

## Seriál – Matematická biologie I: Populační modely, dynamika a chaos

Druhá polovina dvacátého století přinesla revoluci v našem poznání živých systémů, počínaje rozluštěním struktury DNA a lidským genomem konče. Vznikaly nové mezioborové vědecké disciplíny a v neposlední řadě došlo k postupnému pronikání matematiky do biologie a medicíny. A proto ani korespondenční seminář nemůže stát stranou a připravil pro letošní seriál zajímavou novinku. Budeme si povídat o tom, že biologie a matematika mohou najít společnou řeč, o oboru, který budeme nazývat matematická biologie.

Biologie je věda zabývající se studiem velice různorodých živých systémů. Abychom pochopili život, musíme znát strukturu a funkci jednotlivých molekul, které se vyskytují v tělech rostlin a živočichů. Na druhou stranu, biologové také zkoumají objekty mnohem větší – například jednotlivé buňky, orgány, organismy či v neposlední řadě celá společenstva organismů. Stručně řečeno, biolog může (na jedné straně) přejíždět stovky kilometrů při studiu populace medvědů na Aljašce, nebo se (na druhou stranu) pohybovat ve světě molekul, kde se typické vzdálenosti měří v nanometrech. K pochopení takto rozdílných světů je samozřejmě potřeba rozdílná matematika.

V dnešním díle seriálu se zaměříme na objekty velké – na chování celých populací organismů (např. bakterií, mravenců, medvědů, lidí) či skupin buněk (např. růst rakovinného nádoru). Ve druhém díle seriálu se zaměříme na objekty malé – může se těšit na povídání o DNA, bílkovinách a jejich úloze v genetice, vývoji embryí, evoluci a medicíně. Téma závěrečného třetího dílu seriálu budiž prozatím překvapením.

Abychom mohli použít matematiku ke studiu živých objektů, musíme si nejprve ujasnit, s jakými problémy nám matematika může pomoci. Chceme-li studovat aljašské medvědy grizzly, otázek nás napadne spousta. Čím se medvěd živí? Jak rychle běhá? Proč v zimě spí? Kolik je medvědů na Aljašce? Atd. Najít odpověď na podobné otázky vyžaduje spoustu pozorování, experimentů a studia biologické literatury. To samozřejmě platí i pro další biologické systémy – chceme-li přispět matematikou k pochopení biologie, musíme se nejprve naučit biologii! A i když se nakonec biologii naučíme, vystačíme mnohdy jen se slovní popisnou odpovědí. Například pozorováním zodpovíme naši poslední otázku větou, že na Aljašce žije přibližně 35 tisíc medvědů grizzly. Na druhou stranu, ptáme-li se, kolik bude medvědů na Aljašce za rok, za deset let nebo za sto let, pak s pouhým pozorováním nevystačíme a můžeme začít zapojovat své matematické myšlení. Podobné otázky o budoucnosti populace nemusíme klást jen pro populaci medvědů, ale pro jakoukoliv skupinu jedinců, zvířat či rostlin. Studujeme-li například rostoucí zhoubný nádor, zajímá nás kolik buněk bude obsahovat v budoucnosti. Bojujeme-li s epidemií, zajímá nás vývoj počtu nakažených lidí. Pokusme se tedy zformulovat podobné otázky obecně matematicky.

Uvažujme, že charakteristická jednotka času pro naši populaci je jeden rok, a označme si počet jedinců dnes jako  $N_0$ , počet jedinců za rok jako  $N_1$ , počet jedinců za dva roky jako  $N_2$ , obecně počet jedinců za  $t$  let jako  $N_t$ . Předpokládejme pro jednoduchost, že na základě pozorování víme, že velikost populace v roce  $t + 1$  můžeme spočítat z velikosti populace v roce  $t$  podle vzorečku

$$N_{t+1} = f(N_t), \quad (1)$$

kde  $f$  je nějaká známá funkce. Rovnice typu (1) matematici nazývají *diferenční rovnice*. Chování populace je jednoznačně dáno rovnicí (1) a hodnotou  $N_0$ , kterou budeme nazývat *počáteční podmínka*.

Jako jednoduchý příklad rovnice (1) předpokládejme, že velikost populace v roce  $t + 1$  je dvojnásobkem velikosti populace v roce  $t$ , pak (1) zní  $N_{t+1} = 2N_t$ . Budeme-li tedy v čase  $t = 0$

mít sto jedinců (tzn.  $N_0 = 100$ ), bude jich za rok  $N_1 = 200$ , za dva roky  $N_2 = 400$ , za tři roky  $N_3 = 800$ , za čtyři roky  $N_4 = 1600$  atd. Obecně za  $t$  let bude populace obsahovat  $N_t = 2^t \cdot 100$  jedinců, jak si čtenář sám dokáže matematickou indukci. Jak vidíme z obecného vyjádření, roste počet jedinců s časem nade všechny meze. To ovšem v reálném problému nastat nemůže, neboť organismy mají omezený zdroj potravy, omezený prostor, ve kterém žijí, apod. Realističtější je předpokládat, že s rostoucím počtem jedinců klesá jejich rychlost množení. Příkladem může být diferenční rovnice

$$N_{t+1} = \left(2 - \frac{N_t}{500}\right) N_t. \quad (2)$$

Rovnice (2) je příkladem takzvané logistické rovnice. Pro malý počet jedinců  $N_t$  můžeme na chvíli zanedbat malé číslo  $\frac{N_t}{500}$  a dostaneme, že chování populace je podobné jako chování dané dříve zkoumanou rovnicí  $N_{t+1} = 2N_t$ . Na druhou stranu, zvyšuje-li se počet  $N_t$ , snižuje se faktor  $\left(2 - \frac{N_t}{500}\right)$  a tím dochází ke snižování růstu populace, což řeší problém s nerealistickým přemnožením populace. V neposlední řadě modely ve tvaru rovnice (2) byly i v nedávné době s úspěchem použity na modelování rozličných biologických systémů, například pro modelování růstu rakovinného nádoru (kde  $N_t$  označuje počet rakovinných buněk), či vývoje počtu obyvatel různých států. Podotkneme, že rovnice (2) nám dává obecně posloupnost reálných čísel  $N_t$  a chceme-li finální výsledky interpretovat jako počty jedinců, musíme je zaokrouhlit. V každém případě čísla  $N_t$ ,  $t = 0, 1, 2, \dots$  musí být reálná nezáporná čísla, to znamená, že  $N_t \geq 0$  a  $N_{t+1} = \left(2 - \frac{N_t}{500}\right)N_t \geq 0$ . Řešením těchto dvou nerovnic dostaneme nutnou podmínku pro  $N_t$  ve tvaru  $N_t \in \langle 0, 1000 \rangle$ .

Zajímá nás tedy chování *posloupnosti reálných čísel* daných rovnicí (2) při počáteční podmínce  $N_0 \in \langle 0, 1000 \rangle$ . Jako první krok analýzy diferenčních rovnic ve tvaru (1) je užitečné spočítat si takzvané *pevné body*, které jsou definovány jako řešení rovnice  $N = f(N)$  a mají následující interpretaci. Je-li  $N_0$  rovno pevnému bodu, potom  $N_0 = N_1 = N_2 = N_3 = \dots$ , tzn. že biologický systém má v každém čase  $t$  stejný počet  $N_t = N_0$  jedinců. V případě modelu (2) jsou tedy pevné body dány řešením rovnice  $N = \left(2 - \frac{N}{500}\right)N$ . To je kvadratická rovnice, která má dvě řešení  $N = 0$  a  $N = 500$ . To znamená, že pokud v čase  $t = 0$  žilo  $N_0 = 500$  jedinců, nebude se jejich počet měnit a dostaneme  $N_t = 500$  pro libovolný čas  $t$ . Podobně, pokud na začátku nebyl žádný jedinec v systému ( $N_0 = 0$ ), budeme mít  $N_t = 0$  pro libovolné  $t$ . Tím jsme odhalili všechno zajímavé o modelu (2) pro speciální počáteční podmínky  $N_0 = 0$  a  $N_0 = 500$ . Co však můžeme říci pro obecnou počáteční podmínku  $N_0$  z intervalu  $(0, 1000)$ ?

Zvolme například  $N_0 = 100$  a počítejme. S využitím rovnice (2) dostaneme (při zaokrouhlení na jedno desetinné místo)  $N_1 = 180$ ,  $N_2 = 295,2$ ,  $N_3 = 416,1$ ,  $N_4 = 485,9$ ,  $N_5 = 499,6$ ,  $N_6 = 500$ ,  $N_7 = 500$ ,  $N_8 = 500$  atd. Vidíme, že se hodnoty  $N_t$  s rostoucím časem přibližují k hodnotě 500, což je hodnota jednoho z pevných bodů. Ke stejnému výsledku vede i jakákoliv jiná počáteční podmínka z intervalu  $(0, 1000)$ , jak se může čtenář sám přesvědčit. Tímto pozorováním jsme jednak popsali dynamické chování modelu (2) pro všechny zajímavé počáteční podmínky, ale také jsme zjistili, že pevný bod 500 je mnohem významnější než pevný bod 0, neboť členy všech posloupností začínajících číslem  $N_0 \in (0, 1000)$  jsou pro velká  $t$  prakticky rovné číslu 500. Naproti tomu jenom posloupnosti začínající čísly 0 a 1000 vedou k druhému pevnému bodu 0. Poznamenejme, že posloupnosti začínající číslem mimo interval  $\langle 0, 1000 \rangle$  vedou k záporným číslům, nemají žádnou biologickou interpretaci a nemá smysl se jimi zabývat.

Na základě předcházejícího pozorování pro rovnici (2) zkusme vybudovat obecnou teorii pro rovnici (1). Rovnice (1) a počáteční podmínka  $N_0$  nám definují nekonečnou posloupnost reálných čísel  $N_0, N_1, N_2, N_3, \dots, N_t, \dots$ . Důležitým matematickým objektem, který se hodí ke studiu posloupností, je takzvaná limita posloupnosti. Intuitivně je *limita posloupnosti* číslem, ke kterému se blíží členy posloupnosti  $N_t$  pro velké časy, toto číslo se standardně označuje  $\lim N_t$ .

Například jsme již dříve objevili, že pro libovolnou počáteční podmínku  $N_0 \in (0, 1000)$  dostaneme pro posloupnost danou modelem (2) hodnotu  $\lim N_t = 500$ . Obecnou definici limity posloupnosti nalezneme čtenář ve středoškolských učebnicích matematiky. Poznamenejme, že pro některé posloupnosti limita nemusí existovat.

Nyní již můžeme definovat, že pevný bod  $\bar{N}$  rovnice (1) je *globálně stabilní*, pokud pro libovolnou počáteční podmínku  $N_0$  existuje limita posloupnosti  $N_t$  a platí  $\lim N_t = \bar{N}$ . Naše definice v podstatě říká, že počet jedinců v populaci se s rostoucím časem blíží k pevnému bodu  $\bar{N}$ . Takový globálně stabilní pevný bod je bezesporu velmi významný, neboť v tomto jednom čísle je ukryto chování populace pro velké časy pro libovolnou počáteční podmínku. Na druhou stranu není příliš časté, aby takový globálně stabilní pevný bod existoval. Například v případě rovnice (2) není pevný bod 500 globálně stabilní, protože například posloupnost začínající 0 se nikdy k číslu 500 nepřiblíží. V tomto případě je proto lepší definovat jen takzvanou lokální stabilitu pevného bodu, která požaduje, že  $\lim N_t = \bar{N}$  platí jen pro počáteční podmínky  $N_0$  z nějakého otevřeného intervalu obsahujícího zkoumaný pevný bod  $\bar{N}$ . Ve smyslu této definice má model (2) lokálně stabilní pevný bod rovný číslu  $\bar{N} = 500$ . Vskutku, pro jakoukoliv počáteční podmínku  $N_0 \in (0, 1000)$  platí  $\lim N_t = 500$ . Na druhou stranu pevný bod 0 rovnice (2) není stabilní, tudíž vidíme, že definice stability nám pomohla rozlišit významné a méně významné pevné body. Pochopení definice stability si můžeš vyzkoušet v příkladě 2, k jehož řešení si vystačíš s jednoduchou středoškolskou matematikou. V obecném případě se k vyšetřování stability pevných bodů hodí elementární znalost derivací. Pokud jsi již o derivacích slyšel, pak by neměl být pro Tebe problém dokázat stabilitu pevného bodu 500 v našem předcházejícím případě. Abychom nediskriminovali mladší řešitele, nebudeme v tomto seriálu derivace potřebovat.

Dynamika (rozuměj chování řešení) rovnic podobných rovnici (2) může být mnohem komplikovanější než dynamika, kterou jsme již viděli. Uvažujme například diferenční rovnici

$$N_{t+1} = R \left( 1 - \frac{N_t}{1000} \right) N_t, \quad (3)$$

kde  $R$  je dané známé číslo. Zvolíme-li  $R = 2$  v modelu (3), dostaneme model (2), jehož chování už známe. Zvolme proto  $R = 3,2$  v modelu (3) a sledujme, jak se změní chování této diferenční rovnice. Zvolíme-li počáteční podmínku  $N_0 = 500$ , dostaneme  $N_1 = 800$ ,  $N_2 = 512$ ,  $N_3 = 799,5$ ,  $N_4 = 512,9$ ,  $N_5 = 799,5$ ,  $N_6 = 513,0$ ,  $N_7 = 799,5$ ,  $N_8 = 513,0$  atd. Vidíme, že řešení po čase osciluje mezi hodnotami 799,5 a 513,0. K takovému výsledku dojdeme pro libovolnou počáteční podmínku z intervalu  $(0, 1000)$ . Tím jsme zjistili, že rovnice (3) pro  $R = 3,2$  nemá stabilní pevný bod (ona má dva nestabilní pevné body, které čtenář snadno najde), ale má oscilující řešení s periodou délky 2. Taková řešení se dají pro libovolnou rovnici (1) odhalit jako řešení rovnice  $N = f(f(N))$ .

Dosadíme-li  $R = 3,5$  do modelu (3), objevíme periodické řešení o délce 4, podobně, dosadíme-li  $R = 3,55$  do modelu (3), objevíme periodické řešení o délce 8, jak si čtenář může sám vyzkoušet. Dynamika daná modelem (3) začne být ještě překvapivější, když dosadíme čísla  $R > 3,570$ . Jak si čtenář sám ověří, můžeme dostat jednak periodická řešení, ale také řešení zcela neperiodická. Například pro  $R = 4$  a počáteční podmínku  $N_0 = 50$  model (3) dává  $N_1 = 190$ ,  $N_2 = 615,6$ ,  $N_3 = 946,5$ ,  $N_4 = 202,4$ ,  $N_5 = 645,7$ ,  $N_6 = 915,1$ ,  $N_7 = 310,8$  atd. Tím jsme objevili deterministický chaos – neperiodické omezené řešení dané jednoduchým vzorcem, které velice citlivě závisí na počáteční podmínce  $N_0$ . Čtenář si může ještě chvíli experimentovat s rovnicí (3) a my se mezitím vrátíme k biologii.

Model (1) nám říká, že počet jedinců v čase  $t+1$  je funkcí počtu jedinců v čase  $t$ . V některých případech může být takový model daleko od reality. Uvažujme-li například populaci medvěďů,

pak model (1) má v sobě zakotven předpoklad, že populace 100 mládat se bude vyvíjet úplně stejně jako populace 100 starých medvěďů. To je samozřejmě velice zjednodušený předpoklad. Abychom dostali věrohodnější popis skutečnosti, je dobré studovat nejen počet jedinců v populaci, ale také věkové rozložení populace.

Abychom si zjednodušili použitou matematiku, bude nás zajímat jen počet samic v populaci. Navíc, abychom si ještě více zjednodušili použitou matematiku, budeme předpokládat, že se samice dožívají maximálně dvou let (čtenář si může jako cvičení rozmyslet, jak se naše úvahy změní, pokud se samice dožívají třeba třiceti let). Budeme tedy rozlišovat samice s věkem do jednoho roku (nula-leté), jejichž počet v roce  $t$  budeme označovat  $N_t^0$ , a samice s věkem od jednoho do dvou let (jedno-leté), jejichž počet v roce  $t$  budeme označovat  $N_t^1$ . Při dovršení dvou let samice umírá. Samice mají potomky v daném období roku (například na jaře) a víme, že nula-leté samice se v průměru narodí 2 mladé samičky a jedno-leté samice se v průměru narodí 4 mladé samičky. Pouze 75% nula-letých samic přežije první rok života a všechny samice umírají dovršením druhého roku života. Potom můžeme spočítat počet nula-letých a jedno-letých samic v roce  $t + 1$  z počtu nula-letých a jedno-letých samic v roce  $t$  podle vzorečku

$$\begin{aligned} N_{t+1}^0 &= 2N_t^0 + 4N_t^1, \\ N_{t+1}^1 &= 0,75N_t^0. \end{aligned} \quad (4)$$

Model (4) je příkladem takzvaného *Leslieho modelu*. V každém případě je vývoj populace jednoznačně dán rovnicí (4) a počáteční podmínkou – uspořádanou dvojicí  $(N_0^0, N_0^1)$ . Rovnice (4) nám potom dává posloupnost uspořádaných dvojic  $(N_0^0, N_0^1)$ ,  $(N_1^0, N_1^1)$ ,  $(N_2^0, N_2^1)$ ,  $(N_3^0, N_3^1)$ ,  $\dots$ ,  $(N_t^0, N_t^1)$ ,  $\dots$

K tomu, abychom pochopili chování modelu (4), je dobré nejprve nalézt jisté speciální počáteční podmínky, které budeme nazývat *vlastními vektory* modelu (4) a definujeme je takto: Nenulová počáteční podmínka  $(N_0^0, N_0^1)$  je *vlastním vektorem*, pokud existuje reálné číslo  $\lambda$  takové, že  $N_1^0 = \lambda N_0^0$  a  $N_1^1 = \lambda N_0^1$ . Číslo  $\lambda$  budeme pak nazývat *vlastním číslem*. Definice vlastního vektoru nám říká, že populace v následujícím roce je  $\lambda$ -násobkem roku předcházejícího. Je-li tedy populace v čase  $t = 0$  rovna vlastnímu vektoru modelu (4), pak počet jedinců v populaci v čase  $t$  můžeme jednoduše spočítat jako  $N_t^0 = \lambda^t N_0^0$  a  $N_t^1 = \lambda^t N_0^1$ , jak si čtenář snadno dokáže matematickou indukcí. Využijeme-li tedy předcházející definice, pak všechny vlastní vektory a vlastní čísla modelu (4) splňují

$$\begin{aligned} \lambda N_0^0 &= 2N_0^0 + 4N_0^1, \\ \lambda N_0^1 &= 0,75N_0^0, \end{aligned}$$

což po malé úpravě dává

$$\begin{aligned} (2 - \lambda)N_0^0 + 4N_0^1 &= 0, \\ 0,75N_0^0 - \lambda N_0^1 &= 0. \end{aligned} \quad (5)$$

Uvažujeme-li nyní číslo  $\lambda$  jako parametr a čísla  $N_0^0$  a  $N_0^1$  jako souřadnice ve dvou-dimenzionálním prostoru, potom dvě rovnice (5) popisují dvě přímky procházející bodem o souřadnicích  $N_0^0 = 0$  a  $N_0^1 = 0$ . Uspořádaná dvojice  $(N_0^0, N_0^1) = (0, 0)$  je samozřejmě vždy řešením rovnice (5). Přečteme-li si ovšem pozorně definici vlastního vektoru, zjistíme, že hledáme *nenulová* řešení soustavy (5). Taková nenulová řešení existují pouze v případě, když přímky dané oběma rovnicemi jsou totožné, jinými slovy, když první rovnice v soustavě (5) je násobkem druhé. Z tohoto pozorování vyplývá, že vlastní čísla  $\lambda$  splňují rovnici  $\lambda(2 - \lambda) + 0,75 \cdot 4 = 0$ . To je kvadratická rovnice, která má dvě řešení  $\lambda = 3$  a  $\lambda = -1$ . Tím jsme objevili, že model (4) má dvě vlastní čísla.

Uvažujme nyní vlastní číslo  $\lambda = 3$  a spočítejme si příslušný vlastní vektor. Dosazením  $\lambda = 3$  do (5) dostaneme jednu rovnici  $-N_0^0 + 4N_0^1 = 0$  pro dvě neznámé  $N_0^0$  a  $N_0^1$ . Taková rovnice má samozřejmě nekonečně mnoho řešení, nám bude stačit jen jedno řešení, například  $N_0^0 = 4$  a  $N_0^1 = 1$ . Tím jsme objevili vlastní vektor  $(N_0^0, N_0^1) = (4, 1)$  příslušející vlastnímu číslu  $\lambda = 3$ . Zkusme nyní použít uspořádanou dvojici  $(N_0^0, N_0^1) = (4, 1)$  jako počáteční podmínku pro model (4). Dostaneme posloupnost uspořádaných dvojic  $(N_1^0, N_1^1) = (12, 3) = (3 \cdot 4, 3)$ ,  $(N_2^0, N_2^1) = (36, 9) = (3^2 \cdot 4, 3^2)$ ,  $(N_3^0, N_3^1) = (108, 27) = (3^3 \cdot 4, 3^3)$  atd. Obecně tedy v roce  $t$  bude mít populace strukturu  $N_t^0 = 3^t \cdot 4$  a  $N_t^1 = 3^t$ . Tím jsme ovšem našli vzoreček pro obecný člen posloupnosti dané modelem (4) pro speciální počáteční podmínku  $(N_0^0, N_0^1) = (4, 1)$ .

Podobně můžeme nalézt vlastní vektor pro vlastní číslo  $\lambda = -1$ . Dosazením  $\lambda = -1$  do (5) dostaneme jednu rovnici  $3N_0^0 + 4N_0^1 = 0$  pro dvě neznámé  $N_0^0$  a  $N_0^1$ , jejímž jedním řešením je například  $N_0^0 = 4$  a  $N_0^1 = -3$ . Tím jsme objevili vlastní vektor  $(N_0^0, N_0^1) = (4, -3)$  příslušející vlastnímu číslu  $\lambda = -1$ . Čtenáře by mělo samozřejmě ihned zarazit, že vlastní vektor obsahuje záporné číslo, proto vlastnímu vektoru  $(4, -3)$  nemůžeme dát biologickou interpretaci. Na druhou stranu bude vlastní vektor  $(4, -3)$  spolu s vlastními vektorem  $(4, 1)$  užitečným pomocným prostředkem k pochopení chování modelu (4). K tomu bude velice užitečné rovněž následující pozorování. Uvažujme libovolné dvě reálné konstanty  $C_1$  a  $C_2$  a uvažujme počáteční podmínku danou součtem  $(N_0^0, N_0^1) = C_1(4, 1) + C_2(4, -3)$ , což znamená  $N_0^0 = C_1 \cdot 4 + C_2 \cdot 4$  a  $N_0^1 = C_1 - C_2 \cdot 3$ . Potom je počet nula-letých a jedno-letých samic čase  $t$  dán vztahem  $(N_t^0, N_t^1) = C_1 3^t(4, 1) + C_2(-1)^t(4, -3)$ , což znamená

$$N_t^0 = C_1 3^t \cdot 4 + C_2(-1)^t \cdot 4 \quad \text{a} \quad N_t^1 = C_1 3^t - C_2(-1)^t \cdot 3. \quad (6)$$

Vzoreček (6) čtenář snadno nahlédne dosazením počáteční podmínky do rovnic (4), použitím vlastností vlastních vektorů  $(4, 1)$  a  $(4, -3)$  a využitím matematické indukce. S využitím vztahu (6) můžeme spočítat počet jedinců daných modelem (4) pro libovolný čas  $t$  a pro libovolnou počáteční podmínku  $(N_0^0, N_0^1)$ .

Typická úloha může tedy znít takto: Uvažujme populaci obsahující 8 mladších samic a 22 starších samic, která se vyvíjí podle modelu (4). Kolik bude mladších a starších samic za sto let? V tomto případě je počáteční podmínka  $(N_0^0, N_0^1) = (8, 22)$ . Nejprve nalezneme takové konstanty  $C_1$  a  $C_2$ , které splňují  $(8, 22) = C_1(4, 1) + C_2(4, -3)$ , neboli budeme řešit soustavu rovnic

$$\begin{aligned} 4C_1 + 4C_2 &= 8, \\ C_1 - 3C_2 &= 22. \end{aligned}$$

Řešením této soustavy rovnic je  $C_1 = 7$  a  $C_2 = -5$ . Využitím vzorce (6) proto dostaneme, že počet mladších samic v roce  $t = 100$  je roven  $N_{100}^0 = 7 \cdot 3^{100} \cdot 4 - 5 \cdot (-1)^{100} \cdot 4 = 28 \cdot 3^{100} - 20$  a počet starších samic v roce  $t = 100$  je roven  $N_{100}^1 = 7 \cdot 3^{100} + 5 \cdot (-1)^{100} \cdot 3 = 7 \cdot 3^{100} + 15$ . Čtenář si může sám dokázat, že poměr počtu mladších a starších samic pro velké  $t$  je přibližně roven 4:1, tzn.  $\lim_{t \rightarrow \infty} \frac{N_t^0}{N_t^1} = 4$ .

V předcházejících odstavcích jsme prezentovali metodu, jejíž pochopení si můžeš otestovat v příkladě 1 a která nám pro libovolnou počáteční podmínku a pro libovolný čas  $t$  dá jednoduchý výpočet počtu mladších samic  $N_t^0$  a starších samic  $N_t^1$ . Jedna z věcí, která by nás měla zarazit, je, že počet jedinců roste s rostoucím časem nade všechny meze, což není příliš realistické. Se zvyšujícím se počtem organismů roste potřeba živin, prostoru atd., což má v konečném důsledku za následek zpomalení růstu populace způsobené větší úmrtností, menší porodností apod. S problémem přemnožení populace jsme se již dříve setkali, když jsme zkoumali rovnici  $N_{t+1} = 2N_t$ . Tehdy jsem problém vyřešili uvažováním komplikovanějšího modelu (2), který obsahoval kvadratickou funkci. Úplně stejná myšlenka se dá uplatnit i zde. Zavedením vhodných komplikovanějších

funkcí do modelu pak dostaneme omezená řešení, která se budou buď blížit k nějakému bodu, či budou oscilovat, nebo odhalíme opět deterministický chaos jako v případě rovnic (3).

Doposud jsme studovali modely ve tvaru (1) a model s věkově strukturovanou populací (4), všechny tyto modely zatím popisovaly jeden živočišný druh. Většinou ovšem spolu v přírodě žije mnoho různých druhů pohromadě, které spolu soupeří o potravu, o prostor, o přežití. Uvažujme například les, ve kterém žijí vlci a zajíci, a zjednodušíme si situaci tím, že vlci žerou jenom zajíce a zajíci mají trávy vždycky dost. Pak si můžeme slovně popsat dynamiku tohoto systému takto: Pokud je v lese hodně zajíců, začnou se množit vlci, neboť mají dost potravy. Tím, že se začali množit vlci, začnou ubývat zajíci. Tím, že začali ubývat zajíci, nemají vlci co žrát a začne jejich počet klesat. Tím, že klesá počet vlků, začne zase přibývat zajíců, neboť je nemá kdo lovit. Tím, že je hodně zajíců, začne se opět po čase zvyšovat počet vlků atd. Shrnuto, počty zajíců a vlků budou patrně oscilovat. Samozřejmě, že bychom teď mohli po studiu relevantní biologické literatury napsat model pro takový ekosystém a jako správní matematikové bychom dokázali odhalit podobné oscilatorní chování. Na to již zde není místo, čtenář si může ovšem sám vyšetřit jednoduchý model chování soupeřících druhů v příkladě 3. Zde je dynamika mnohem jednodušší než v případě vlků a zajíců a k řešení příkladu 3 není potřeba nic víc, než co jsme si v dnešním díle seriálu pověděli. Pokud bys nevěděl/a, jak si s příkladem 3 poradit, zvol si nějakou počáteční podmínku, napiš si několik prvních členů posloupností a ono Tě už něco napadne.

Závěrem nutno podotknout, že jsme se v prvním dílu seriálu dotkli rozdílných partií matematiky a zvědavý čtenář může v tomto trendu pokračovat. Od diferenčních rovnic, pevných bodů a chaosu je jen krůček k rovnicím diferenciálním, od Leslieho modelu k maticím a lineární algebře. Tak věříme, že Tě naše povídání bude motivovat k dalšímu samostudiu. V příštím díle seriálu si představíme matematiku zcela jinou, neboť zavítáme do mikrosvěta mezi bílkoviny a nukleové kyseliny.

## Matematická biologie II: DNA, bílkoviny a embrya

V minulém díle seriálu jsme se zaměřili na studium skupin velkého počtu organismů nebo buněk. Ukázali jsme si příklady a vlastnosti jednoduchých matematických modelů vývoje biologických společenstev. Studovali jsme diferenční rovnice tvaru  $N_{t+1} = f(N_t)$ , které nám dávaly závislost velikosti populace v roce  $t+1$  na velikosti populace v roce  $t$ . Jedním z přirozených pokračování našeho seriálu by proto mohlo být detailnější studium vlastností diferenčních rovnic. My ne zvolíme tento postup, neboť nechceme psát seriál o diferenčních rovnicích, ale chceme ukázat čtenáři různé oblasti biologie, které vyžadují pomocnou ruku matematiky. Dlužno také podotknout, že diferenční rovnice nebývají tím nejčastějším nástrojem ke studiu otázek položených v minulém díle seriálu. Větší popularitu mají takzvané (obyčejné, parciální, či stochastické) diferenciální rovnice nebo různé typy počítačových simulací. To bychom se ale již dostávali příliš daleko od středoškolského učiva. Nám šlo o vysvětlení základních myšlenek matematického studia populací organismů a k tomu byly z pedagogického hlediska diferenční rovnice nejvhodnější. Dnes však ponecháme populace stranou a zabrouzdáme do úplně jiné oblasti biologie. Budeme si povídat o DNA, bílkovinách a jejich úloze v genetice, vývoji embryí, evoluci a medicíně. Začneme trochou biologie.

V roce 1953 rozluštili Watson a Crick strukturu molekuly DNA, nositelky genetické informace, a svým výzkumem (za který obdrželi v roce 1962 Nobelovu cenu) otevřeli dveře k dalšímu převratným objevům. Po padesáti letech vše vyústilo až k přečtení lidského genetického kódu, lidského genomu, který v roce 2001 publikovaly hned dva velké vědecké týmy. Rozluštění lidského genomu (rozuměj stanovení přesné posloupnosti tří miliard základních kamenů nukleové kyseliny DNA obsažené v lidské buňce) nám do budoucna otevírá netušené možnosti v diagnostikování, prevenci a léčbě nemocí. V neposlední řadě i matematika má k výzkumu genomu co říci. Abychom si to osvětlili, pojďme se nyní podívat, jakou informaci nám biologie dává.

Nukleové kyseliny (DNA a RNA) jsou společně s bílkovinami jedny z nejvýznamnějších složek buněk. V molekulách DNA je uložena dědičná informace buňky. Vznikají propojováním velkého počtu základních chemických kamenů, tzv. nukleotidů. Zjednodušeně řečeno, můžeme si deoxyribonukleovou kyselinu (DNA) představit jako dlouhé slovo složené ze čtyř písmen A,C,T,G. Ta písmena odpovídají začínajícím písmenům názvů příslušných nukleosidů (adenosin(A), guanosin(G), cytidin(C) a thymidin(T)), ze kterých se daná DNA molekula skládá. Odhlédneme-li od chemické struktury DNA, můžeme si tedy genetickou informaci obsaženou v DNA zapsat jako dlouhý sled písmen

... GACGGTGCTGTATTAGCAGCAGTCAATATA ... (1)

Pomocí takto zakódované informace je buňka schopna sestavovat proteiny (česky se také nazývají bílkoviny). Zjednodušeně řečeno, nukleové kyseliny obstarávají veškeré řízení biologických systémů a jsou zodpovědné za předávání genetické informace z generace na generaci. Pro život neméně důležité makromolekulární látky jsou bílkoviny sestavené pomocí informace obsažené v DNA.

Zhruba řečeno, zatímco DNA má moc řídit, mají proteiny veškerou výkonnou moc. Proteiny totiž nesou ve své struktuře informace pro velkou řadu biologických procesů, jako je transport látek a jejich chemické přeměny, řízení rychlosti dějů probíhajících v buňkách atd. Základním stavebním kamenem proteinů jsou tzv. aminokyseliny. O aminokyselinách jsi možná již slyšel a nebo uslyšíš v hodinách chemie. Z biologických systémů jich bylo izolováno několik set, některé se vyskytují jen v určitých druzích, jiné jsou rozšířeny obecně. Daleko nejdůležitějších z nich je dvacet, ze kterých jsou zbudovány molekuly všech bílkovin na této planetě. Říkáme jim *kódované aminokyseliny* a nazývají se např. glycin (zkratka Gly), alanin (Ala), valin (Val), arginin (Arg) atd. Abychom se pomalu dostali k matematice, můžeme si velmi zjednodušeně představit protein jako dlouhou posloupnost sestavenou z různě se opakujících dvaceti základních jednotek, například

... -Asp-Gly-Ala-Val-Leu-Ala-Ala-Val- ... (2)

kde Asp, Gly, Ala, Val, Leu atd. jsou symboly pro některou z kódovaných aminokyselin. Vlastnosti a funkci bílkoviny určuje v první řadě pořadí aminokyselin dané v posloupnosti (2). Na základě chemických vlastností těchto aminokyselin a okolního prostředí zaujme protein (2) svou prostorovou strukturu. Zkratka různé

aminokyseliny se mají různě rády, a tak řetězec aminokyselin zaujme takové prostorové uspořádání, že to všem nejvíce vyhovuje. Záleží samozřejmě i na vlastnostech okolního prostředí. Výsledný prostorový tvar molekuly bílkoviny pak určuje její funkci.

Přechod od řetězce typu (2) k prostorové struktuře a posléze k funkci výsledné bílkoviny je proces významný, velmi studovaný a matematické modelování vyžadující. My se jím ale zde nebudeme zabývat, neboť bychom museli jít do větších biochemických detailů a záhy bychom opustili středoškolské učivo. Zaměříme se místo toho na přechod od DNA dané ve tvaru (1) k posloupnosti aminokyselin daných (2). Tento proces je schematicky znázorněn v následujícím obrázku.



Nejprve se informace obsažená v DNA přepíše do RNA. RNA je také nukleová kyselina s významnými biologickými vlastnostmi. Nám bude stačit vědět, že informaci obsaženou v RNA si můžeme opět představit jako posloupnost čtyř různě se opakujících písmen ve tvaru (1), kde písmeno T je nahrazeno písmenem U (thymidin je nahrazen jiným nukleosidem nazývaným uridin). Přepis tedy s informací obsaženou v řetězci (1) neudělá nic převratného. Druhým procesem je překlad, který na základě řetězce RNA vytvoří posloupnost aminokyselin (2).

Jak však může buňka jednoznačně z posloupnosti tvořené čtyřmi základními kameny (1) vytvořit posloupnost opakujících se 20-ti základních kamenů? Za tímto účelem čte překládací mechanismus písmena v RNA molekule po trojicích. Například, když RNA obsahuje posloupnost následujících dvanácti písmen ... ACGAAUGCACCG ... , rozdělí si pomyslně buňka tato slova na trojice ACG, AAU, GCA, CCG. Pak se podívá do tabulky genetického kódu a zjistí, že trojici ACG odpovídá aminokyselina prolin (Pro), trojici AAU odpovídá aminokyselina tyrosin (Tyr), trojici GCA odpovídá aminokyselina serin (Ser) a trojici CCG odpovídá zase aminokyselina prolin (Pro). Výsledný úsek v proteinu proto zní ... -Pro-Tyr-Ser-Pro- ...

Ze čtyř písmen A,C,G,U můžeme vytvořit celkem  $4^3 = 64$  trojic, proto má buňka dostatek možností k zakódování všech dvaceti kódovaných aminokyselin. Mnoho aminokyselin je kódováno více trojicemi a některé ze 64 trojic také znamenají konec čtení RNA. O tom, jaké má překládací mechanismus z RNA do proteinu důsledky, si čtenář může popřemýšlet při řešení čtvrtého příkladu. Podotkneme, že zvidavý čtenář může genetický kód, tj. překládací tabulku, která říká, jaká aminokyselina odpovídá dané trojici písmenek v RNA, nalézt v běžných učebnicích biochemie. Tuto tabulku ani jiné biologické detaily však čtenář k řešení příkladu 4 nepotřebuje, vystačí si jen se středoškolským kombinatorickým uvažováním.

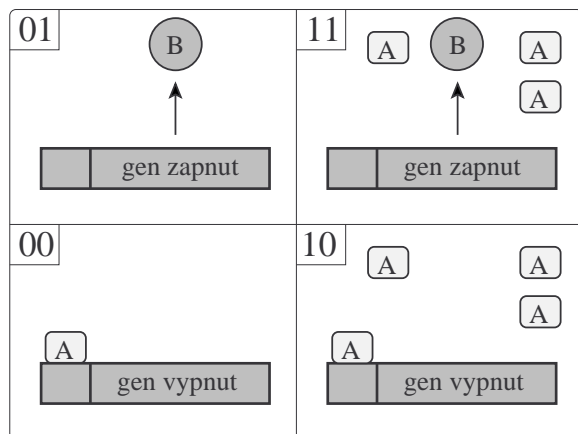
Je zajímavé podotknout, že vynechání jednoho písmenka z posloupnosti RNA může vést k úplně jiné výsledné posloupnosti aminokyselin. Například dříve zmíněná posloupnost ... ACGAAUGCACCG ... po vynechání prvního písmenka A vede k posloupnosti ... CGAAUGCACCG ... , která při čtení po trojicích dává CGA, AUG, CAC, ... , které kódují aminokyseliny po řadě arginin (Arg), methionin (Met) a histidin (His). Tedy výsledný úsek v proteinu po vynechání jednoho písmenka v RNA zní -Arg-Met-His, což je úplně jiný výsledek než ... -Pro-Tyr-Ser-



Pro- ... , který jsme obdrželi prve. Ačkoliv tato vlastnost vypadá jako nevýhoda, proti které musí mít buňka kontrolní mechanismy, využívají zmíněnou vlastnost ke svému užítku některé viry, které pomocí jedné posloupnosti DNA nebo RNA dokáží kódovat více proteinů.

Genom obsahuje dvě velmi důležité informace – geny, které kódují proteiny, jak jsme popsali výše. Druhou informací jsou regulační mechanismy (anglicky gene regulatory networks) které říkají kdy, kde a jak mnoho bílkovin daný gen produkuje. To je samozřejmě velmi důležitá informace, neboť většina buněk lidského těla obsahuje stejnou genetickou informaci. Uvažujme geny, které nesou informaci k tvorbě prstů na lidské ruce. Je naprosto žádoucí, aby se tyto geny překládaly u buněk na správném místě (tj. na ruce), ve správném čase (tj. v daném okamžiku embryonického vývoje) a ve správné intenzitě (tj. aby celý proces nakonec vedl k pěti normálně vypadajícím prstům na lidské ruce). Selhání kontrolních mechanismů vede při embryonickém vývoji k velmi smutným abnormalitám mnohdy neslučitelným se životem, u dospělého člověka k rakovině.

Abychom si vysvětlili, jak to všechno funguje, uvažujme gen produkující protein B. Zjednodušeně řečeno, tento gen může existovat ve dvou stavech. Buď je zapnut a protein B je produkován (stav 1), nebo je vypnut a protein se nepřekládá (stav 0). Stav genu je ovšem kontrolován dalšími proteiny. Na následujícím obrázku předpokládáme, že přepis informace z genu je ovlivňován proteinem označeným A. Zhruba řečeno, pokud je protein A navázaný k DNA, tak se překládání genu vypne, v opačném případě je protein B produkován.



Rozlišme tedy čtyři stavy systému podle toho, zda-li je přítomen protein A, protein B, žádný protein, či oba dva.<sup>1</sup> Tyto stavy jsou dány schematicky na předcháze-

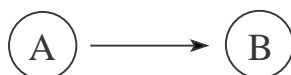
<sup>1</sup>Poznamenejme, že tento model neuvazuje, kolik přesně molekul proteinu je v systému. Zkrátka, pokud je dost molekul daného proteinu v systému, označujeme jeho stav jako 1, a pokud jich je zanedbatelné množství, pak stav systému označíme jako 0. To poznamenáváme proto, aby čtenář nemátl, že i ve stavu 00 máme na obrázku jednu molekulu proteinu A. Slovy „protein

jícím obrázku. Stav 00 znamená, že není přítomen ani jeden protein (zde první číslice popisuje stav proteinu A a druhá číslice popisuje stav proteinu B). Stav 10 znamená, že je přítomen protein A, ale není produkován protein B. Stav 01 znamená, že není přítomen protein A a je produkován protein B a konečně stav 11 značí, že oba dva proteiny jsou v systému.

Jak jsme již poznamenali, je velmi důležité pochopit, jak zapínání a vypínání různých genů závisí na čase. Pro jednoduchost budeme stejně jako v minulém díle seriálu popisovat časovou proměnnou v diskrétních časových krocích. To znamená, že budeme udávat vývoj systému v čase  $t + 1$  na základě vývoje systému v čase  $t$ , kde  $t \in \mathbb{N}$ , jako v následující tabulce:

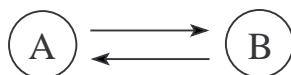
stav systému v čase $t$	00	01	10	11
stav B v čase $t + 1$	1	1	0	0

Předcházející tabulka nám dává možnost spočítat stav genu/proteinu B v čase  $t + 1$  ze stavu v čase  $t$ . Tabulka jenom v matematické terminologii shrnula, co by člověk očekával. Pokud v předcházejícím čase nebyl přítomen protein A, pak se gen B překládá a protein B produkuje. Naopak, pokud v čase  $t$  byl přítomen protein A, pak je v následujícím čase gen B vypnut a protein se nepřekládá. Vidíme tedy, že koncentrace proteinu B závisí jen na koncentraci proteinu A v předcházejícím časovém kroku. Tj., vše můžeme schématicky vyjádřit pomocí následujícího grafu:



V tomto grafu kolečko znamená příslušný gen/RNA/protein a šipka interakci, tj. že stav B je ovlivněn přítomností A v systému. Toto ovlivnění může být pozitivní (přítomnost A zvyšuje produkci proteinu B), nebo negativní (jako v našem případě).

K popsání systému nyní ještě potřebujeme udat, jak se mění stav A v čase. Protein A je samozřejmě produkován příslušným genem a tato produkce musí být rovněž kontrolována ostatními proteiny. Pokud předpokládáme, že stav A je ovlivněn stavem B úplně stejným způsobem jako B je ovlivněno A, můžeme tuto interakci popsat schématicky obrázkem



a tabulka vývoje systému bude dána

stav v čase $t$	00	01	10	11
stav v čase $t + 1$	11	01	10	00

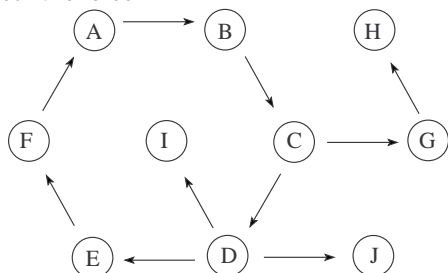
Oprostíme se nyní od biologie a podívejme se na vývoj systému z matematického hlediska. Začneme-li v čase  $t = 0$  se stavem rovným 01, bude stav roven 01 pro

---

je přítomen v systému“ zkrátka rozuměj vždy nezanedbatelné množství.

všechny následující časy. Takové neměnné chování systému jsme v prvním díle seriálu nazývali pevným bodem a tuto počáteční podmínku budeme nazývat *pevným bodem* i zde. Na druhou stranu, začneme-li s počáteční podmínkou rovnou 00, bude stav systému v následujících časech roven 11, 00, 11, 00. Tj. objevili jsme tzv. *cyklus* délky 2.

Zkusme nyní trochu abstrahovat. Komponenty systému, ať už to byly geny, či proteiny, jsme popisovali pomocí koleček a interakce pomocí šipek. Matematikové většinou takové struktury nazývají grafy. Potom kolečkům říkají *vrcholy* a šípkám *hrany*. Předchozí systém měl tedy dva vrcholy a jeho stav byl popsán uspořádanou dvojicí čísel 0 nebo 1. Uvažujme nyní graf o více vrcholech, příkladem může být následující graf o deseti vrcholech.



Uvažujme-li jako dříve, že každý vrchol může být jen v jednom ze stavů 0 nebo 1, bude celý systém popsán uspořádanou posloupností nul a jedniček o délce 10. Takových posloupností je celkem  $2^{10} = 1024$ . To ovšem znamená, že tabulka popisující vývoj systému, která by nám dávala stav v čase  $t + 1$  na základě znalosti stavu systému v čase  $t$ , by musela obsahovat 1024 sloupečků.

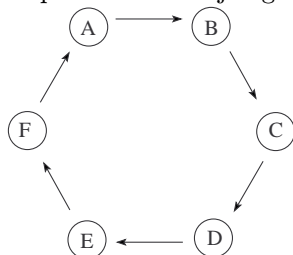
Takovou tabulkou zde nebudeme plýtvat místo, neboť můžeme vývoj systému popsat i jednodušeji a přehledněji. Náš graf je totiž příkladem speciálního typu grafů, kde do každého vrcholu vstupuje jen jedna hrana. Například do vrcholu D vstupuje jen hrana z C, proto chování vrcholu D může být podle jednoho z těchto čtyř pravidel.

název pravidla pro D v čase $t + 1$	stav D v čase $t + 1$ , když je stav C v čase $t$ roven 0	stav D v čase $t + 1$ , když je stav C v čase $t$ roven 1
Identita	0	1
Inverze	1	0
Nula	0	0
Jedna	1	1

Pravidlo Identita neříká nic jiného, než že hodnota stavu D je rovna hodnotě stavu C v předcházejícím časovém okamžiku. Inverze naopak změnil vstupní hodnotu z nuly na jedničku a naopak. Pravidla Nula a Jedna vstup úplně ignorují a vždy zvolí jako stav stejné číslo. To ovšem neznamená nic jiného, než že hrana v grafu je celkem nadbytečná, neboť nepopisuje žádnou interakci.

Když nyní zadáme, jak graf vypadá, a zadáme-li rovněž pro každý z vrcholů nějaké z pravidel Identita, Inverze, Nula nebo Jedna, je časový vývoj stavu systému

jednoznačně dán. Uvažujme například následující graf o šesti vrcholech.



To je opět graf, ve kterém každý vrchol má jen jediný vstup. Zvolme například tato pravidla pro vrcholy: pravidlo Nula pro vrchol A, pravidlo Inverze pro vrchol B, pravidlo Identita pro C, pravidlo Nula pro D a pravidlo Identita pro vrcholy E a F. Zvolme počáteční stav systému jako uspořádanou šestici 111111 (zde první číslo značí hodnotu ve vrcholu A, druhé hodnotu ve vrcholu B atd.). Pak stavy systému v následujících časových krocích jsou 001011, 010001, 011000, 011000, 011000, ... Vidíme, že se stav systému po chvíli ustálí na jednom pevném bodě. Zde můžeme vidět nádhernou analogii s matematickými vlastnostmi diferenčních rovnic, kde jsme také viděli, že se řešení systému může blížit k nějakému pevnému bodu.

Uvažujme nyní tentýž šestiúhelníkový graf, ale zvolme jiná pravidla pro vrcholy. Například zvolme pravidlo Identita pro vrcholy A, C, E a pravidlo Inverze pro vrcholy B, D, F. Zvolme stejnou počáteční podmínku jako prve, tj. počáteční stav systému je 111111. Potom stav systému v následujících časech je roven 101010, 000000, 010101, 111111, 101010, 000000, 010101, 111111, ... Objevili jsme tudíž periodické řešení s periodou délky čtyři, tj. cyklus délky 4. Čtenář si opět může všimnout zajímavé analogie s periodickými řešeními diferenčních rovnic, která jsme objevili v první části seriálu.

Modely plné nul a jedniček, které jsme doposud studovali, se nazývají *booleovské modely* (anglicky boolean networks). Není zase příliš obtížné pokračovat v jejich analýze. To si můžeš ostatně sám vyzkoušet při řešení příkladů 5 a 6 seriálové série. Nic víc než středoškolské znalosti k jejich řešení není potřeba.

Jeden by samozřejmě mohl namítnout, jakou použitelnost mají modely založené jen na nulách a jedničkách. Krásným příkladem je modelování vývoje embrya mouchy *Drosophila melanogaster*. To je muška, která nepořádným studentům na kolejích běžně vylétá z nevynesených odpadkových košů a kterou National Institutes of Health zvolil za jeden z modelových systémů pro biologických výzkum. Její obrázek následuje.