

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

Stochastické modelovanie génovej expresie

Diplomová práca

Bratislava, 2013

Bc. Hana Miklóssyová

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY



Stochastické modelovanie génovej expresie

Diplomová práca

Evidenčné číslo: 176f1c56-941a-4776-be87-b07a0217cd1f

Študijný program: Ekonomická a finančná matematika
Študijný odbor: 1114 Aplikovaná matematika
Školiteľ: Mgr. Pavol Bokes, PhD.

Bratislava, 2013

Bc. Hana Miklóssyová



Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Hana Miklóssyová
Študijný program: ekonomická a finančná matematika (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: 9.1.9. aplikovaná matematika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: slovenský

Názov: Stochastické modelovanie génovej expresie

Cieľ: Cieľom práce bude analyzovať matematický model pre dynamiku stochastickej génovej expresie. Model bude popísaný Markovovským procesom so spojitým časom a s diskretnou stavovou premennou, a skúmaný vhodnými simulačnými a analytickými metódami. V práci nás bude zaujímať efekt oneskorenia, vzniklého v dôsledku prepisu genetickej informácie, na kvalitatívne vlastnosti rozdelenia počtu génových produktov a aktivity génu.

Vedúci: Mgr. Pavol Bokes, PhD.
Katedra: FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky
Vedúci katedry: prof. RNDr. Daniel Ševčovič, CSc.
Dátum zadania: 25.01.2012

Dátum schválenia: 26.01.2012
prof. RNDr. Daniel Ševčovič, CSc.
garant študijného programu

.....
študent

.....
vedúci práce

Čestné prehlásenie

Prehlasujem, že som túto prácu vypracovala samostatne, s použitím uvedenej literatúry a svojich poznatkov.

V Bratislave, apríl 2013

.....
Hana Miklóssyová

Pod'akovanie

Rada by som pod'akovala vedúcemu práce Pavlovi Bokesovi za cenné rady a pripomienky pri písaní tejto práce. Taktiež ďakujem svojim rodičom a priateľom za podporu.

Abstrakt

MIKLÓSSYOVÁ, Hana: Stochastické modelovanie génovej expresie [Diplomová práca], Univerzita Komenského v Bratislave, Fakulta matematiky, fyziky a informatiky, Katedra aplikovanej matematiky a štatistiky; Vedúci diplomovej práce: Mgr. Pavol Bokes, PhD., Bratislava, 2013.

Práca sa zaoberá stochastickým modelovaním génovej expresie, ktorá je viacstupňovým procesom prebiehajúcim na molekulovej úrovni. Oboznámime sa s teóriou stochastických Markovovských procesov a základným modelom génovej expresie, ktorý popisuje aktivitu génu a počet bielkovín. Opierajúc sa o výsledky experimentálnych štúdií sa pokúsime obohatiť model o prvok oneskorenia s cieľom ovplyvniť výsledné stacionárne pravdepodobnostné rozdelenie. Analýzou kvantitatívnych a kvalitatívnych vlastností modelov bez oneskorenia a s oneskorením rozoberieme všetky možné prípady, ktoré môžu nastať. Zvláštnu pozornosť venujeme prezentovaniu výsledkov modelu pre hodnoty parametrov, ktoré boli získané z publikovaných experimentálnych štúdií.

Kľúčové slová: génová expresia, stochastický model, Markovov proces, model s oneskorením, model so spojeným časom, stacionárne pravdepodobnostné rozdelenie

Abstract

MIKLÓSSYOVÁ, Hana: Stochastic modeling gene expression - [Master Thesis], Comenius University in Bratislava, Faculty of Mathematics, Physics and Informatics, Department of Applied Mathematics and Statistics; Supervisor: Mgr. Pavol Bokes, PhD., Bratislava, 2013.

The thesis deals with stochastic modeling of gene expression, which is a multistage process at the molecular level. We review the fundamentals of the stochastic theory of Markov processes and introduce a basic gene expression model, which describes the temporal dynamics of the activity of a gene and the level of its protein product. Based on the results of experimental studies, we integrate the mechanism of transcriptional delay into the model, aiming to characterize its impact on the resulting stationary probability distribution. With the help of quantitative and qualitative model analysis, we discuss all possible cases that may arise. Specifically, we focus on the behavior of the model in the parameter regimes which are consistent with the results of experimental studies.

Key words: gene expression, stochastic model, Markov process, delayed model , continuous-time model, stationary probability distribution

Obsah

Úvod	3
1 Génová expresia	5
1.1 Transkripcia	5
1.2 Translácia	5
1.3 Výsledky experimentálnych štúdií	6
1.3.1 Aktivita génu	6
1.3.2 Korelácia počtu mRNA a počtu molekúl bielkovín	7
2 Teoretický úvod	8
2.1 Exponenciálne rozdelenie	8
2.2 Markovove procesy	9
2.2.1 Homogénne Markovove procesy	10
2.2.2 Kolmogorove diferenciálne rovnice	10
2.2.3 Limitné vlastnosti	11
2.3 Elementárne Markovove procesy	12
2.3.1 Poissonov proces	12
2.3.2 Procesy vzniku a zániku	13
2.4 Vlastnosti vybraných funkcií	14
3 Stochastický model génovej expresie	16
3.1 Základný Markovovský model génovej expresie	17
3.1.1 Kolmogorovove rovnice	17
3.1.2 Vytvárajúca funkcie	19
3.1.3 Pravdepodobnostné rozloženie	22
3.2 Markovovský model génovej expresie s oneskorením	23
3.2.1 Vplyv oneskorenia na model	23
3.2.2 Kolmogorovove rovnice	25
4 Analýza citlivosti modelov	27
4.1 Určenie časovej jednotky	27
4.2 Relatívny vzťah mierových konštánt	28
4.3 Prototypy simulácia	28
4.4 Simulácie s oneskorením	29
4.5 Testovanie na reálnych dátach	30
Záver	32

Literatúra	34
A Príloha	36

Úvod

S pokrokom technológií sme schopní hlbšie preniknúť do chápania okolitého sveta a zužitkovať novonadobudnuté znalosti. Okrem toho, že nám progres umožnil nájsť odpoveď na doposiaľ nezodpovedané otázky, otvára nám aj dvere do doteraz nepreskúmaných oblastí.

Génová expresia je v biológii relatívne nová a nie úplne preskúmaná oblasť. Jedná sa o viacstupňový proces syntetizácie bielkovín na molekulovej úrovni. Je zaujímavé, že susedné bunky toho istého organizmu preukazujú medzi sebou veľké rozdiely v počtoch bielkovín. Pokiaľ sa nám podarí bližšie pochopiť tento proces a jeho správanie, môžeme dospieť k zaujímavým prelomovým výsledkom. Zavedenie nových experimentálnych postupov umožnilo presnejšie preskúmať proces génovej expresie a prinieslo prekvapivé výsledky, ktoré nás nútia prehodnotiť doposiaľ vyvinuté modely. V našej práci si jeden takýto model predstavíme a budeme sa snažiť vylepšiť jeho vlastnosti zavedením prvku oneskorenia tak, aby sa modelované výsledky priblížili k reálnym dátam.

Na začiatku tejto práce sa bližšie oboznámime so samotným procesom génovej expresie. Popíšeme si jednotlivé stupne a zameriame sa na biologické javy, ktoré môžu spôsobiť oneskorenie medzi začatím procesu tvorby bielkovín a dokončením hotového produktu. Predstavíme si nedávne výsledky experimentálnych štúdií, ktoré budú pre nás relevantné pri budovaní modelu.

Po predstavení biologickej stránky potrebnej pri položení základov modelu sa zameriame na samotnú matematickú teóriu. Základné procesy génovej expresie budeme modelovať s využitím stochastickej teórie Markovových procesov so spojitým časom. Pozrieme sa na elementárne Markovove procesy, ako je proces vzniku a zániku alebo Poissonov proces. V nasledujúcich výpočtoch pravdepodobnostného rozdelenia sa stretneme s potrebou aplikovania vlastností niektorých konkrétnych funkcií, ktoré si tiež uvedieme v tejto časti.

Dostávame sa k základnému Markovovskému modelu. Bude nás zaujímať vplyv dynamiky génovej expresie na stacionárne pravdepodobnostné rozdelenie bielkovín. Počet bielkovín bude závislý od aktivity génu. Pokiaľ je gén aktívny, tak dochádza k syntetizácii bielkovín. V prípade neaktívneho stavu nedochádza k žiadnej tvorbe bielkovín a počet bielkovín len degraduje. Po dopočítaní stacionárneho pravdepodobnostného rozdelenia prechádzame na model obohatený o oneskorenie. Biologickú stránku oneskorenia sme už načrtli v prvej kapitole, ďalej sa zameriame na dopad zavedenia takejto zmeny na celkové vlastnosti. Pri výpočtoch pravdepodobnostného rozdelenia modelu s onekorením využijeme metódu neurčitých koeficientov a Kolmogorovove rovnice.

Predtým, ako začneme voliť jednotlivé parametre modelu, zvolíme si časovú škálu, ktorá nám najlepšie vystihuje model a pomôže interpretovať výsledky. Nakoľko stationárne pravdepodobnostné rozloženie je také, ktoré je nezávislé od času, zameriame sa na analýzu vzťahov jednotlivých konštantných mier modelu. Nakoniec si predstavíme aj jednu simuláciu vytvorenú z dostupných dát. Táto kapitola nám poskytne odpoveď na otázku, či sa nám podarilo zavedením oneskorenia vylepšiť vlastnosti modelu a priblížiť sa k reálnym pozorovaniam.

1 Génová expresia

Génová expresia je viacstupňový proces prebiehajúci na molekulovej úrovni. V tomto procese dochádza k prepisu informácie zo samotného génu do mediátorovej RNA (*mRNA*), vďaka ktorému dochádza k tvorbe génových produktov. Týmito produktami sú najčastejšie bielkoviny. V tomto procese dochádza aj k tvorbe RNA.

Proces génovej exprese prebieha vo všetkých organizmoch od eukaryotov cez prokaryoty. Do určitej miery je tento proces pozorovateľný aj vo vírusoch. Základné stupne, do ktorých môžeme génovú expresiu rozdeliť, sú prepis (*transkripcia*), *posttranskripčné modifikácia*, preklad (*translácia*) a *posttranslačná modifikácia*. Proces génovej exprese je zložitý a musí byť regulovaný na všetkých úrovniach. V genetike je génová expresia fundamentálnym kameňom *genotypu*, ktorý je deterministický pre všetky pozorovateľné vlastnosti a znaky organizmu (*fenotyp*). Najväčšiu úlohu v ovplyvnení fenotypu majú bielkoviny a ich syntetizácia.

1.1 Transkripcia

V procese transkripcie dochádza k tvorbe jednovláknovej molekuly mRNA, ktorá je komplementárna k prepisovanému DNA. K prepisu dochádza postupne, komplementárne sa kopíruje väzba kodónu a antikodónu a vlákno mRNA sa postupne predlžuje. Tento proces sa nazýva *elongácia* transkripcie. V prokaryotických organizmoch dochádza k transkripcii v cytoplazme. V eukaryotických organizmoch je proces o niečo zložitejší. K samotnému prepisu informácie dochádza v jadre bunky, ale predtým ako sa zrelé mRNA dostane do cytoplazmy za účelom translácie, musí primárny transkript absolvovať posttranskripčné úpravy. Typickou posttranskripčnou modifikáciou pre eukaryoty je zostrih, počas ktorého dochádza k odstráneniu intrónov. Zostrihovanie je jedným z hlavných regulátorov eukaryotických buniek.

1.2 Translácia

Translácia je proces, v ktorom zrelé vlákno molekuly mRNA je použité ako šablóna pri tvorbe bielkovín. Podobne ako v transkripcii dochádza k tomuto procesu postupne na základe komplementarity v procese elongácie. V prípade eukaryotov obsahuje mRNA sekvenciu potrebnú pre tvorbu jednej molekuly bielkoviny, zatiaľ čo v prípade prokaryotov obsahuje viacero bielkovinových sekvencií. Počas a po syntéze bielkoviny dochádza k zmene štruktúry samotnej molekuly bielkoviny. Molekula prechádza od svojho primárneho stavu k sekundárnemu a nakoniec k svojmu terciárnemu stavu, kde dochádza k zmene štruktúry na trojdimeziálnu. Vo svojej terciárnej štruktúre

sú bielkoviny schopné plniť rôzne funkcie od molekulárnych identifikátorov až po katalyzátory.

1.3 Výsledky experimentálnych štúdií

1.3.1 Aktivita génu

Nedávne experimentálne štúdie ukázali, že identické populácie baktérii a kvasiniek môžu vykazovať veľkú variáciu medzi susednými bunkami v počte proteínov, ktoré sú produkované génmi. Tieto rozdiely spôsobujú zvýšenú fenotypickú rozmanitosť. Faktory, ktoré spôsobujú tieto variácie, môžeme klasifikovať na základe ich pôvodu do dvoch skupín:

1. okolité, vonkajšie faktory ako je napríklad rozdielny počet transkripčných aktivátorov,
2. náhodné, vnútorné, molekulárne činitele ako je transkripcia mRNA a translácia bielkovín.

Autori štúdií [7, 8] sa zamerali na vnútorné náhodné faktory a odôvodnili odlišnosť medzi jednotlivými bunkami náhodnými prechodmi samotného génu medzi aktívnym a neaktívnym stavom. V aktívnom stave dochádza vo vyššej miere k transkripcii, v neaktívnom stave dochádza k prepisu v oveľa menšej miere [2]. Iné štúdie robené na bunkách vyšších eukaryotov tiež poukazujú na významné výchyľky v počte vyprodukovaných bielkovín medzi susednými bunkami týchto organizmov [9, 10, 11].

Avšak priamo určiť aktivitu a neaktivitu génu nie je možné na základe počtu proteínov. Novovytvorené bielkoviny z poslednej aktivity génu nebolo možné oddelene počítat, nakoľko sa v bunke ešte stále vyskytovali bielkoviny, ktoré boli produktami predošlej aktivity génu. Toto prekrytie je spôsobené dlhou životnosťou sledovaných bielkovín. Ďalší problém sa vyskytol pri samotnom počítaní fluorescenčných proteínov. Pretože individuálne molekuly fluorescenčných bielkovín produkujú len malé množstvo fluorescence, je ťažké odhaliť ich nízke počty medzi ostatnými produktami mnohých génov.

Vhodnejší spôsob poukázania aktivity génu je zamerať sa na počet jednotlivých vlákien mRNA, nakoľko majú v porovnaní s bielkovinami oveľa kratšiu životnosť. Hladina ich počtu nám presnejšie popisuje pozorovateľný stav aktivity génu.

V ďalšej štúdií vedci ukázali priamy dôkaz prechodu génu medzi aktívnym stavom a neaktívnym. Použitím fluorescenčnej in situ hybridizácie (FISH) spočítali molekuly mRNA v bunkách cicavcov. Získaním presného počtu mRNA, ktoré sa rýchlo rozpadá, boli schopní ukázať, že zriedkavo dochádza k zmene stavov z aktívneho (s veľkým počtom mRNA) do stavu neaktívneho (nízky počet mRNA) a naopak [6].

1.3.2 Korelácia počtu mRNA a počtu molekúl bielkovín

Ako sme už uviedli, aktivitu génu môžeme odsledovať na základe momentálneho počtu mRNA s relatívne krátkou životnosťou v bunke. Dalo by sa predpokladať, že bude existovať nejaká signifikantná korelácia medzi počtom mRNA a počtom molekúl bielkovín v bunke, nakoľko mRNA slúži ako vzor pri syntéze bielkovín, a teda čím viac by bolo prítomných mRNA v bunke, tým viac sa zvýši produkcia bielkovín. Toto, z hľadiska priebehu génovej expresie, logické tvrdenie bolo z časti vyvrátené v štúdiu [12].

Taniguchiho a kol. spravili kvantitatívnu analýzu proteínov a mRNA v jednotlivých bunkách baktérií *Escherichia coli* [12]. Jeden z výsledkov ich štúdie potvrdil vysokú fluktuáciu počtu bielkovín susedných buniek v isogénnom prostredí. Ale pre nás zaujímavejší a prekvapivejší záver ich práce bol, že v samotnej bunke neexistuje žiadna korelácia medzi hladinou mRNA a proteínov v konkrétnom čase.

Táto absencia korelácie by sa dala z časti zdvôvodniť relatívne rýchlou degradáciou mRNA oproti bielkovinám. V *E. Coli*, mRNA zaniká zvyčajne v priebehu pár minút, zatiaľ čo väčšina proteínov vrátane fluorescenčných má životnosť dlhšiu ako je samotný cyklus bunky. Z toho nám vyplýva, že počet mRNA v čase nám zobrazuje nedávny priebeh aktivity (pár minút), kým hladina molekúl bielkovín v tom istom čase zaznamenáva dlhodobejší stav (v časovom rozsahu bunkového cyklu) reprezentujúci aj predošlú aktivitu génov. Toto ale nestačí na vysvetlenie skoro nulovej korelácie v konkrétnom čase, musí byť ešte prítomný nejaký externý translačný šum alebo regulačný systém. Treba podotknúť, že merania jednej bunky sa prevádzali vo fixnom bode v čase, čo neprotirečí tomu, že počet vzniknutých mRNA za dlhší časový interval by mal byť korelovaný počtu molekúl syntetizovaných za tú istú časovú periódu, ako bolo pozorované v [13].

Jedna z ďalších príčin skoro nulovej korelácie môže byť v oneskorení syntézy bielkovín, ktoré je spôsobené posttranskripčnými modifikáciami alebo elongáciou translácie. Špeciálne v prípade eukaryotických organizmov toto časové posunutie spôsobuje zostih intónov a v prípade prokaryotických organizmov je za to zodpovedná aj väčšia dĺžka prepisovaných informačných sekvencií. Netreba zabúdať na posttranslačnú modifikáciu.

2 Teoretický úvod

Predtým ako sa začneme zaoberať samotným modelovaním génovej expresie, je potrebné sa oboznámiť s teoretickým základom, na ktorom sú modely postavené a na ktorý sa aj budeme neskôr odvolávať. Okrem samotnej teórie Markovovských reťazcov so spojitým časom si uvedieme aj niektoré vlastnosti funkcií, ktoré nám pomôžu pri neskorších výpočtoch. Obsirnejšie informácie k teórii môžeme nájsť v literatúre [4, 5]

2.1 Exponenciálne rozdelenie

Ako budeme modelovať náhodný čas aktivity génu a počet bielkovín v bunke?

Z abstraktného hľadiska sa jedná o udalosti rovnakého typu, ktoré sa opakujú, akou je dĺžka aktivity/neaktivity génu alebo vznik/zánik bielkovín. Tieto udalosti sú náhodné a navzájom celkom nezávislé. Výskyt udalosti v minulosti nám nevie presne povedať, kedy nastane udalosť v budúcnosti. Pravdepodobnosť nastania udalosti narastá s dĺžkou doby, ktorá prešla od posledného výskytu.

Spolu s nezávislosťou udalostí predpokladáme, že podmienky ich výskytu sa v čase nemenia a označíme $p_0(t)$ pravdepodobnosť toho, že udalosť nenastane v časovom intervale dĺžky t . Zrejme bude platiť

$$p_0(0) = 1, \quad (2.1)$$

$$p_0(t + s) = p_0(t)p_0(s), \quad t, s \geq 0. \quad (2.2)$$

Teda udalosť nenastane v intervale dĺžky $t + s$ práve vtedy, ak nenastane na intervale dĺžky t , ani na intervale dĺžky s a tieto udalosti sú podľa predpokladu nezávislé. Pravdepodobnosť výskytu viacerých udalostí v tom istom čase je nulová.

K predpokladu nezávislosti v čase a nemennosti podmienok v čase (homogénosti) pridáme podmienku, že funkcie $p_0(t)$ je spojitá. Potom platí

$$p_0(t) = e^{-\lambda t} \quad (2.3)$$

pre všetky $t \geq 0$ reálne.

Pravdepodobnosť toho, že udalosť nastane v intervale dĺžky t aspoň jeden raz, je

$$1 - p_0(t) = 1 - e^{-\lambda t}. \quad (2.4)$$

Inak povedané, je to pravdepodobnosť toho, že dĺžka intervalu τ medzi dvoma po sebe nasledujúcimi udalosťami je $\leq t$, a teda distribučná funkcia τ .

Jej priemerná hodnota je

$$\begin{aligned}
 E(\tau) &= \int_0^\infty \tau \frac{\partial}{\partial \tau} (1 - e^{-\lambda \tau}) d\tau \\
 &= - \int_0^\infty \tau \frac{\partial}{\partial \tau} e^{-\lambda \tau} d\tau \\
 &= [\tau e^{-\lambda \tau}]_0^\infty + \int_0^\infty e^{-\lambda \tau} d\tau \\
 &= \frac{1}{\lambda}.
 \end{aligned}$$

To značí, že $1/\lambda$ má význam priemernej dĺžky intervalu medzi dvoma udalosťami. Parameter λ má teda význam priemerného počtu udalostí za jednotku intervalu.

Dĺžku aktivity a neaktivity génu vieme modelovať ako náhodnú premennú s exponenciálnym rozložením. Hoci čas medzi posledným vznikom aj zánikom bielkovín vieme modelovať tiež použitím exponenciálneho rozdelenia, na opis dynamiky ich počtu to nestačí. Potrebujeme poznať ich počet ako funkciu času, označíme ju $X(t)$. Pre každé t je $X(t)$ náhodnou premennou, nadobúdajúcou celé nezáporné čísla.

2.2 Markovove procesy

Stochastický proces sa v literatúre uvádza ako súbor (systém) náhodných premenných $X = X(t), t \in T$, kde T je množina alebo podmnožina nezáporných reálnych čísel. Množina T sa najčastejšie interpretuje ako čas, tak je tomu aj v našom prípade. Ak $T = t_0, t_1, t_2, \dots$, hovoríme o *procese s diskrétnym časom*. Ak $T = \langle 0, \infty \rangle$, hovoríme o *procese so spojitým časom*. $X(t)$ popisuje stav procesu v čase t . Procesy s konečnou alebo spočítateľnou množinou všetkých hodnôt S nazývame aj *reťazcami*. Množinu S nazývame *množinou stavov reťazca*. Náhodný reťazec so spojitým časom $X = X(t), t \in T$ s množinou stavov S nazveme *Markovov proces*, ak

1. množina $T = \langle 0, \infty \rangle$,
2. platí Markovova vlastnosť:

$$\begin{aligned}
 &\forall t_0, \dots, t_{n+1} \in T : t_0 < t_1 < \dots < t_{n+1}, \forall i_0, \dots, i_{n-1}, i, j \in S : \\
 &P(X(t_{n+1}) = j | X(t_n) = i, X(t_{n-1}), \dots, X(t_0) = i_0) = \\
 &= P(X(t_{n+1}) = j | X(t_n) = i).
 \end{aligned} \tag{2.5}$$

Ak $X(t) = i$, potom hovoríme, že *proces je v čase t v stave i* .

2.2.1 Homogénne Markovove procesy

Markovov proces nazveme homogénny (v čase), ak platí

$$\begin{aligned} \forall t_1, t_1 + h, t_2, t_2 + h \in T, \quad \forall i, j \in S : \\ (X(t_1 + h) = j | X(t_1) = i) = P(X(t_2 + h) = j | X(t_2) = i). \end{aligned} \quad (2.6)$$

Pravdepodobnosť prechodu zo stavu i do stavu j za čas h označíme

$$p_{ij}(h) = \begin{cases} P(X(t+h) = j | X(t) = i) & \text{ak } h > 0, \\ \delta_{ij} & \text{ak } h = 0. \end{cases} \quad (2.7)$$

Pravdepodobnosť prechodu zo stavu i do stavu j za čas h označíme

$$\mathbf{P}(h) = (p_{ij}(h))_{i,j \in S}. \quad (2.8)$$

Pravdepodobnosť, že proces sa v čase t nachádza v stave j , označíme

$$p_j(t) = P(X(t) = j). \quad (2.9)$$

Ich usporiadanie v tvare vektora

$$\mathbf{p}(t) = (p_j(t))_{j \in S} \quad (2.10)$$

nazývame *pravdepodobnostné rozdelenie procesu* (v čase t). Počiatočným rozdelením procesu rozumieme $\mathbf{p}(0)$. Z elementárnych vlastností podmienených pravdepodobnosťami vyplýva pre ľubovoľné $r, s \geq 0$

$$\sum_{k \in S} p_{ik}(r) p_{kj}(s) = p_{ij}(r+s), \quad (2.11)$$

čo značí

$$\mathbf{P}(r+s) = \mathbf{P}(r)\mathbf{P}(s) \quad \forall r, s \in T. \quad (2.12)$$

Ďalej zrejme platí

$$\mathbf{P}(0) = \mathbf{I}. \quad (2.13)$$

2.2.2 Kolmogorove diferenciálne rovnice

Naším cieľom je vypočítať pravdepodobnostné rozdelenie Markovovho procesu $\mathbf{p}(t)$. Výpočet rozdelenia reťazca v ľubovoľnom okamihu dostaneme z riešenia sústavy homogénnych lineárnych diferenciálnych rovníc s konštantnými koeficientami. Na popis homogénneho Markovovho reťazca so spojitým časom použijeme *intenzity prechodu zo stavu i do stavu j*

$$\lambda_{ij}(h) = \begin{cases} \lim_{h \rightarrow 0^+} \frac{p_{ij}(h)}{h} & \text{ak } i \neq j, \\ \lim_{h \rightarrow 0^+} \frac{1 - p_{ii}(h)}{h} & \text{ak } i = j. \end{cases} \quad (2.14)$$

Ich usporiadanie v tvare matice $\mathbf{Q} = (q_{ij})_{i,j \in S}$, kde

$$q_{ij} = \begin{cases} \lambda_{i,j} & \text{ak } i \neq j, \\ -\lambda_{i,j} & \text{ak } i = j \end{cases} \quad (2.15)$$

sa nazýva *maticou intenzít prechodu*.

Ďalej platí

$$\mathbf{p}(t+h) = \mathbf{p}(t)\mathbf{P}(h) \quad \forall t, h \in T, \quad (2.16)$$

z čoho vyplýva, že ak je známe pravdepodobnostné rozloženie $\mathbf{p}(0)$, potom pre rozloženie $\mathbf{p}(t)$ procesu $X(t)$ platí

$$\mathbf{p}(t) = \mathbf{p}(0)\mathbf{P}(t). \quad (2.17)$$

Dostávame sa k diferenciálnym rovniciam

$$\frac{\partial}{\partial t} \mathbf{p}(t) = \mathbf{p}(0)\mathbf{P}(t)\mathbf{Q} = \mathbf{p}(t)\mathbf{Q}, \quad (2.18)$$

ktoré nazývame *Kolmogorovovými diferenciálnymi rovnicami*.

2.2.3 Limitné vlastnosti

Stacionárnym nazveme také rozloženie π , pre ktoré z $\mathbf{p}(0) = \pi$ vyplýva $\mathbf{p}(t) = \pi$ pre každé $t \geq 0$. Zrejme pre takéto rozloženie musí platiť $\partial \mathbf{p}(t)/\partial t = 0$ a preto je *stacionárne rozdelenie* $\pi = (\pi_j)_{j \in S}$ určené riešením

$$\pi \mathbf{Q} = 0, \quad \sum_{j \in S} \pi_j = 1, \quad \pi_j > 0. \quad (2.19)$$

Ak má Markovov proces so spojitým časom stacionárne rozloženie, všetky jeho stavy sú navzájom dosiahnuteľné a jeho intenzity prechodu sú konečné, potom existuje jediné pravdepodobnostné rozloženie π , spĺňajúce (2.19) a pre všetky pravdepodobnostné rozloženia $\mathbf{p}(t) = \mathbf{p}(0)\mathbf{P}(t)$ platí

$$\lim_{t \rightarrow \infty} \mathbf{p}(t) = \pi > 0. \quad (2.20)$$

Stav j je dosiahnuteľný zo stavu i , ak platí $p_{ij}(t) > 0$, pre nejaké t , čo je ekvivalentné vzťahu

$$\lim_{t \rightarrow \infty} p_{ij}(t) = \pi_j \quad (2.21)$$

pre ľubovoľné i .

2.3 Elementárne Markovove procesy

2.3.1 Poissonov proces

Nech $X(t)$ je bodový proces, t.j. má zmysel výskytu náhodnej premennej udalosti v intervale dĺžky t . Predpokladáme, že udalosti sú na seba nezávislé a homogénne. Ako $p(i)$ označíme pravdepodobnosť výskytu i udalostí v intervale $[0, t)$. Predpokladajme ďalej, že priemerný počet výskytu udalostí za čas t je λt . Potom intenzity prechodu procesu budú spĺňať

$$\begin{aligned} q_{ii} &= -\lambda, \\ q_{i,i+1} &= \lambda, \\ \text{a } q_{ij} &= 0 \text{ pre } j \neq i, i+1 \end{aligned} \tag{2.22}$$

a podľa (2.18) pre pravdepodobnosť $p_i(t)$ platí

$$\begin{aligned} \frac{\partial}{\partial t} p_0(t) &= -\lambda p_0(t), \\ \frac{\partial}{\partial t} p_i(t) &= -\lambda p_i(t) + \lambda p_{i-1}(t) \text{ pre } i > 0(t). \end{aligned} \tag{2.23}$$

Zrejme platí $p_0(0) = 1, p_i(0) = 0$ pre $i > 0$. Postupným riešením rovníc (2.22) pri týchto počiatočných podmienkach dostávame

$$\begin{aligned} p_0(t) &= e^{-\lambda t}, \\ p_i(t) &= \frac{(\lambda t)^i}{i!} e^{-\lambda t} \text{ pre } i > 0. \end{aligned} \tag{2.24}$$

Pre každé t je rozloženie $\{p_0(t), p_1(t), \dots\}$ Poissonovo, preto tomuto procesu hovoríme *Poissonov proces*.

Ako $r_k(t)$ označíme pravdepodobnosť toho, že v intervale dĺžky t nastane najviac k udalostí. Zrejme platí

$$r_k(t) = e^{-\lambda t} \sum_{i=0}^k \frac{(\lambda t)^i}{i!}. \tag{2.25}$$

Rozloženie $t \rightarrow 1 - r_k(t)$ je *Erlangovo* s parametrami λ, k . Opisuje nám rozdelenie času čakania na k -tu udalosť.

2.3.2 Procesy vzniku a zániku

Medzi elementárne procesy s množstvom praktických aplikácií v demografii, hromadnej obsluhu a teórii kozmického žiarenia patria procesy *vzniku a zániku*, známe aj ako procesy *rodenia a úmrtí*. Jedná sa o model populácie, v ktorej z niektorých jedincov vznikajú noví jedinci a iní postupne zanikajú.

Procesom vzniku a zániku rozumieme homogénny Markovov proces s množinou stavov $S = \{i, i \in \mathbb{Z}^+\}$, s počiatočným rozdelením $\mathbf{p}(0) = (0, \dots, 0, 1, 0, \dots)$ a intenzitami prechodu

$$q_{ij}(h) = \begin{cases} \lambda_i & \text{ak } j = i + 1, \\ \mu_i & \text{ak } j = i - 1, \\ -\lambda_i - \mu_i & \text{ak } j = i, i > 0, \\ -\lambda_i & \text{ak } j = 0, i = 0, \\ 0 & \text{ak } |i - j| \geq 2. \end{cases} \quad (2.26)$$

Najznámejšie špeciálne prípady procesu vzniku a zániku sú:

1. *lineárny proces vzniku a zániku*, u ktorého sú intenzity prechodu lineárnymi funkciami stavov, v ktorých bol systém v predošlom okamihu, t.j. $\lambda_i = i\lambda, \lambda > 0$ pre $i = 0, 1, \dots$ a $\mu_i = i\mu, \mu > 0$ pre $i = 0, 1, \dots$,
2. *proces rastu*, keď počet jedincov v systéme s časom len rastie, t.j. $\lambda_i > 0$ pre $i = 0, 1, \dots$,
3. *proces zániku*, keď počet jedincov v systéme s časom len klesá, t.j. $\mu_i > 0$ pre $i = 0, 1, \dots$,
4. *konečný proces vzniku a zániku*, u ktorého je $S = \{0, 1, 2, \dots, n\}$ množina stavov je konečná, a tak $\lambda_i > 0$ pre $i = 0, 1, \dots, n$ a $\mu_i > 0$ pre $i = 0, 1, \dots, n$.

Aby boli splnené rovnice (2.19), musí pre stacionárne rozloženie π platiť

$$\lambda_0 \pi_0(t) = \mu_1 \pi_1(t), \quad (2.27)$$

$$(\lambda_i + \mu) \pi_i(t) = \lambda_{i-1} \pi_{i-1}(t) + \mu_{i+1} \pi_{i+1}(t). \quad (2.28)$$

V prípade homogénneho prípadu sa rovnice redukujú na

$$\lambda_i \pi_i(t) = \mu_{i+1} \pi_{i+1}(t), \quad i = 0, 1, 2, \dots \quad (2.29)$$

Použitím týchto rovníc sa vieme dostať k rekurentnému riešeniu.

Konečný proces vzniku a zániku nazveme ergodický (*regulárny*) so stacionárnym rozdelením $\pi = (\pi_j)_{j=0}^n$

$$\pi_j = \pi_0 \frac{\lambda_0 \lambda_1 \dots \lambda_j - 1}{\mu_1 \mu_2 \dots \mu_j} \text{ pre } 1 \leq j \leq n, \quad (2.30)$$

kde

$$\pi_0 = \left(1 + \sum_{j=1}^n \frac{\lambda_0 \lambda_1 \dots \lambda_j - 1}{\mu_1 \mu_2 \dots \mu_j} \right)^{-1}. \quad (2.31)$$

Tento rekurentný vzťah platí aj pre proces vzniku a zániku s nekonečnou množinou stavov ak nerovnosť

$$\sum_{j=1}^n \frac{\lambda_0 \lambda_1 \dots \lambda_j - 1}{\mu_1 \mu_2 \dots \mu_j} < \infty \quad (2.32)$$

je splnená.

2.4 Vlastnosti vybraných funkcií

Vlastnosti degenerovanej hypergeometrickej funkcie

Degenerovaná hypergeometrická funkcia je definovaná ako nekonečný rad:

$${}_1F_1(a, b, z) = \sum_{n=0}^{\infty} \frac{(a)_n z^n}{(b)_n n!}, \text{ kde}$$

$$(a)_n = a.(a+1).(a+2).\dots.(a+n-1), \quad (a)_0 = 1,$$

je Pochhammerov symbol (stúpajúci faktoriál). Pre deriváciu ${}_1F_1$ platí:

$$\frac{\partial}{\partial z} {}_1F_1(a, b, z) = \frac{a}{b} {}_1F_1(a+1, b+1, z).$$

Všeobecne pre $k \in \mathbb{N}$ sa derivácie správajú nasledovne:

$$\frac{\partial^k}{\partial z^k} {}_1F_1(a, b, z) = \frac{(a)_k}{(b)_k} {}_1F_1(a+k, b+k, z). \quad (2.33)$$

Tento vzorec nám umožňuje popísať hypergeometrickú funkciu ako mocninový rad so stredom rôznym od nuly.

Analytické funkcie

Analytickou funkciou nazývame funkciu, ktorú vieme napísať ako konvergenčný mocninový rad tvaru:

$$\begin{aligned} f(x) &= \sum_{n=0}^{\infty} a_n (x-b)^n \\ &= a_0 + a_1(x-b) + a_2(x-b)^2 + a_3(x-b)^3 + \dots \end{aligned} \tag{2.34}$$

Derivovaním funkcie podľa x n -krát za podmienky, že $x = b$ sa dostávame k prepisu

$$a_n = \frac{1}{n!} \left. \frac{\partial^n f(x)}{\partial x^n} \right|_{x=b} \tag{2.35}$$

pre výpočet koeficientov radu.

3 Stochastický model génovej expresie

Výhoda súčasných počítačov a pokroku v oblasti numerických algoritmov spočíva aj v tom, že kedysi zdĺhavé a únavné postupy numerických výpočtov trvajú v dnešnej dobe pár sekúnd. To viedlo k nárastu popularity matematického modelovania procesov, či už zo sveta biológie, mechaniky alebo astronómie. Veľká výhoda matematických modelov spočíva v ich potenciále nahradiť časovo a finančne náročné pozorovania v testovacom prostredí. Matematický model postavený na základe týchto pozorovaní nám vie rozumne a kvantitatívne popísať výsledky testovania. V niektorých prípadoch je matematický model schopný nahradiť samotné testovacie merania.

Vo všeobecnosti matematický model nie je schopný reprezentovať samotný proces so všetkými podrobnosťami, ale na druhú stranu je schopný opísať najdôležitejšie a najpodstatnejšie pozorované rysy procesu. Samotné modelovanie je adaptívny proces, ktorý nám umožňuje vylepšovať na základe predošlých výsledkov vlastnosti ďalšej generácie modelov.

Pri modelovaní procesu génovej expresie využijeme poznatky nadobudnuté z experimentálnych štúdií spomenutých v druhej kapitole. Nakoľko hladina mRNA so svojou krátkou životnosťou, nám indikuje len krátkodobú históriu aktivity génu, sústredíme sa radšej na prechod génu medzi aktívnym a neaktívnym stavom. Z dlhodobého hľadiska by malo takéto pozorovanie vykazovať podobnú výpovednú hodnotu ako samotný vývoj populácie mRNA v čase. Samozrejme, nevyhnutná návaznosť vzniku bielkovín na mRNA ostáva. V modeli k syntetizácii bielkovín môže dôjsť len vtedy, ak gén je aktívny.

Najskôr si predstavíme základny model, kde gén musí byť aktívny, aby v tom istom čase došlo k vzniku bielkovín. Toto tvrdenie ale nesúhlasí s pozorovaniami v [12]. Na to, aby bol model viac konzistentný s experimentálnymi výsledkami, predstavíme si ďalší model v ktorom zakomponujeme prvok oneskorenia, t.j. na to, aby sa syntetizovali nové bielkoviny netreba, aby gén bol aktívny v tom istom časovom okamihu. Stačí, aby gén bol aktívny v minulosti. Konkrétny moment, kedy musel byť gén aktívny v predošlom čase, bude závisieť od hodnoty oneskorenia.

3.1 Základný Markovovský model génovej expresie

V tejto časti si bližšie rozobereme model, ktorým sa zaoberali Peccoud a Ycart v článku [2]. Tento model zohľadňuje jednotlivé stupne a fázy génovej expresie ako je tvorba a zánik bielkovín, aktivita a neaktivita génu. Model bude popísaný vďaka nasledujúcim vzťahom:

$$I \xrightarrow{\lambda} A, \quad (3.1)$$

$$A \xrightarrow{\mu} I, \quad (3.2)$$

$$A \xrightarrow{\nu} A + P, \quad (3.3)$$

$$P \xrightarrow{\delta} \emptyset, \quad (3.4)$$

kde (3.1) predstavuje prechod génu z aktívneho stavu na stav neaktívny, (3.2) prechod z neaktívneho stavu do aktívneho, (3.3) syntetizáciu molekúl bielkovín a (3.4) degradáciu bielkovín. Symbol $\emptyset$ nám vyjadruje to, že nedochádza k žiadnym chemickým reakciám. Došlo k transformácii proteínu P na nejaký produkt, ktorý už nezapočítavame do nášho modelu. Časový vývoj syntézy bielkovín je modelovaný Markovovským procesom so spojitým časom. Množina stavov tohto procesu je:

$$S = \{(i, n), i \in \{0, 1\}, n \in \mathbb{N}_+\}.$$

V každom stave (i, n) nám i popisuje aktivitu génu:

$i = 0$ keď je gén neaktívny,

$i = 1$ keď je gén aktívny.

V množine stavov nám n predstavuje počet molekúl (kópií) bielkovín P v bunke. Predpokladáme, že časový vývoj stavov génu je Markovovský. Gén zotrúva aktívny priemerne $1/\mu$ časových jednotiek. Dĺžka tejto doby je exponenciálne rozdelená s parametrom μ . Potom gén prejde do neaktívneho stavu, v ktorom ostane priemerne $1/\lambda$ (exponenciálne rozdelenie s parametrom λ). Počas neaktivity génu nedochádza k žiadnej syntéze bielkovín P . Pokiaľ je gén aktívny, dochádza k syntéze bielkovín s konštantnou mierou ν . Životnosť každej bielkoviny P je navzájom nezávislá, rovnako rozdelená a ich spoločné rozdelenie je exponenciálne s parametrom δ .

3.1.1 Kolmogorovove rovnice

Zo stavu $(0, n)$, neaktívny gén s n prvkami P , sa môžeme dostať do stavov:

$(1, n)$ s intenzitou λ (gén prejde do aktívneho stavu),

$(0, n - 1)$ s intenzitou $n\delta$ (degradácia jedného P).

Zo stavu $(1, n)$ aktívny gén s n prvkami P , sa môžeme dostať do stavov:

- $(0, n)$ s intenzitou ν (gén prejde do neaktívneho stavu),
- $(1, n - 1)$ s intenzitou $n\delta$ (degradácia jedného P),
- $(1, n + 1)$ s intenzitou μ (syntéza jedného P).

Takto stanovené prechody medzi stavmi nám definujú Markovovský proces na množine S , ktorý pomenujeme ako IAP proces. Nech $p_0(m, t)$ (respektíve $p_1(m, t)$) je pravdepodobnosť toho, že v čase t je gén aktívny (respektívne neaktívny) a máme prítomných n molekúl P . Ku Kolmogorovým rovniciam IAP procesu sa dostaneme cez intenzity prechodu:

$$\begin{aligned}
\forall n \geq 0, \quad \frac{\partial p_{0,n}(t)}{\partial t} &= -(\lambda + n\delta)p_{0,n}(t) + (n+1)\delta p_{0,n+1}(t) + \mu p_{1,n}(t), \\
\frac{\partial p_{1,0}(t)}{\partial t} &= -(\mu + \nu)p_{1,0}(t) + \delta p_{1,1}(t) + \lambda p_{0,0}(t), \\
\forall n \geq 1, \quad \frac{\partial p_{1,n}(t)}{\partial t} &= -(\mu + \lambda + n\delta)p_{1,n}(t) + (n+1)\delta p_{1,n+1}(t) \\
&\quad + \nu p_{1,n-1}(t) + \lambda p_{0,n}(t).
\end{aligned} \tag{3.5}$$

Na IAP proces sa môžeme pozrieť ako na proces vzniku a zániku s dvoma stavovými Markovovskými prostrediami v spojitom čase. Z takéhoto uhla pohľadu môžeme chápať časový vývoj aktivity a neaktivity génu ako proces prostredia. Stav tohoto procesu nám jednoznačne určuje rozloženie populácie bielkovín v procese vzniku a zániku. Pokiaľ je gén aktívny a miera degradácie bielkovín δ je kladná, tak počet bielkovín sa vyvíja podľa lineárneho procesu vzniku a zániku. Pokiaľ $\delta = 0$, tak bielkovinový proces je Poissonovým procesom, nakoľko nedochádza k žiadnej denaturácii bielkovín.

Stacionárne rozloženie procesu prostredia vieme vypočítať z intenzít prechodu a podmienok (2.19) a (2.29). Dostávame sa k

$$\begin{aligned}
\lambda p_0(m, t) &= \mu p_1(m, t), \\
p_0(m, t) + p_1(m, t) &= 1.
\end{aligned} \tag{3.6}$$

Budeme predpokladať, že rozloženie procesu prostredia bude v čase $t = 0$ v tomto rovnovážnom stave

$$p_0(0, 0) = \frac{\mu}{\lambda + \mu}, \tag{3.7}$$

$$p_1(0, 0) = \frac{\lambda}{\lambda + \mu}. \tag{3.8}$$

3.1.2 Vytvárajúca funkcie

Nech

$$\{p_{i,n}, i \in \{0, 1\}, n \in \mathbb{N}\}$$

je stacionárnym rozdelením IAP procesu, kde $p_{i,n}$ vyjadrujú konštanty, ktoré sú riešeniami Kolmogorových rovníc (3.5). Predstavíme si vytvárajúce funkcie g_0 a g_1 , ktoré nám pomôžu pri hľadaní predpisov pre $p_{0,n}$ a $p_{1,n}$:

$$g_0(z) = \sum_{n=0}^{\infty} z^n p_{0,n}, \quad (3.9)$$

$$g_1(z) = \sum_{n=0}^{\infty} z^n p_{1,n}. \quad (3.10)$$

Na to, aby sme sa dostali k predpisom pre funkcie g_0 a g_1 k predpisom pre $p_{0,n}$ a $p_{1,n}$, využijeme vlastnosť analytických funkcií (2.35).

Namiesto funkcie $f(x)$ v (2.35) si dosadíme funkcie (3.9) a (3.10), kde $b = 0$, $x = z$ a $a_n = p_{i,n}$. Dostávame sa k predpisom:

$$p_{0,n} = \frac{1}{n!} \left. \frac{\partial^n g_0(z)}{\partial z^n} \right|_{z=0}, \quad (3.11)$$

$$p_{1,n} = \frac{1}{n!} \left. \frac{\partial^n g_1(z)}{\partial z^n} \right|_{z=0}. \quad (3.12)$$

Prenásobením Kolmogorových rovníc (3.5) členom z^n a zosumovaním cez $n \geq 0$ zistíme, že vytvárajúce funkcie $g_0(z)$ a $g_1(z)$ spĺňajú systém dvoch lineárnych rovníc s nekonštantnými koeficientami:

$$\delta(z-1)g'_0 = -\lambda g_0(z) + \mu g_1(z), \quad \delta(z-1)g'_1 = \lambda g_0(z) - \mu g_1(z) - \nu(1-z)g_1(z). \quad (3.13)$$

Riešenie budeme hľadať v tvare mocninového radu so stredom v $z = 1$:

$$g_0(z) = \sum_{n=0}^{\infty} a_n(z-1)^n, \quad g_1(z) = \sum_{n=0}^{\infty} b_n(z-1)^n. \quad (3.14)$$

Derivácie vyzerajú nasledovne

$$\frac{\partial g_0(z)}{\partial z} = \sum_{n=0}^{\infty} n a_n (z-1)^{n-1} \quad \frac{\partial g_1(z)}{\partial z} = \sum_{n=0}^{\infty} n b_n (z-1)^{n-1}$$

Snažíme sa zistiť, ako bude vyzeráť a_n a b_n po dosadení do rovníc z (3.13)

$$\begin{aligned} \delta(z-1) \sum_{n=0}^{\infty} n a_n (z-1)^{n-1} &= -\lambda \sum_{n=0}^{\infty} a_n (z-1)^n + \mu \sum_{n=0}^{\infty} b_n (z-1)^n, \\ \delta \sum_{n=0}^{\infty} n a_n (z-1)^n &= -\lambda \sum_{n=0}^{\infty} a_n (z-1)^n + \mu \sum_{n=0}^{\infty} b_n (z-1)^n, \\ \sum_{n=0}^{\infty} \left[(\delta n a_n + \lambda a_n - \mu b_n) (z-1)^n \right] &= 0. \end{aligned}$$

$$\begin{aligned}
\delta(z-1) \sum_{n=0}^{\infty} n b_n (z-1)^{n-1} &= \lambda \sum_{n=0}^{\infty} a_n (z-1)^n - \mu \sum_{n=0}^{\infty} b_n (z-1)^n + \nu \sum_{n=0}^{\infty} b_n (z-1)^{n+1}, \\
\delta \sum_{n=0}^{\infty} n b_n (z-1)^n &= \lambda \sum_{n=0}^{\infty} a_n (z-1)^n - \mu \sum_{n=0}^{\infty} b_n (z-1)^n + \nu \sum_{n=0}^{\infty} b_n (z-1)^{n+1}, \\
\sum_{n=0}^{\infty} \left[(\delta n b_n + \mu b_n - \lambda a_n) (z-1)^n - \nu b_n (z-1)^{n+1} \right] &= 0.
\end{aligned}$$

Podľa (3.7) a (3.8) bude pre $n = 0$ platiť:

$$a_n = \frac{\mu}{\lambda + \mu}, \quad b_n = \frac{\lambda}{\lambda + \mu}.$$

Pre $n \geq 1$ v prípade a_n bude platiť :

$$\begin{aligned}
\delta n a_n + \lambda a_n - \mu b_n &= 0, \\
a_n (\delta n + \lambda) &= \mu b_n, \\
a_n &= \frac{\mu}{\delta n + \lambda} b_n
\end{aligned}$$

V prípade b_n budú rovnice vyzeráť nasledovne:

$$\delta n b_n + \mu b_n - \lambda a_n - \nu b_{n-1} = 0.$$

Dosadíme a_n z predošlej rovnice

$$\begin{aligned}
\delta n b_n + \mu b_n - \lambda \frac{\mu}{\delta n + \lambda} b_n - \nu b_{n-1} &= 0, \\
b_n \left(\delta n + \mu - \frac{\lambda \mu}{\delta n + \lambda} \right) &= \nu b_{n-1}, \\
b_n \left(\frac{(\delta n + \mu)(\delta n + \lambda) - \lambda \mu}{\delta n + \lambda} \right) &= \nu b_{n-1}, \\
b_n \left(\frac{\delta^2 n^2 + (\lambda + \mu) \delta n + \lambda \mu - \lambda \mu}{\delta n + \lambda} \right) &= \nu b_{n-1}, \\
b_n \left(\frac{\delta n (\delta n + \lambda + \mu)}{\delta n + \lambda} \right) &= \nu b_{n-1}.
\end{aligned}$$

To sa dá prepísať do tvaru

$$\begin{aligned}
b_n &= \frac{\delta n + \lambda}{\delta n + \lambda + \mu} \frac{\nu}{\delta n} b_{n-1}, \\
b_n &= \frac{\delta n + \lambda}{\delta(n-1) + \delta + \lambda + \mu} \frac{\nu}{\delta n} b_{n-1}, \\
b_n &= \frac{1}{n} \frac{n \frac{\nu}{\delta} + \frac{\lambda \nu}{\delta^2}}{(n-1) + \frac{\delta + \lambda + \mu}{\delta}} b_{n-1}.
\end{aligned}$$

Vykonaním nasledujúcej substitúcie

$$\begin{aligned}a &= \frac{\lambda + \mu + \delta}{\delta}, \\b &= \frac{\nu}{\delta}, \\c &= \frac{\lambda\nu}{\delta^2}\end{aligned}$$

dostaneme sa k rekurentnému tvaru:

$$b_n = \frac{1}{n} \frac{nb + c}{(n-1) + a} b_{n-1}.$$

Riešenie rekurentného vzťahu má tvar

$$b_n = \frac{1}{n!} \cdot \frac{(c+b)\dots(c+nb)}{a(a+1)\dots(a+n-1)} \cdot \frac{\lambda}{\lambda + \mu}.$$

Podobným spôsobom sa budeme snažiť upraviť a_n

$$\begin{aligned}a_n &= \frac{\mu}{\delta n + \lambda} b_n, \\a_n &= \frac{\mu}{\delta n + \lambda} \cdot \frac{1}{n!} \cdot \frac{\lambda}{\lambda + \mu} \cdot \frac{(c+b)\dots(c+nb)}{a(a+1)\dots(a+n-1)}, \\a_n &= \frac{\mu}{\lambda + \mu} \cdot \frac{1}{n!} \cdot \frac{\lambda}{\delta n + \lambda} \cdot \frac{\frac{\nu}{\delta^2}}{\frac{\nu}{\delta^2}} \cdot \frac{(c+b)\dots(c+nb)}{a(a+1)\dots(a+n-1)}, \\a_n &= \frac{\mu}{\lambda + \mu} \cdot \frac{1}{n!} \cdot \frac{\frac{\lambda\nu}{\delta^2}}{\frac{n\nu}{\delta} + \frac{\lambda\nu}{\delta^2}} \cdot \frac{(c+b)\dots(c+nb)}{a(a+1)\dots(a+n-1)}, \\a_n &= \frac{\mu}{\lambda + \mu} \cdot \frac{1}{n!} \cdot \frac{c}{nb + c} \cdot \frac{(c+b)\dots(c+nb)}{a(a+1)\dots(a+n-1)}, \\a_n &= \frac{\mu}{\lambda + \mu} \cdot \frac{1}{n!} \cdot \frac{c(c+b)\dots(c+(n-1)b)}{a(a+1)\dots(a+n-1)}.\end{aligned}$$

Podarilo sa nám zistiť, ako bude vyzeráť a_n a b_n . Dosadením do (3.14) sa dostávame k tvaru

$$\begin{aligned}g_0 &= \frac{\mu}{\lambda + \mu} \left[1 + \sum_{n=1}^{\infty} \frac{1}{n!} \frac{c(c+b)\dots(c+(n-1)b)}{a(a+1)\dots(a+n-1)} (z-1)^n \right], \\g_0 &= \frac{\mu}{\lambda + \mu} \left[1 + \sum_{n=1}^{\infty} \frac{1}{n!} \frac{\frac{c}{b}(\frac{c}{b}+1)\dots(\frac{c}{b}+(n-1))}{a(a+1)\dots(a+n-1)} b^n (z-1)^n \right], \\g_0 &= \frac{\mu}{\lambda + \mu} {}_1F_1 \left(\frac{c}{b}, a, b(z-1) \right), \\g_0 &= \frac{\mu}{\lambda + \mu} {}_1F_1 \left(\frac{\lambda}{\delta}, \frac{\lambda + \mu + \delta}{\delta}, \frac{\nu}{\delta} (z-1) \right),\end{aligned} \tag{3.15}$$

$$\begin{aligned}
g_1 &= \frac{\lambda}{\lambda + \mu} \left[1 + \sum_{n=1}^{\infty} \frac{1}{n!} \frac{(c+b)\dots(c+nb)}{a(a+1)\dots(a+n-1)} (z-1)^n \right], \\
g_1 &= \frac{\lambda}{\lambda + \mu} \left[1 + \sum_{n=1}^{\infty} \frac{1}{n!} \frac{(\frac{c}{b}+1)\dots(\frac{c}{b}+n)}{a(a+1)\dots(a+n-1)} b^n (z-1)^n \right], \\
g_1 &= \frac{\lambda}{\lambda + \mu} {}_1F_1 \left(\frac{c}{b} + 1, a, b(z-1) \right), \\
g_1 &= \frac{\lambda}{\lambda + \mu} {}_1F_1 \left(\frac{\lambda}{\delta} + 1, \frac{\lambda + \mu + \delta}{\delta}, \frac{\nu}{\delta} (z-1) \right).
\end{aligned} \tag{3.16}$$

kde ${}_1F_1$ je degenerovanou hypergeometrickou funkciou.

3.1.3 Pravdepodobnostné rozloženie

Poznáme už správny tvar pre vytvárajúce funkcie g_0 a g_1 . Na to, aby sme sa dostali k n -tým deriváciám g_0 a g_1 v bode $z = 0$, využijeme vlastnosť ${}_1F_1$ (2.33):

$$\begin{aligned}
\left. \frac{\partial^n g_0(z)}{\partial z^n} \right|_{z=0} &= \frac{\left(\frac{c}{b}\right)_n}{(a)_n} \cdot b^n \cdot {}_1F_1 \left(\frac{c}{b} + n, a + n, b(z-1) \right), \\
\left. \frac{\partial^n g_0(z)}{\partial z^n} \right|_{z=0} &= \frac{\left(\frac{\lambda}{\delta}\right)_n}{\left(\frac{\lambda + \mu + \delta}{\delta}\right)_n} \cdot \left(\frac{\nu}{\delta}\right)^n \cdot {}_1F_1 \left(\frac{\lambda}{\delta} + n, \frac{\lambda + \mu + \delta}{\delta} + n, -\frac{\nu}{\delta} \right), \\
\left. \frac{\partial^n g_1(z)}{\partial z^n} \right|_{z=0} &= \frac{\left(\frac{c}{b} + 1\right)_n}{(a)_n} \cdot b^n \cdot {}_1F_1 \left(\frac{c}{b} + n + 1, a + n, b(z-1) \right), \\
\left. \frac{\partial^n g_1(z)}{\partial z^n} \right|_{z=0} &= \frac{\left(\frac{\lambda}{\delta} + 1\right)_n}{\left(\frac{\lambda + \mu + \delta}{\delta}\right)_n} \cdot \left(\frac{\nu}{\delta}\right)^n \cdot {}_1F_1 \left(\frac{\lambda}{\delta} + n + 1, \frac{\lambda + \mu + \delta}{\delta} + n, -\frac{\nu}{\delta} \right).
\end{aligned} \tag{3.17}$$

Po dosadení derivácií (3.17) a (3.18) sa konečne dostávame k všeobecnému predpisu pre $p_{0,n}$ a $p_{1,n}$

$$p_{0,n} = \frac{1}{n!} \cdot \frac{\left(\frac{\lambda}{\delta}\right)_n}{\left(\frac{\lambda + \mu + \delta}{\delta}\right)_n} \cdot \left(\frac{\nu}{\delta}\right)^n \cdot {}_1F_1 \left(\frac{\lambda}{\delta} + n, \frac{\lambda + \mu + \delta}{\delta} + n, -\frac{\nu}{\delta} \right), \tag{3.19}$$

$$, \quad p_{1,n} = \frac{1}{n!} \cdot \frac{\left(\frac{\lambda}{\delta} + 1\right)_n}{\left(\frac{\lambda + \mu + \delta}{\delta}\right)_n} \cdot \left(\frac{\nu}{\delta}\right)^n \cdot {}_1F_1 \left(\frac{\lambda}{\delta} + n + 1, \frac{\lambda + \mu + \delta}{\delta} + n, -\frac{\nu}{\delta} \right). \tag{3.20}$$

3.2 Markovovský model génovej expresie s oneskorením

Vzťahy budú veľmi podobné ako v pôvodnom modeli:

$$I \xrightarrow{\lambda} A, \quad (3.21)$$

$$A \xrightarrow{\mu} I, \quad (3.22)$$

$$A(t) \xrightarrow{\nu} A(t) + P(t + d), \quad (3.23)$$

$$P \xrightarrow{\delta} \emptyset. \quad (3.24)$$

kde (3.21) predstavuje prechod génu z aktívneho stavu na stav neaktívny, (3.22) prechod z neaktívneho stavu do aktívneho, (3.23) syntetizáciu molekúl bielkovín s oneskorením d a (3.24) degradáciu bielkovín. Symbol $\emptyset$ nám vyjadruje to, že nedochádza k žiadnym chemickým reakciám. Došlo k transformácii P na nejaký produkt, ktorý už nezapočítavame do nášho modelu. Časový vývoj syntézy bielkovín je modelovaný Markovovským procesom so spojitým časom. Množina stavov tohto procesu je:

$$S = \{(a, m), a \in \{0, 1\}, m \in \mathbb{N}_+\}.$$

V každom stave (a, m) nám i popisuje aktivitu génu:

$a = 0$ keď je gén neaktívny,

$a = 1$ keď je gén aktívny.

Ak $(A(t), M(t))$ je definovaný ako stacionárny proces bez oneskorenia, kde $A(t) \in \{0, 1\}$ a $M(t) \in \mathbb{N}_0$. Potom ako stacionárny proces s oneskorením budeme nazývať proces definovaný ako $(A(t), M(t - d))$, kde d je konštantná hodnota oneskorenia.

3.2.1 Vplyv oneskorenia na model

Jedna z výhod základného Markovovského modelu spočíva v jeho jednoduchosti, vďaka ktorej sme boli schopní presne vyjadriť pravdepodobnostné rozloženie procesu. Pri modeli s oneskorením sa jednoduchosť vyjadrenia zachováva, nakoľko proces vzniku a zániku bielkovín dostávame posunutím toho istého procesu základného modelu v čase o konštantnú hodnotu d . Proces prostredia v porovnaní so základným modelom ostáva nezmenený.

Z čisto matematického hľadiska sa ukazuje takéto posunutie v modeli ako korektné, problém by však mohol vzniknúť pri biologickom opodstatnení modelu z pohľadu splnenia podmienok homogenity a nezávislosti. Podobný problém však bol riešený v článku Bokesa a Gedeona [3], ktorí ukázali, že za určitých predpokladov sú základné podmienky splnené.

Zaoberali sa podobným modelom ako je uvedený v tejto práci. Rozdiel nastáva vo formulácii procesu prostredia. V našej verzii modelu používame na popis procesu prostredia aktivitu a neaktivitu génu $A(t)$, ktorá je popísaná množinou stavov $S = \{a, a \in (0, 1)\}$, pričom v modeli z [3] je tento proces nahradený procesom vzniku a zániku $mRNA(t)$ s množinou stavov $S = \{m, m \in \mathbb{N}_+\}$. Ostatné parametre modelu sú rovnaké. Nakoľko náš model spadá do podmnožín modelu, ktorým sa zaoberali Bokes a Gedeon, sú ich predpoklady po určitej adaptácii použiteľné aj v našom prípade.

Pokiaľ platí, že

1. posunutie d je rovnaké a v čase konštantné pre všetky molekuly bielkovín,
2. množstvo bielkovín neovplyvňuje proces zmeny aktivity génu a vzniku bielkovín (t.j. nedochádza k žiadnej spätnej reakcii) ,
3. nekompletné molekuly bielkovín nepodliehajú rozkladu,

tak potom realizáciu procesu vzniku a zániku bielkovín modelu s oneskorením dostaneme posunutím v čase realizácie tohto procesu v modeli bez oneskorenia o hodnotu d . Realizácia $A(t)$ ostáva rovnaká bez ohľadu na veľkosť oneskorenia. Kvôli oneskoreniu dochádza k vzniku každej molekuly o d jednotiek času neskôr od začiatku procesu. Nakoľko neexistuje žiadna spätna väzba, miera syntézy bielkovín ostáva nezmenená a v zmysle pravdepodobnostného rozdelenia rovnaká ako miera modelu bez oneskorenia. Pokiaľ určíme trajektóriu stochastického modelu tvorby bielkovín v čase a tento čas posunieme o hodnotu d , dostaneme trajektóriu modelu s posunutím.

Vznik molekuly bielkoviny nie je jednorázový proces. V čase medzi začatím procesu a dokončením syntézy molekula bielkoviny prechádza zmenou svojej primárnej stavby na sekundárnu a následne terciárnu štruktúru. Až po dosiahnutí terciárnej stavby považujeme molekulu za plne rozvinutú a schopnú biologických interakcií s okolím. Medzi tieto biologické interakcie zahrňame aj degradáciu bielkovín, preto môžeme predpokladať, že k zániku bielkoviny z nášho modelu dochádza až po dokončení oneskorenej syntézy. Životnosť každej molekuly (čas medzi syntézou a zánikom) bude taká istá ako v základnom modeli. Celkovú dobu rozkladu v modeli s oneskorením dostaneme pridaním konštanty d k dobe rozkladu v modeli bez oneskorenia.

Pokiaľ časové okamihy modelu bez oneskorenia, v ktorých dochádza k zmene počtu molekúl bielkovín, či už kvôli procesu vzniku alebo zániku, označíme ako t_i , kde $t_i = 1, 2, \dots$, tak potom časové body $t_i + d$ budú zodpovedať okamihom, keď nastane zmena v modeli s oneskorením. Trajektóriu procesu vzniku a zániku bielkovín modelu

s oneskorením budeme vedieť zapísať ako

$$M_d(t) = M(t - d). \quad (3.25)$$

Na základe vzťahu (3.25) môžeme tvrdiť, že jediný vplyv oneskorenia na populáciu bielkovín je, že všetky zmeny v populácii sa dejú o d jednotiek neskôr. Podmienka homogénnosti prostredia a nezávislosti udalostí naďalej ostáva splnená.

3.2.2 Kolmogorovove rovnice

Poznáme pravdepodobnostné rozloženie $(A(t), M(t))$, budeme chcieť určiť pravdepodobnostné rozloženie $(A(t), M(s))$, kde $d = t - s$ a $t \geq s$. Pravdepodobnosť $p_0(m; t, s)$ a $p_1(m; t, s)$ nám bude vyjadrovať

$$p_0(m; t, s) := P[A(t) = 0 \& M(s) = m], \quad (3.26)$$

$$p_1(m; t, s) := P[A(t) = 1 \& M(s) = m]. \quad (3.27)$$

Pre model s oneskorením budú Kolmogorovove rovnice procesu prostredia vyjadrené ako

$$\frac{\partial}{\partial t} p_0(m; t, s) = -\lambda p_0(m; t, s) + \mu p_1(m; t, s), \quad (3.28)$$

$$\frac{\partial}{\partial t} p_1(m; t, s) = \lambda p_0(m; t, s) - \mu p_1(m; t, s). \quad (3.29)$$

$$(3.30)$$

Sčítaním sa dostávame k :

$$\frac{\partial}{\partial t} (p_0(m; t, s) + p_1(m; t, s)) = 0,$$

z čoho nám vyplýva, že súčet pravdepodobnostných rozložení $p_0(m; t, s)$ a $p_1(m; t, s)$ je konštantný bez ohľadu na veľkosť zmeny v čase.

$$\begin{aligned} p_0(m; t, s) + p_1(m; t, s) &= p_0(m; s, s) + p_1(m; s, s) = \\ &= p_0(m) + p_1(m) = p(m). \end{aligned}$$

Z prvej rovnice systému (3.28) máme:

$$\begin{aligned} \frac{\partial}{\partial t} p_0(m; t, s) &= -\lambda p_0(m; t, s) + \mu(p(m) - p_0(m; t, s)), \\ &= -(\lambda + \mu)p_0(m; t, s) + \mu p(m). \end{aligned} \quad (3.31)$$

Z druhej rovnice (3.29) sa dostaneme k podobnému výsledku:

$$\begin{aligned} \frac{\partial}{\partial t} p_1(m; t, s) &= \lambda(p(m) - p_1(m; t, s)) - \mu p_1(m; t, s), \\ &= -(\lambda + \mu)p_1(m; t, s) + \lambda p(m), \end{aligned} \quad (3.32)$$

$$\frac{\partial}{\partial t}p_0 = \frac{\partial p_0}{\partial d} \cdot \frac{\partial(t-s)}{\partial t} = \frac{\partial p_0}{\partial d},$$

$$\frac{\partial}{\partial d}p_0(m, d) = -(\lambda + \mu)p_0(m, d) + \lambda p(m).$$

V našom prípade hodnoty $p(m)$ poznáme z modelu bez oneskorenia a môžeme ich brať ako konštanté. Diferenciálna rovnica pre $p_0(m, d)$ je lineárna a nehomogénna, preto jej riešenie hľadáme v tvare:

$$p_0(m, d) = p_0^{\text{hom}}(m, d) + p_0^{\text{part}}(m, d), \quad (3.33)$$

kde $p_0^{\text{hom}}(m, d)$ je riešením homogénnej úlohy, t.j.

$$p_0^{\text{hom}}(m, d) = ce^{-(\lambda+\mu)d},$$

kde c je konštanta, zatiaľ čo $p_0^{\text{part}}(m, d)$ je partikulárnym riešením nehomogennej úlohy. Keďže homogenita je konštantná v čase, podľa metódy neurčitých koeficientov [1] hľadáme partikulárne riešenie tiež v tvare konštanty. Dosadením $p_0^{\text{part}}(m, d) \equiv \tilde{c}$ do rovnice získame

$$p_0^{\text{part}}(m, d) = \tilde{c} = \frac{\mu}{\lambda + \mu}p(m).$$

Preto

$$p_0(m, d) = ce^{-(\lambda+\mu)d} + \frac{\mu}{\lambda + \mu}p(m).$$

Integračnú konštantu c určíme z podmienok $p_0(m, d) = p_0(m)$, ktorá vyjadruje skutočnosť, že pri modeli s oneskorením sa pre nulovú hodnotu oneskorenia redukuje na pôvodný IAP model.

$$p_0(m; s, s) = p_0(m, d = 0) = p_0(m),$$

$$p_0(m, 0) = ce^{-(\lambda+\mu) \cdot 0} + \frac{\mu}{\lambda + \mu}p(m),$$

$$c = p_0(m) - \frac{\mu}{\lambda + \mu}p(m).$$

Po dopočítaní integračnej konštanty c sa dostávame ku konečnému tvaru $p_0(m; t, s)$, ktoré bude vyzeráť nasledovne:

$$p_0(m; t, s) = \frac{\mu}{\lambda + \mu}p(m) + \left(p_0(m) - \frac{\mu}{\lambda + \mu}p(m)\right)e^{-(\lambda+\mu)(t-s)}, \quad (3.34)$$

$$p_0(m; t, s) = p_0(m)e^{-(\lambda+\mu)(t-s)} + (1 - e^{-(\lambda+\mu)(t-s)})\frac{\mu}{\lambda + \mu}p(m). \quad (3.35)$$

Obdobným spôsobom by sme sa dostali k riešeniu diferenciálnej rovnice pre $p_1(m; t, s)$

$$\frac{\partial}{\partial t}p_1(m; t, s) = -(\lambda + \mu)p_1(m; t, s) + \lambda p(m),$$

ktoré vyzerá nasledovne:

$$p_1(m; t, s) = p_1(m)e^{-(\lambda+\mu)(t-s)} + (1 - e^{-(\lambda+\mu)(t-s)})\frac{\lambda}{\lambda + \mu}p(m). \quad (3.36)$$

4 Analýza citlivosti modelov

Hodnoty oneskorenia a mier, ktoré používame v našich modeloch, boli už experimentálne odhadované v prácach [15, 16]. Výsledky týchto prác ukázali veľkú diverzifikáciu odhadovaných mier v závislosti nielen od druhu organizmu, ale aj druhu sledovaných bielkovín. Rozdielnosť sa vzťahovala nielen na konštantné miery, ale aj na hodnotu oneskorenia d , v extrémnych prípadoch bolo oneskorenie odhadované na 16 hodín od začatia transkripcie. Preto sa pri analýze kvantitatívnych a kvalitatívnych vlastností modelov zameriame radšej na všetky možné prípady, ktoré môžu nastať. Po vykonaní analýzy si na ilustráciu ukážeme pravdepodobnostné rozloženie pochádzajúce z reálne namerných dát.

4.1 Určenie časovej jednotky

Predtým, ako začneme rozmýšľať nad určením konštantných mier v modeli, sa musíme zamyslieť nad samotným výberom jednotky času. Rovnice (3.19), (3.20) nám udávajú stacionárne pravdepodobnostné rozloženie pre pôvodný IAP model

$$p_{0,n} = \frac{1}{n!} \cdot \frac{\left(\frac{\lambda}{\delta}\right)_n}{\left(\frac{\lambda + \mu + \delta}{\delta}\right)_n} \cdot \left(\frac{\nu}{\delta}\right)^n \cdot {}_1F_1\left(\frac{\lambda}{\delta} + n, \frac{\lambda + \mu + \delta}{\delta} + n, -\frac{\nu}{\delta}\right),$$

$$p_{1,n} = \frac{1}{n!} \cdot \frac{\left(\frac{\lambda}{\delta} + 1\right)_n}{\left(\frac{\lambda + \mu + \delta}{\delta}\right)_n} \cdot \left(\frac{\nu}{\delta}\right)^n \cdot {}_1F_1\left(\frac{\lambda}{\delta} + n + 1, \frac{\lambda + \mu + \delta}{\delta} + n, -\frac{\nu}{\delta}\right).$$

Stacionárne pravdepodobnostné riešenie modelu s oneskorením môžeme vyčítať z rovníc (3.35) a (3.36)

$$p_0(m; t, s) = p_0(m) e^{-(\lambda + \mu)(t-s)} + (1 - e^{-(\lambda + \mu)(t-s)}) \frac{\mu}{\lambda + \mu} p(m),$$

$$p_1(m; t, s) = p_1(m) e^{-(\lambda + \mu)(t-s)} + (1 - e^{-(\lambda + \mu)(t-s)}) \frac{\lambda}{\lambda + \mu} p(m).$$

Treba si uvedomiť, že stacionárne rozloženia sú nezávislé od času. Preto zmena voľby časovej jednotky nám neovplyvní konečný tvar rozloženia. Voľba časovej jednotky v sekundách, minútach alebo hodinách nám zbytočne komplikuje interpretáciu výsledkov a nepopisuje priamo relatívne vzťahy mierových konštánt.

Prirodzenejšie a transparentnejšie bude, ak budeme merať čas v jednotkách δ^{-1} , t.j. ako časovú jednotku zoberieme životnosť bielkoviny. Táto voľba je odôvodnená aj faktom, že sa δ nachádza v riešení hlavne v menovateli, z čoho vyplýva, že výsledný

tvar rozloženia bude určený pomerom hodnôt k δ . Budeme sledovať vplyv honôt $\lambda/\delta, \mu/\delta$ a ν/δ na stacionárnu distribúciu.

4.2 Relatívny vzťah mierových konštánt

Rovnovážne rozloženie pravdepodobností medzi aktívnym a neaktívnym stavom IAP procesu sme definovali ako:

$$p_0 = \frac{\mu}{\lambda + \mu},$$

$$p_1 = \frac{\lambda}{\lambda + \mu}.$$

Z toho nám vyplýva, že čím je väčšia mierová konštanta λ v pomere k μ , tým viac gén zotrúva v aktívnom stave ako v neaktívnom. Produkcia génov súvisí priamo s aktivitou. Ak gén ostane pomerne dlhšie aktívny, tak sa zvýši počet vyprodukovaných bielkovín a s tým sa zvýši pravdepodobnosť výskytu väčšieho počtu bielkovín. Pokiaľ by bola μ , miera prechodu z neaktívneho stavu do aktívneho vyššia ako λ , tak by sa zvýšila pravdepodobnosť výskytu menšieho počtu bielkovín.

Spomínali sme vyšší a nižší počet bielkovín v géne, otázne je, aký maximálny počet bielkovín môžeme očakávať. Na objasnenie využijeme vzťah medzi mierami vzniku a zániku ν a δ . Za predpokladu, že gén je stále v aktívnom stave, nám ν/δ predstavuje priemernú hodnotu počtu bielkovín. Nakoľko sme si ako časovú jednotku zvolili δ , tak nás bude zaujímať hlavne relatívna hodnota ν k δ . Zvýšením tejto hodnoty sa nám nezmení samotný tvar distribúcie, ale zmení sa nám jej rozpätie. To znamená, že ak by sme si zvolili ν rovné 100δ , tak môžeme očakávať, že pravdepodobnosť výskytu 100 molekúl bielkovín v aktívnom géne bude nenulová. Pokiaľ by sme si tento pomer zvolili $\nu/\delta = 50$, tak nemôžeme očakávať dosiahnuteľnosť takýchto výsledkov a pravdepodobnosť bude nulová.

4.3 Prototypy simulácia

Za jednotku času sme si zvolili δ . Na to, aby sme lepšie mohli preskúmať vlastnosti rozdelenia pravdepodobností, budeme sa snažiť uskutočňovať tieto pozorovania na širšom intervale dosiahnuteľného počtu bielkovín. Ako sme vyššie spomínali, toto docielime vhodným zvolením pomeru ν/δ . Pre našu štúdiu položíme $\nu = 100\delta$.

Pri výpočtoch stacionárneho pravdepodobnostného rozdelenia sme využili programové prostredie MATLAB. Na overenie správnosti výsledkov a prípadnú kalibráciu programu sme použili také isté parametre IAP procesu, ako boli skúmané v práci[14]. Výsledné rozdelenia, ktoré sme porovnávali, je možné vidieť v prílohe na obrázku 2.

Grafy na obrázku 2 nám opisujú pravdepodobnosť výskytu m molekúl v aktívnom ($p_1(m)$) a v neaktívnom géne ($p_0(m)$). Celková pravdepodobnosť výskytu m bielkovín je určená pravdepodobnosťou $p(m)$.

Zamerali sme sa na 5 majoritných prípadov, ktoré mohli nastať, a to $\lambda, \mu \leq \delta$ a $\lambda = \mu = \delta$. Názvy jednotlivých grafov nám udávajú použité hodnoty pre (λ, μ) .

V prípade $\lambda > \delta$ má pravdepodobnostné rozdelenie v okolí nuly stúpajúcu tendenciu. Je to spôsobené tým, že počas aktivity génu nedôjde k zániku novovytvorených proteínov tejto aktivity a bielkoviny ostanú v géne prítomné aj počas neaktivity. Pokiaľ je ešte gén pomerne viac v aktívnom stave ako neaktívnom ($\lambda > \mu$), tak sa celé rozdelenie posunie smerom k strednej hodnote pravdepodobnostného rozdelenia počtu bielkovín stále aktívneho génu ($\nu/\delta = 100$). Niektoré bielkoviny sa dožijú aj ďalšieho prechodu z neaktívneho do aktívneho stavu.

Pokiaľ sa miery rovnajú $\lambda = \mu = \delta$, tak celkové pravdepodobnostné rozloženie je vyrovnané v okolí hodnoty 0,01. Pravdepodobnosť výskytu m bielkovín začne klesať až v okolí $m = 90$. Takýto výsledok sa dal očakávať, nakoľko pravdepodobnosť výskytu v aktívnom alebo neaktívnom stave je rovnaká a priemerná dĺžka zotrvania génu v jednom zo stavov je taká istá ako priemerná živostnosť jednej bielkoviny. To znamená, že za jednu periódu aktivity a neaktivity génu dôjde k vzniku a zániku rovnakého počtu bielkovín.

Posledné 2 grafy obrázku 2 nám znázorňujú situáciu $\lambda < \delta$. Celkové pravdepodobnostné rozdelenie nadobúda maximum v okolí nuly. Životnosť bielkoviny oproti aktivite génu je nižšia. Už počas aktívneho stavu dochádza k degradácii bielkovín. Pravdepodobnosť, že sa nejaké bielkoviny dožijú ďalšieho prechodu z neaktívneho na aktívny gén, je veľmi nízka. Táto pravdepodobnosť sa ešte viac zníži, ak gén zotráva pomerne dlhšie v neaktívnom stave ako aktívnom ($\mu > \lambda$). Vzhľadom na postavenie IAP modelu výsledky, ktoré sme dostali, sa dali očakávať. Simulované stacionárne pravdepodobnostné rozdelenia preukazujú veľký súvis medzi počtom bielkovín a aktivitou génu. Toto protirečí so závermi štúdií [12].

4.4 Simulácie s oneskorením

Aby sme sa vyhli korelácii počtu bielkovín s aktivitou, resp. neaktivitou génu, zakomponovali sme do pôvodného IAP modelu prvok oneskorenia. Ako pri predošlom modeli, pozrieme sa na všetky možné prípady relatívnych vzťahov mierových konštánt, ktoré môžu nastať. Budeme manipulovať s relatívnou hodnotou d k δ .

Výsledky našich simulácií si môžeme pozrieť v prílohe, jedná sa o obrázky 3 až 7. Názvy obrázkov nám predstavujú zvolené hodnoty pre miery (λ, μ) . Ako jednotku

času sme ponechali δ a platí $\mu = 100\delta$. Hodnoty oneskorenia sme skúmali v rozpätí 0.1 až 2, konkrétne ako d sme si zvolili postupne hodnoty 0.1, 0.3, 0.6, 1, 1.5 a 2.

Všetky simulácie majú niekoľko spoločných vlastností. Vychádzajúc z rovníc (3.35) a (3.36) môžeme vidieť, že celkové pravdepodobnostné rozloženie $p(m) = p_0(m) + p_1(m)$ sa nemení. Tento fakt súvisí aj s experimentálnymi poznatkami, z ktorých vieme, že posunutie času by nemalo celkové rozdelenie ovplyvňovať. Oneskorenie spôsobí len to, že veci sa dejú o d jednotiek času neskôr. Ako vidieť, podarilo sa nám znížiť závislosť pravdepodobnostného rozloženia od aktivity génu. S dostatočne veľkou hodnotou posunutia sa tvary pravdepodobnostných rozdelení skoro rovnajú. V prípade, keď $\lambda = \mu$, tak po zahrnutí posunutia sú pravdepodobnostné rozloženia rovnaké nielen tvarovo, ale dosahujú aj rovnaké hodnoty. Pokiaľ $\lambda > \mu$ tak pri dostatočne veľkej hodnote posunutia dosahuje pravdepodobnostné rozloženie aktívneho génu väčšia hodnoty ako v prípade, keď gén je neaktívny. Je to logické, pokiaľ sú rozdelenia tvarovo rovnaké, tak proporcionálne rozdelenie času génu stráveného buď v aktívnom alebo neaktívnom stave spôsobí zmenu v dosahovaných hodnotách, nakoľko celkové rozdelenie sa nemení.

Ďalej si môžeme povšimnúť neúmernú zmenu rozloženia voči veľkosti odchýlky. Rozdiel medzi pôvodným modelom a modelom s posunutím $d = 0.1$ je oveľa markantnejší ako v prípade rozdelenia pravdepodobností s posunom $d = 1$ a $d = 2$. Z toho nám vyplýva, že s malou zmenou pôvodného modelu dostávame model s lepšími kvalitatívnymi výsledkami.

Na základe rozboru modelu a poznatkov, z ktorých sme vychádzali, môžeme byť z výsledkami spokojný. Model nám poskytuje kvantitatívnu výpoveď o vzájomných vzťahoch jednotlivých mier a informáciu o možných výsledkoch v závislosti od relatívnych hodnôt. Výsledné rozdelenie sa nemení, je závislé len od konštantných mier a neovplyvní sa zmenou posunutia. Je stále zachovaná priama súvislosť aktivity génu a syntézy bielkovín, len sa prejaví o d jednotiek času neskôr.

4.5 Testovanie na reálnych dátach

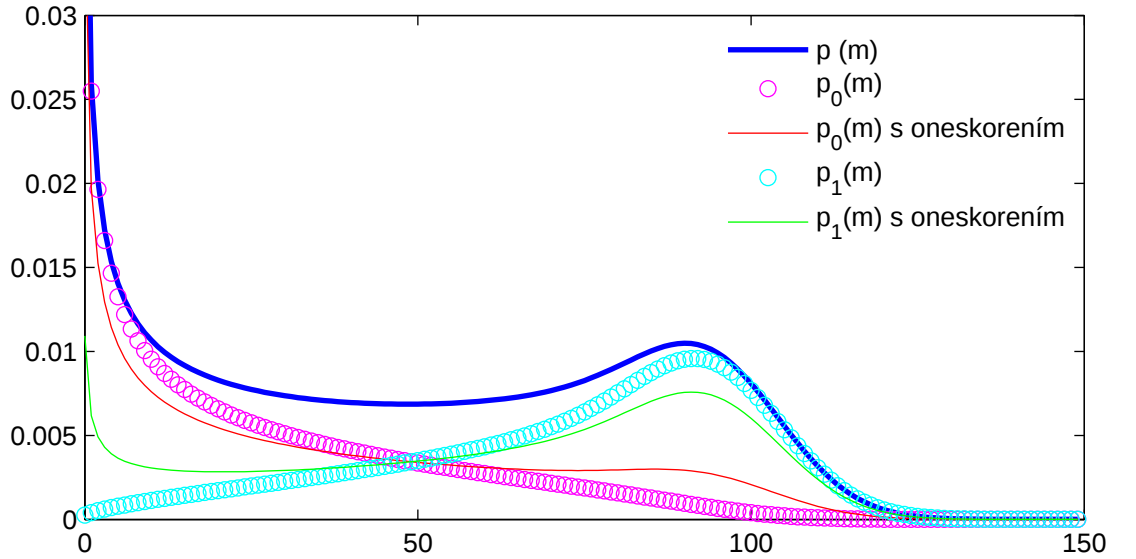
Analyzovali sme všetky prípadné vzájomné vzťahy mierových konštánt. Odstránili sme problém vysokej korelácie zavedením posunutia do modelu. Nakoľko model je postavený na základe experimentálnych výsledkov, zaujíma nás, ako sa bude správať v prípade nameraných mierových konštánt. Bohužiaľ, nestretli sme sa s odhadmi pre všetky parametre, preto niektoré zvolíme na základe predošlej analýzy.

V práci [15] sme sa stretli s odhadmi pre *Hes1* bielkoviny eukaryotických buniek.

Odhadovaná hodnota miery pre degradácie bielkovín je $\delta = 0.03/min$ a oneskorenie $d = 18.5 min$. V tejto práci okrem odhadov bola sledovaná oscilácia počtu mRNA a počtu bielkovín. Pre bielkoviny Hes1 bola nameraná oscilačná perióda 2 hodiny. Počet bielkovín v priebehu jednej hodiny narástol a dosiahol svoje maximum, potom začal klesať. Na základe týchto výsledkov môžeme predpokladať, že gén bol aktívny, resp. neaktívny priemerne jednu hodinu. Ako odhad mierových konštánt použijeme $\lambda = \mu = 1/60 min \doteq 0.017 min$.

Ako jednotku času si zoberieme $\delta^{-1} \doteq 33 min$. Bude nás zaujímať hodnota mierových konštánt za novozvolenú jednotku času. Chýba nám ešte určiť mieru vzniku bielkovín ν . Po analýze modelov sme došli k záveru, že táto hodnota len rozšíri pozorovateľný interval molekúl bielkovín. Kvôli prehľadnosti si za mierovú konštantu dosadíme $\mu = 100\delta$. Upravené miery dosahujú hodnoty:

$$\delta = 1, \quad \lambda = 0.55, \quad \mu = 0.55, \quad \nu = 100.$$



Obr. 1: Realistické hodnoty mierových konštánt

Podobný prípad sme už rozoberali. Jedná sa o prípad, keď μ a λ sú menšie ako δ . Priemerná dĺžka aktivity, resp. neaktivity génu je dlhšia ako priemerná životnosť bielkovín. Niektoré novosyntetizované bielkoviny sa začnú degradovať ešte počas aktivity génu a nedožijú sa ďalšej aktivity génu.

Model vykazuje relevantné výsledky aj s použitím experimentálnych dát. Zavedením oneskorenia sa nám podarilo vylepšiť vlastnosti aj tejto simulácie. Je vidieť, že vzájomná korelácia medzi aktivitou a neaktivitou sa znížila.

Záver

Na základe nedávnych výsledkov z oblasti experimentálnych technológií a s nimi spolusúvisiacimi štúdiami sme sa rozhodli upraviť už známy a využívaný model génovej expresie. Účelom tejto úpravy bolo priblížiť výsledné vlastnosti modelu k pozorovaným experimentálnym vlastnostiam.

Na začiatku sme obohatili svoje vedomosti o znalosti z oblasti génovej expresie, ako je transkripcia a translácia, a o znalosti aktuálnych experimentálnych výsledkov. Potrebovali sme sa oboznámiť lepšie s problematikou, aby sme mohli navrhnúť relevantné riešenie na zlepšenie kvalitatívnych vlastností modelu.

Na popísanie modelu sme použili stochastickú teóriu Markovových procesov, konkrétne sme využívali procesy so spojitým časom, ako je proces vzniku a zániku a Poissonov proces. Po preskúmaní experimentálnych výsledkov sme ako jeden parameter modelu zvolili aktivitu génu. Aktivita génu nám popisuje dynamiku počtu mRNA v géne, ktoré majú relatívne krátku životnosť a k ich vzniku dochádza nárazovo, v určitých časových intervaloch so zvýšenou intenzitou mRNA. Tento stav sme označili ako aktívny gén a iba v tomto stave sme uvažovali možnú transláciu bielkovín. Gén, ktorý má nízku transkripciu mRNA, sme nazvali neaktívnym. Do nášho základného modelu IAP procesu vstupovali parametre popisujúce miery prechodu medzi aktívnym a neaktívnym stavom, miery vzniku a zániku bielkovín. Úspešne sme vyjadrili explicitné riešenie pre stacionárne rozloženie aktívneho a neaktívneho génu vzhľadom na počet bielkovín prítomných v géne.

Pred zakompovaním oneskorenia do modelu sme najskôr odôvodnili správnosť našej voľby a ukázali sme, že oneskorenie nebude mať vplyv na potrebnú homogenitu a vzájomnú nezávislosť výskytu udalostí, ako je vznik a zánik proteínov a prechod medzi stavmi aktivity. Oneskorenie spôsobí len to, že sa translácia dokončuje o d jednotiek neskôr od jej iniciácie. Stacionárne pravdepodobnostné rozdelenie sme vyjadrili pomocou rozloženia samotného IAP procesu a veľkosti oneskorenia, lebo pokiaľ si za d zvolíme hodnotu nula, tak musíme dostať tie isté výsledky, ako sú v IAP procese. Z výsledkov sme mohli vyčítať, že čím vyššiu zvolíme hodnotu oneskorenia, tým viac sa začne tvarom naše výsledne rozdelenie podobáť celkovému rozdeleniu pravdepodobnosti výskytu m molekúl bielkovín v géne.

V štvrtej kapitole sme sa zamerali na analýzu vlastností oboch modelov, aj s oneskorením, aj bez oneskorenia. Po zvážení možností sme ako časovú jednotku zvolili priemerný život jednej molekuly bielkoviny. Takto sa nám podarilo poskytnúť lepšiu kvantitatívnu analýzu modelu už len na základe vstupných mierových konštánt a vzťahov medzi jednotlivými mierami. Zistili sme, že hodnota mierovej konštanty

vzniku bielkovín ν spôsobí len zmeny v šírke intervalu dosiahnuteľného počtu bielkovín. Na tvar majú vplyv vzťahy medzi mierami prechodu z aktívneho a neaktívneho stavu λ a μ a mierou degradácie bielkovín δ . Prešli sme si päť prototypov simulácií, ktoré mohli nastať a analyzovali sme ich výsledky. Podarilo sa nám docieľiť, čo sme chceli, a to je zníženie vzájomného vplyvu medzi počtom bielkovín a samotnou aktivitou génu. Ostatné pozorované vlastnosti taktiež súhlasili s prezentovanou teóriou.

Jeden z problémov momentálnej teórie génovej expresie je, že zatiaľ chýbajú ucelené odhady pre niektoré parametre modelu a rozdiely medzi jednotlivými organizmami a bielkovinami sú priveľké. Preto sme zobrali odhady niektorých mier pre konkrétny typ bielkoviny a zvyšné sme určili na základe pozorovaných vlastností zo zdrojovej štúdie a našich prototypov. Aj v tomto prípade sa nám podarilo znížiť vplyv aktivity na pravdepodobosť výskytu bielkovín.

S celkovými výsledkami tejto práce môžeme byť spokojní. Okrem toho, že sme dosiahli náš prvotný cieľ zníženia korelácie medzi aktivitou a neaktivitou génu, tak sa nám podarilo na základe relatívnych vzťahov vstupných parametrov podať kvantitatívnu analýzu. Z kvalitatívneho hľadiska modelu sa zavedením malého posunutia d podarilo docieľiť pomerne lepšie výsledky. Ostatné vlastnosti našich simulácií tiež súhlasili s teóriou, z ktorej sme vychádzali.

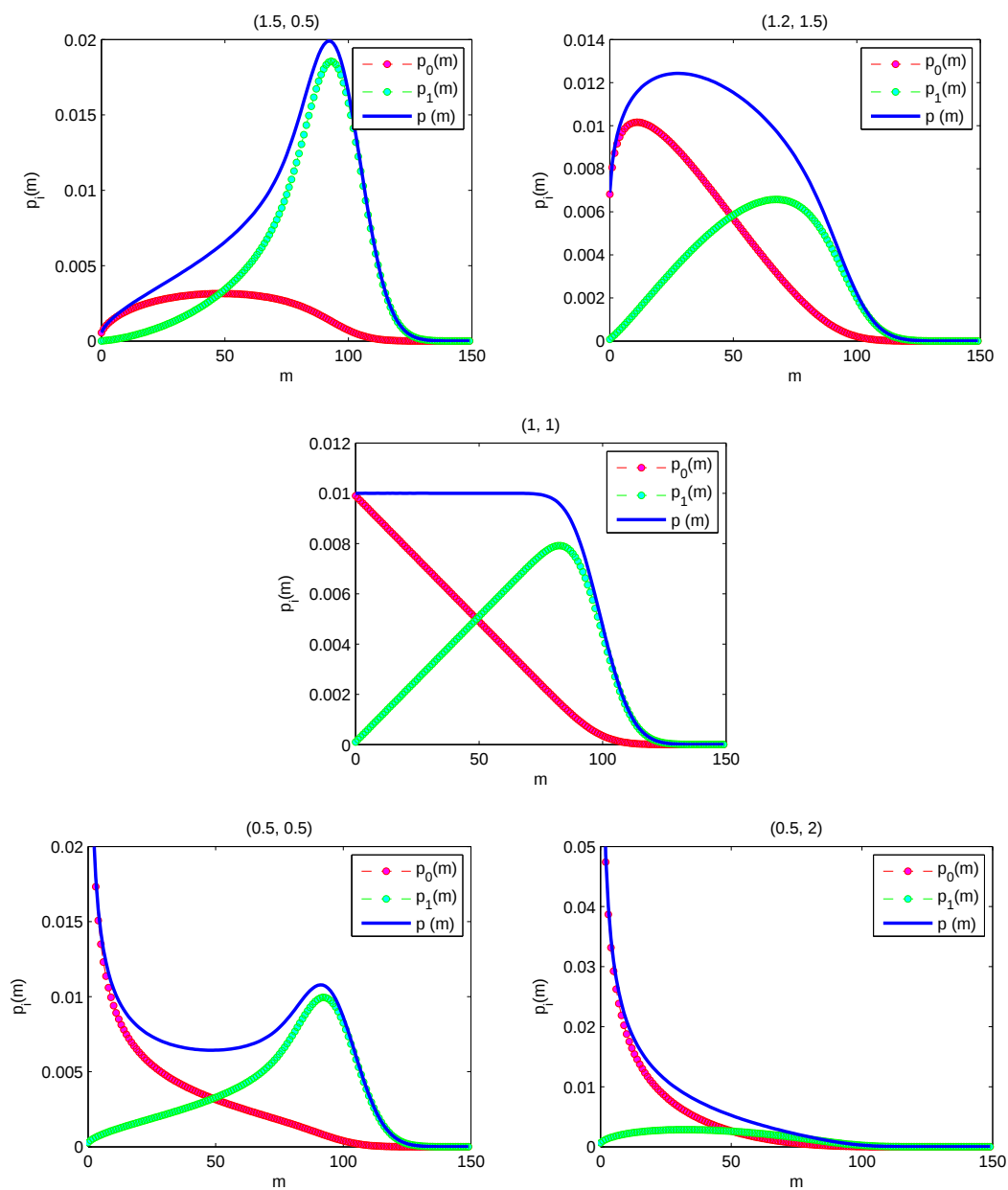
Samozrejme sa jedná len o jednoduchý model, ktorý má možnosť expanzie, ale je vidieť, že v tejto oblasti je ešte veľký priestor pre napredovanie. Aký potenciál prispieť spoločnosti má ešte objav v tejto oblasti, keď zatiaľ nie sme schopní modelovať základný proces na molekulovej úrovni, ktorý ukazuje veľké rozdiely medzi samotnými bunkami? Čo sa vlastne skrýva za obzorom našich momentálnych vedomostí?

Literatúra

- [1] Brunovský P., (2011), *Diferenčné a diferenciálne rovnice*, internetová adresa <http://www.iam.fmph.uniba.sk/skripta/brunovsky/> .
- [2] Peccoud J., Ycart B., (1995), *Markovian modeling of gene-product synthesis*, Theoretical population biology 48:222-34.
- [3] Bokes P., Gedeon T., (2012), *Delayed protein synthesis reduces the correlation between mRNA and protein fluctuations*, Biophys. J. 103:1087-1096.
- [4] Kalas J., (1993), *Markovove reťazce*, Univerzita Komenského, Skriptá.
- [5] Norris J., (1997), *Markov Chain*, Cambridge University Press.
- [6] Raj A., Peskin A., Tranchina D., Vargas D., Tyagi S., (2006), *Stochastic mRNA Synthesis in Mammalian Cells*, PLoS Biology 4:e309.
- [7] Becskei A., Kaufmann B., van Oudenaarden A., (2005), *Contributions of low molecule number and chromosomal positioning to stochastic gene expression*, Nat Genet 37:937-944.
- [8] Raser J., O'Shea K., (2004), *Control of stochasticity in eukaryotic gene expression*, Science 304:1811-1814.
- [9] Bengtsson M., Stahlberg A., Rorsan P., Kubista M., (2005), *Gene expression profiling in single cells from pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels*, Genome Res 15:1388-1392.
- [10] Norris J., Stirland A., McFerran W., Seymour Z., Spiller G., et al. (2003), *Dynamic patterns of growth hormone gene transcription in individual living pituitary cells*, Mol Endocrinol 17:193-202.
- [11] Walters C., Fiering S., Eidemiller J., Margis W., Groudine m., et al., (1995) *Enhancers increase the probability but not level of gene expression*, Proc Natl Acad Sci U S A 92:7125-7129.
- [12] Taniguchi Y., Choi P., Li G., Chen H., Babu M., Hearn J., Emili A, Xie X., (2010), *Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells*, Science 329:533-8.
- [13] Newman J., Ghaemmaghami S., Ihmels J., Breslow D., Noble M., DeRisi J., Weissman J., (2006), *Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise*, Nature 441:840-6.

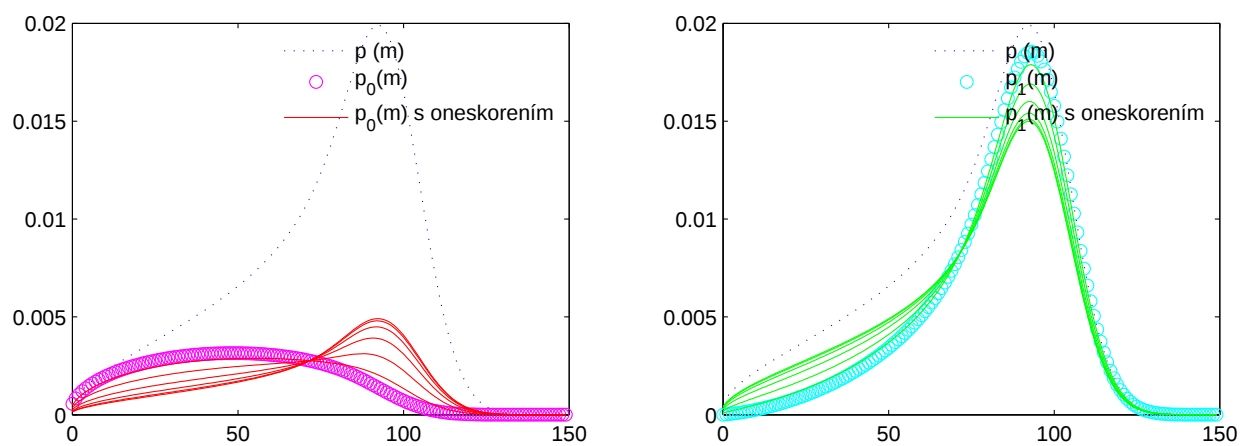
- [14] Iyer-Biswas S., Hayot, F., Jayaprakash C., (2009), *Stochasticity of gene products from transcriptional pulsing*, Phys. Rev. E 79:031911.
- [15] Monk N., (2003), *Oscillatory Expression of Hes1, p53, and NF- κ B Driven by Transcriptional Delays*, Current Biology 13:1409-1413.
- [16] Wang Y., Liu Ch., Storey J., Tibshirani R., Herschlag D., Brown P., (2002), *Precision and functional specificity in mRNA decay*, PNAS 99:5860-5865.

A Príloha

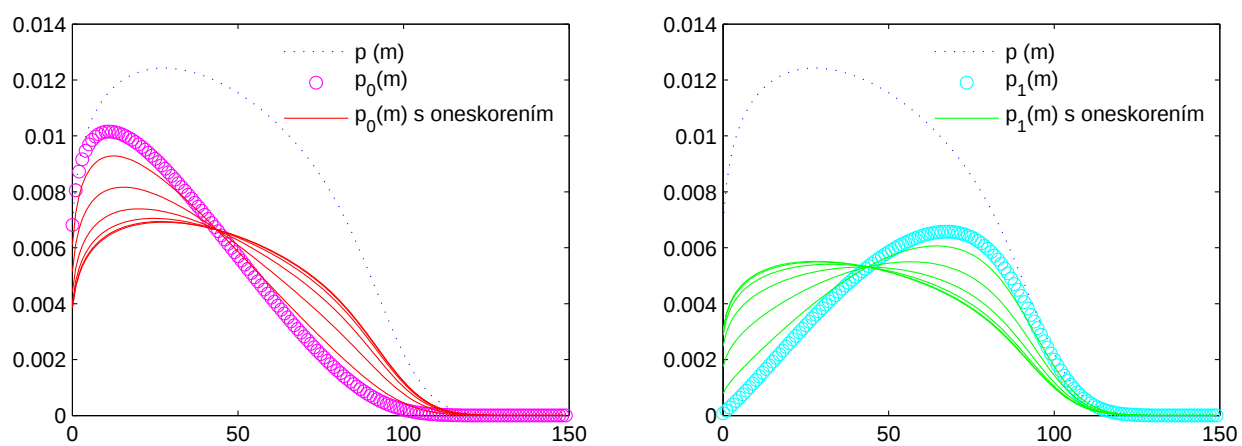


Obr. 2: Stacionárne pravdepodobnostné rozloženie bielkovín génu vzhľadom na aktivitu. Názvy jednotlivých grafov vyjadrujú hodnoty (λ, μ) v jednotkách δ . Miera syntetizácie bielkovín ν má hodnotu 100δ .

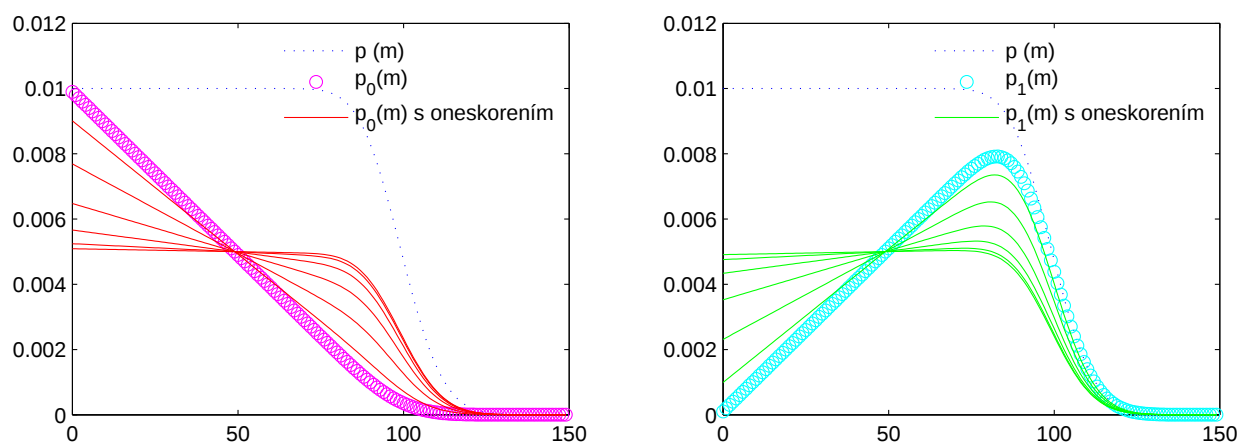
Obr. 3: (1.5, 0.5)



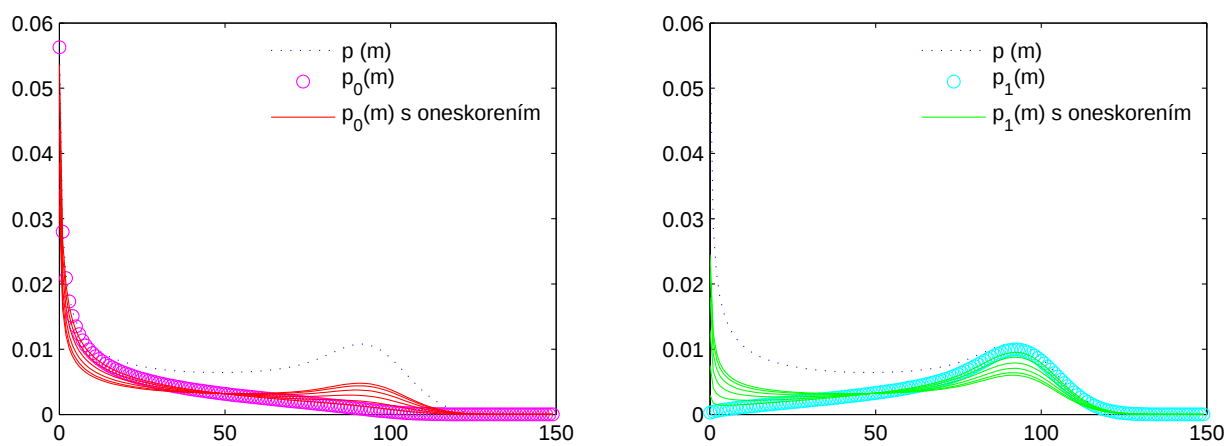
Obr. 4: (1.2, 1.5)



Obr. 5: (1, 1)



Obr. 6: (0.5, 0.5)



Obr. 7: (0.5, 2)

