můžeme vyjádřit stav $\mathbf{x_n}$ v čase n a to vztahem: $\mathbf{x_n = P^n x_0}$

Příklad: Genetika očí

U oči rozlišujeme dva základní faktory a tj. faktor tmavý (hnědá až černá) a faktor světlý (modrá až zelená). Faktor tmavý je dominantní a faktor světlý je recesivní

Dominantní gen je gen, který se po spárováním s recesivním vždy projeví.

Recesivní gen je gen, který se projeví pouze po spárováním s recesivním genem naopak při spárováním s dominantním genem je potlačen.

Pokud jsou v páru dva recesivní nebo naopak dva dominantní geny pak mají oba 50% šanci že přebiji onen druhý gen.

Recesivní x Dominantní => Dominantní

Recesivní x Recesivní => Recesivní

Dominantní x Dominantní => Dominantní

Každý člověk má dva faktory pro barvu svých očí (po jednom od každého z rodičů). Ty podle výše uvedených vztahů určí barvu jeho očí.

Následující tabulka a diagram zobrazují pravděpodobnostní přechody do jednotlivých uzlů. V tabulce vidíme v prvním sloupečku rodiče A ve druhém rodiče B, poté ve třetím je pravděpodobnost, že budou mít potomka s tmavým faktorem a na konec ve čtvrtém, že potomek bude se světlým faktorem. Pravděpodobnosti jsou vyjádřeny zlomkem.

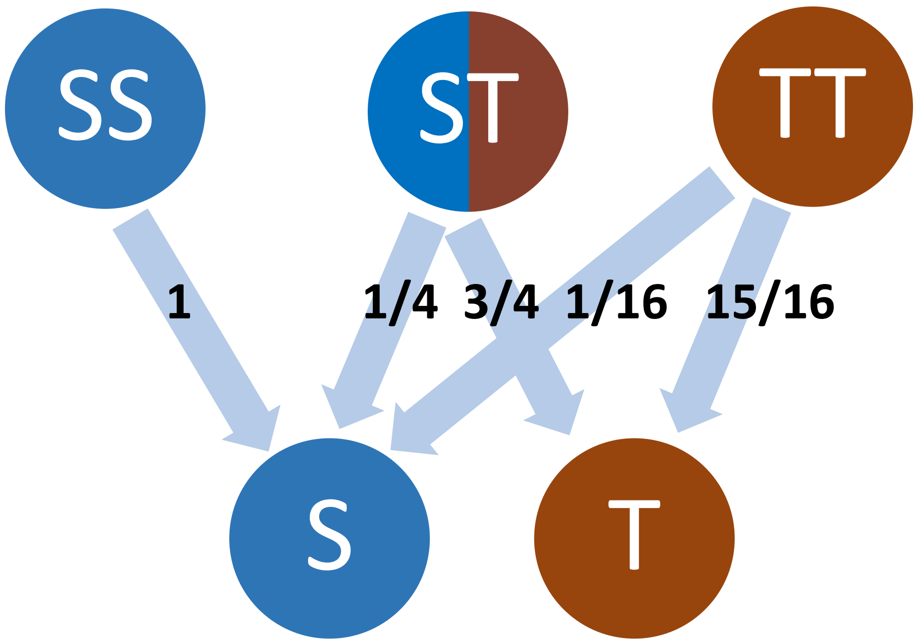
Tabulka 1-Přechod mezi faktory

Rodič A	Rodič B	Potomek s Tmavým f.	Potomek se Světlým f.
Tmavý f.	Tmavý f.	15/16	1/16
Tmavý f.	Světlý f.	3/4	1/4
Světlý f.	Světlý f.	0	1

Poznámka: SS⇔Oba rodiče se Světlým faktorem.

TT⇔Oba rodiče s Tmavým faktorem.

TS⇔Jeden rodič s tmavým faktorem a druhý se světlým.



Obrázek 3-Přechody mezi faktory

Nyní jednoduše sestavíme tabulku, ze které vybudujeme Markovův řetězec. Projdeme všechny možné kombinace, jaké mohou nastat pro oba rodiče a u každé evidujeme pravděpodobnost přechodu do stavu TT, ST a SS. Pravděpodobnosti jsou opět uváděny ve zlomcích

Tabulka 2-Tabulka přechodů TT, ST, SS

Rodič A	Rodič B	TT	ST	SS
TT	TT	1	0	0
TT	ST	0,5	0,5	0
TT	SS	0	1	0
SS	TT	0	1	0
SS	ST	0	0,5	0,5
SS	SS	0	0	1
ST	TT	0,5	0,5	0
ST	ST	0,25	0,5	0,25
ST	SS	0	0,5	0,5

Z údajů v tabulce zrealizujeme Markovův řetězec s třemi stavy TT, SS, ST.

Nejprve sestavíme matici přechodu **P**= $\begin{pmatrix} 0,5 & 0,5 & 0 \\ 0,25 & 0,5 & 0,25 \\ 0 & 0,5 & 0,5 \end{pmatrix}$

Chceme určit matic $\mathbf{P}^n$ pro všechna přirozená n.

Nejprve spočteme vlastní čísla a jim příslušné vlastní vektory. Vlastní čísla nám umožni sledovat, jak se bude libovolný počáteční stav $\mathbf{x}_0$ chovat po n krocích také můžeme zjistit, jak bude vypadat stav, když půjdeme s n do nekonečna (tedy limitu). Vlastní čísla $\lambda_1 = 0, \lambda_2 = 1$ jsou ihned zřejmá, protože matice je singulární tudíž je 0 vlastní číslo a součet všech složek v řádku je 1 pak si stačí uvědomit, že matice transponovaná má stejná vlastní čísla jako ta původní a také co to znamená, pokud má matice ve všech řádcích stejný součet. Spočteme příslušné vlastní vektory v_1 a v_2 , $v_1 = (1 \quad -1 \quad 1)^T$,

$v_2 = (1 \quad 1 \quad 1)^T$ ze stopy matice zjistíme poslední vlastí číslo $\lambda_3 = 0{,}5$ a dopočítáme $v_3 = (-1 \quad 0 \quad 1)^T$.

To nám dává diagonální matici D= $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0,5 \end{pmatrix}$

Nyní potřebujeme najít matici **A** takovou, že $\mathbf{P}=\mathbf{A}\,\mathbf{D}\,\mathbf{A}^{-1}$.

Sloupce matice **A** tvoří vlastní vektory matice **P**. Tedy A= $\begin{pmatrix} 1 & 1 & -1 \\ -1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$

Dopočteme matici $\mathbf{A}^{-1}$= 1/4 * $\begin{pmatrix} 1 & -2 & 1 \\ 1 & 2 & 1 \\ -2 & 0 & 2 \end{pmatrix}$

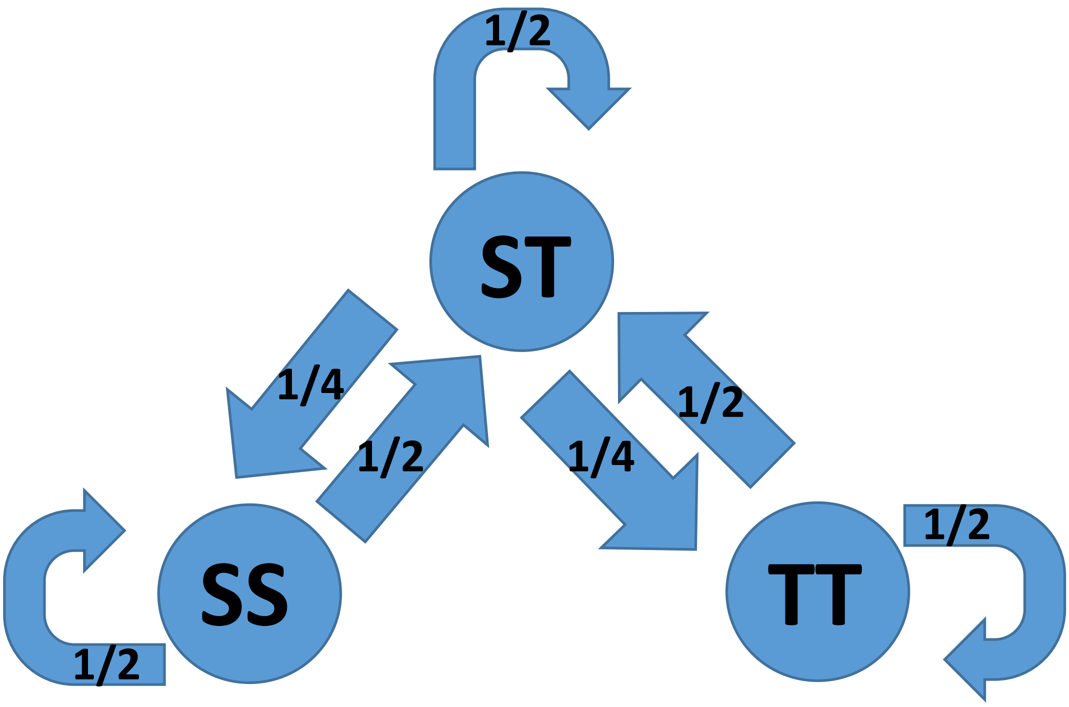
Tudíž máme $\mathbf{P}=\mathbf{A}\,\mathbf{D}\,\mathbf{A}^{-1}$ = 1/4 * $\begin{pmatrix} 1 & 1 & -1 \\ -1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0,5 \end{pmatrix} \begin{pmatrix} 1 & -2 & 1 \\ 1 & 2 & 1 \\ -2 & 0 & 2 \end{pmatrix}$

Poté jest matice $\mathbf{P}^n=\mathbf{A}\,\mathbf{D}^n\,\mathbf{A}^{-1}$ =1/4 * $\begin{pmatrix} 1 & 1 & -1 \\ -1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1^n & 0 \\ 0 & 0 & 0,5^n \end{pmatrix} \begin{pmatrix} 1 & -2 & 1 \\ 1 & 2 & 1 \\ -2 & 0 & 2 \end{pmatrix}$

$$\mathbf{P}^n = 1/4 * \begin{pmatrix} 1+2(0,5)^n & 2 & 1-2(0,5)^n \\ 1 & 2 & 1 \\ 1-2(0,5)^n & 2 & 1+2(0,5)^n \end{pmatrix}$$

$$\mathrm{Lim}_{n \to \text{infinity}} \mathbf{P}^n = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 2 & 1 \\ 1 & 2 & 1 \end{pmatrix}$$

Pomocí údajů z tabulky č. 2 můžeme postavit stavový diagram Markovova řetězce se třemi stavy, tj. SS, ST a TT stačí jen spočítat pro každý stav s jakou pravděpodobností zůstane ve svém stavu a s jakou pravděpodobnosti přejde do ostatních dvou stavů:



Obrázek 3- Stavový diagram markovova řetězce

Citace

Všechny výpočty v příkladu Genetika očí jsem si sám vypočítal a z nich sestavil příslušné stavové diagramy, při tvorbě celého příkladu jsem využíval odborných znalosti a vlastnosti ze zdrojů [2], [3], [5], [6].

[1] Markovov chain. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2017-07-10]. Dostupné z: https://en.wikipedia.org/wiki/Markov_chain

[2] X-linked recessive inheritance. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2017-07-10]. Dostupné z: https://en.wikipedia.org/wiki/X-linked_recessive_inheritance

[3] Genetic algebra. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2017-07-10]. Dostupné z: https://en.wikipedia.org/wiki/Genetic_algebra

[4] Markovův řetězec. *Matematika- Richard Tuček* [online]. 2003 [cit. 2017-07-12]. Dostupné z: www.tucekweb.info/Matem/Vys_mat/St_proc/MPSP03.doc

[5] Genetika 3. *Gymnázium Milady Horákové* [online]. Praha, 2009 [cit. 2017-09-15]. Dostupné z: http://www.gymh.cz/vyuka/biologie/prehledy/9sem_genetika3.pdf

[6] *Mendelistická genetika* [online]. , 1-42 [cit. 2017-09-15]. Dostupné z: http://www.gjs.cz/vedy-o-zemi/Ruda/Sbi/3-SB4-mendelisticka-genetika.pdf