MATEMATICKO-FYZIKÁLNÍ FAKULTA
Univerzita Karlova

BAKALÁŘSKÁ PRÁCE

Šimon Kučera

Větvící procesy a jejich aplikace v biologii

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: doc. RNDr. Daniel Hlubinka, Ph.D.
Studijní program: Bakalářský
Studijní obor: Finanční matematika

Praha 2025

Prohlašuji, že jsem tuto bakalářskou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů. Tato práce nebyla využita k získání jiného nebo stejného titulu.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V ..... dne ....

Podpis autora

Rád bych poděkoval doc. RNDr. Danielu Hlubinkovi, Ph.D. za vedení práce a cenné rady.

Název práce: Větvící procesy a jejich aplikace v biologii

Autor: Šimon Kučera

katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: doc. RNDr. Daniel Hlubinka, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Tato bakalářská práce se zabývá základním typem větvících procesů, konkrétně Galtonovými-Watsonovými (G.-W.) procesy, a jejich aplikacemi v biologii. Cílem práce je srozumitelně představit koncept G.-W. procesů, formulovat a řešit typické problémy s nimi spojené a ilustrovat teoretické výsledky pomocí simulací. Práce definuje základní pojmy a podrobně se věnuje problému vyhynutí procesu a ukazuje, jak střední hodnota počtu potomků (subkritické, kritické a superkritické procesy) ovlivňuje dlouhodobé chování populace. V závěrečné části jsou diskutovány asymptotické vlastnosti procesů. Pro subkritické a kritické procesy je zkoumána podmíněná střední hodnota za předpokladu, že proces ještě nevyhynul. Pro superkritické procesy je definován normalizovaný proces a zmíněna jeho konvergence. Teoretické poznatky jsou průběžně ověřovány pomocí simulací, které potvrzují platnost odvozených vzorců.

Klíčová slova: Větvící se procesy Galtonovy-Watsonovy procesy

Title: Branching processes and their applications in biology

Author: Šimon Kučera

department: Department of Probability and Mathematical Statistics

Supervisor: doc. RNDr. Daniel Hlubinka, Ph.D., Department of Probability and Mathematical Statistics

Abstract: This bachelor's thesis focuses on a fundamental type of branching processes, specifically Galton–Watson (G.-W.) processes, and their applications in biology. The aim of the thesis is to clearly present the concept of G.-W. processes, formulate and solve typical problems associated with them, and illustrate theoretical results through simulations. The thesis defines the basic concepts and examines the extinction problem in detail and demonstrates how the mean number of offspring (in subcritical, critical, and supercritical processes) influences the long-term behavior of a population. In the final part, the asymptotic properties of the processes are discussed. For subcritical and critical processes, the conditional mean value given survival is explored, while for supercritical processes, a normalized process is defined and its convergence is addressed. Theoretical results are continuously verified through simulations, which confirm the validity of the derived formulas.

Keywords: Branching processes Galton-Watson processes

Obsah

Úvod **2**

1 Galtonovy-Watsonovy procesy **3**
1.1 Momenty 4
1.2 Vyhynutí 8
1.2.1 Integrovaný Galtonův-Watsonův proces 10
1.3 Geometrické větvení 12

2 Asymptotické vlastnosti **15**
2.1 Subkritické procesy 15
2.2 Kritické procesy 16
2.3 Superkritické procesy 18

Závěr **21**

Seznam použité literatury **22**

Seznam obrázků **23**

Seznam tabulek **24**

Seznam použitých zkratek **25**

A Přílohy **26**
A.1 První příloha 26

Úvod

Větvící procesy jsou stochastické modely popisující vývoj populací, jejichž částice na konci svého života, který je náhodně dlouhý, vyprodukují náhodný počet potomků, řídící se předepsaným pravděpodobnostním rozdělením. Tyto částice mohou představovat rakovinné buňky, bakterie nebo dokonce mužské potomky vznešeného rodu, u kterého můžeme například počítat pravděpodobnost, že vymře po meči.

V této práci se zaměříme na základní typ větvících procesů, známý jako Galtonovy-Watsonovy procesy, a prozkoumáme jejich klíčové vlastnosti a určité jejich modifikace. Představíme matematické vzorce potřebné k jejich popisu, včetně pravděpodobnostních vytvořujících funkcí, které jsou zásadním nástrojem pro popis těchto procesů.

Cílem této práce není podávat důkazy všech prezentovaných vět týkajících se větvících procesů. Hlavním záměrem je srozumitelně představit koncept G.-W. procesů a formulovat a vyřešit několik typických problémů s nimi spojených. Důraz je kladen také na praktickou ilustraci teoretických výsledků a ověření platnosti odvozených vzorců pomocí simulací. V práci budou popsány základní vlastnosti a typy Galtonových - Watsonových procesů, včetně procesů subkritických, kritických a superkritických.

1. Galtonovy-Watsonovy procesy

Jedním z nejjednodušších modelů zrodu a zániku je Galtonovův-Watsonův proces. Jeho jednoduchost spočívá v tom, že každá buňka žije právě jednu jednotku času a ve chvíli svého zániku vytvoří náhodný počet potomků, který se řídí předepsaným pravděpodobnostním rozdělením. Z toho důvodu můžeme zapsat čas jako diskrétní proměnnou, což bude zároveň i číslo generace procesu. Takové nastavení v reálném světě může odpovídat například množení jepic. Tato kapitola bere za základ kapitolu 3 z Kimmel a Axelrod (2015).

Jeden ze způsobů, jak popsat Galtonův-Watsonův proces, je pomocí zpětné rovnice. Jejím základním principem je, že jsme si schopni rozdělit proces na jednotlivé podprocesy, jejichž novou počáteční buňkou bude jeden z potomků první generace. Celkový proces pak bude sjednocením těchto podprocesů, což lze vidět na 1.1.

Nejdříve si vytvoříme přehlednou tabulku značení, které se budou v práci opakovaně objevovat.

Proměnná	**Definice**
Z_n	Počet jedinců v generaci n
$Z_{1,n+1}^{(j)}$	Počet jedinců v generaci $n+1$, kteří jsou potomky buňky j z první generace
$f(s)$	Pravděpodobnostní vytvořující funkce rozdělení počtu potomků jednoho jedince
$f_n(s)$	Pravděpodobnostní vytvořující funkce generace n
m	Střední hodnota počtu potomků jednoho jedince
e	Pravděpodobnost vyhynutí procesu
e^*	Kořen rovnice $f(s) = s$, který není roven 1

Tabulka 1.1: Tabulka nejpoužívanějších proměnných

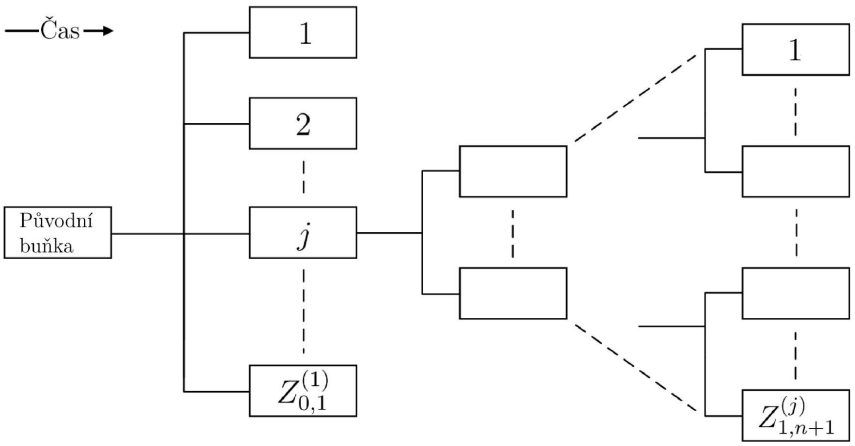
Definovali jsme si Z_n jako počet buněk v n-té generaci a $Z_{1,n+1}^{(j)}$ jako počet buněk v $n+1$ generaci, které vzešly z j-té počáteční buňky první generace. Celkový počet buněk v $n+1$ generaci lze tedy vyjádřit následovně

$$Z_{n+1} = \begin{cases} Z_{1,n+1}^{(1)} + \cdots + Z_{1,n+1}^{(Z_1)}, & \text{je-li } Z_1 > 0, \\ 0, & \text{je-li } Z_1 = 0, \end{cases}$$

nebo alternativně

$$Z_{n+1} = \sum_{j=1}^{Z_1} Z_{1,n+1}^{(j)}.$$

Ukázku těchto vzorců lze pozorovat na následujícím schématu, kde vidíme rozvoj buňky j z první generace.



Obrázek 1.1: Schéma zpětného vzorce

1.1 Momenty

Ke zjistění momentů těchto procesů použijeme alternativní postup a to pomocí pravděpodobnostních vytvořujích funkcí. Ty jsme si již pro první a n-tou označili v tabulce 1.1 jako $f(s)$ a $f_n(s)$.

Definice 1. *Pravděpodobnostní vytvořující funkci $f(s)$ dikrétní náhodné veličiny Z definujeme jako $f(s) := \mathbb{E}[s^Z] = \sum_i s^i P[Z = i]$, pro $Z \in \{0, 1, \cdots\}$.*

Předpokládejme, že proces vzešel z jediné buňky, tudíž

$$f_0(s) = s$$

a dle Grimmet a Stirzaker (2001) 5.4.1 platí následující

$$f_n(s) = \underbrace{f(f(\cdots f(f(s))}_{n \text{ krát}} \cdots)), \tag{1.1}$$

$$f_{n+1}(s) = f_1[f_n(s)] = f[f_n(s)]. \tag{1.2}$$

Jak by ale vypadala pravděpodobnostní vytvořující funkce v n-té generaci, pokud by proces začínal s náhodným počtem buněk, daným pomocí pravděpodobnostní vytvořující funkce $g(s)$? Ta bude nyní vystupovat místo $f_0(s) = s$. Zavedeme si $h_n(s)$ jako pravděpodobnostní funkci generace n tohoto procesu. Poté bude platit podle vzorce 1.2

$$h_n(s) = \underbrace{f(f(\cdots f(g(s))}_{n \text{ krát}} \cdots)).$$

Jelikož již známe pravděpodobnostní vytvořující funkce, můžeme využít vlastnosti, že jejich derivace v bodě 1 zleva udává jejich momenty. Definovali jsme m jako průměrný počet potomků jedné buňky,

$$m = \mathbb{E}(Z_1) = f'(1-)$$

$$\mathbb{E}(Z_n) = f_n'(1-) = f_{n-1}'(1-)f'(1-) = \cdots = m^n,$$

kde $f'(1-)$ značí derivaci funkce f v bodě 1 zleva. Stejným způsobem můžeme odvodit i rozptyl

$$\mathrm{var}(Z_n) = \begin{cases} \frac{\sigma^2 m^{n-1}(m^n-1)}{m-1}, & \text{je-li } m \neq 1, \\ n\sigma^2, & \text{je-li } m = 1, \end{cases}$$

kde σ^2 značí $\mathrm{var}(Z_1)$. Tento vzorec získáme dvojím derivováním $f_n(s)$ v bodě 1

$$f_n''(1-) = f''(1-)f_{n-1}'(1-) + f'(1-)f_{n-1}''(1-)$$

což poté dosadíme do vzorce pro rozptyl

$$\begin{aligned} \mathrm{var}[Z_n] &= \mathbb{E}[Z_1^2] - \mathbb{E}[Z_1]^2 \\ &= \mathbb{E}[Z_1[Z_1 - 1] + Z_1] - \mathbb{E}[Z_1]^2 \\ &= \mathbb{E}[Z_1[Z_1 - 1]] + \mathbb{E}[Z_1] - \mathbb{E}[Z_1]^2 \\ &= f''(1-) + f'(1-) - f'(1-)^2. \end{aligned}$$

Ukážeme, jakou interpretaci mohou pravděpodobnostní vytvořující funkce mít.

1) $f(s) = ps^2 + qs$ - buňka se rozdělí na dvě s pravděpodobností p a přežije do dalšího kroku s pravděpodobností q.

2) $f(s) = ps^2 + q$ - buňka se rozdělí na dvě s pravděpodobností p a zemře s pravděpodobností q.

3) $f(s) = ps^2 + qs + r$ - buňka se rozdělí na dvě s pravděpodobností p, přežije do dalšího kroku s pravděpodobností q a zemře s pravděpodobností r.

Pokud bychom však iterovali takovéto funkce, rychle bychom se dostali k velice složitým tvarům funkcí. Například kdybychom chtěli popsat teprve třetí generaci třetí možnosti výše, získali bychom tvar

$$\begin{aligned} f_3(s) = &p\left(p\left(s^2p + sq + r\right)^2 + q\left(s^2p + sq + r\right) + r\right)^2 \\ &+ q\left(p\left(s^2p + sq + r\right)^2 + q\left(s^2p + sq + r\right) + r\right) + r \end{aligned}$$

Existují ale i netriviální funkce, pro které lze jejich iterace vypočítat explicitně. Nazveme je lineárně-lomené.

Zavedeme

$$p_0 = \frac{1 - b - p}{1 - p},\ p_k = bp^{k-1}, k \in \mathbb{N}, p \in (0{,}1), b \in (0{,}1),$$

kde p_k značí pravděpodobnost, že buňka bude vyprodukuje k potomků. Pak platí

$$f(s) = 1 - \frac{b}{1 - p} + \frac{bs}{1 - ps},$$

což při volbě parametru b na hodnotu $1 - p$ odpovídá geometrickému rozdělení posunutému v nule. Pro n-tou iteraci platí

$$f_n(s) = \begin{cases} 1 - m^n\left(\frac{1-e^*}{m^n-e^*}\right) + \frac{m^n\left(\frac{1-e^*}{m^n-e^*}\right)^2 s}{1-\left(\frac{m^n-1}{m^n-e^*}\right)s}, & \text{pro } m \neq 1, \\ \frac{np-(np+p-1)s}{1-p+np-nps}, & \text{pro } m = 1, \end{cases}$$

kde

$$m = \frac{b}{(1-p)^2}$$

a e^* je vypočteno jako ten kořen funkce rovnice $f(s) = s$, který není roven 1, což je při $m > 1$ pravděpodobnost vymření procesu. To ovšem probereme podrobněji až v následující kapitole. Dokážeme si případ, kdy je střední hodnota počtu potomků jedné buňky rovna jedné pomocí matematické indukce.

Pro $n = 1$ platí

$$f_1(s) = \frac{p-(p+p-1)s}{1-p+p-ps} = \frac{p+s-2ps}{1-ps} = 1 - \frac{b}{1-p} + \frac{bs}{1-ps}.$$

Přepokládejme, že $f_k(s)$ platí pro nějaké $k \in \mathbb{N}$. Náš indukční krok tedy bude $f_k(s) \implies f_{k+1}(s)$, tím získáme

$$\begin{aligned}
f_{k+1}(s) &= f(f_k(s)) \\
&= f\left(\frac{kp-(kp+p-1)s}{1-p+kp-kps}\right) \\
&= \frac{(1-2p)\left(\frac{kp-(kp+p-1)s}{1-p+kp-kps}\right)+p}{1-p\left(\frac{kp-(kp+p-1)s}{1-p+kp-kps}\right)} \\
&= \frac{p+kp+s-2ps-kps}{1+kp-ps+kps} \\
&= \frac{p(k+1)-s(p+kp+p-1)}{1+kp-ps-kps} \\
&= \frac{p(k+1)-s[(k+1)p+p-1]}{1-p+(k+1)p-(k+1)ps},
\end{aligned}$$

čímž jsme vzorec dokázali. S ním jsme nyní schopni vypočítat rozdělení. Stačí využít vzorec pro součet geometrické řady.

$$\begin{aligned}
f(s) &= \frac{np-(np+p-1)s}{1-p+np-nps} \\
&= \frac{np-(np+p-1)s}{1-p+np} \cdot \frac{1}{1-\frac{nps}{1-p+np}} \\
&= \frac{np-(np+p-1)s}{1-p+np} \cdot \sum_{k=0}^{\infty} \left(\frac{nps}{1-p+np}\right)^k \\
&= \sum_{k=0}^{\infty} \frac{(np-(np+p-1)s)(nps)^k}{(1-p+np)^{k+1}} \\
&= \frac{np}{1-p+np} + \sum_{k=1}^{\infty} \frac{(n^{k-1}p^{k-1}-2n^{k-1}p^k+n^{k-1}p^{k+1})s^k}{(1-p+np)^{k+1}}
\end{aligned}$$

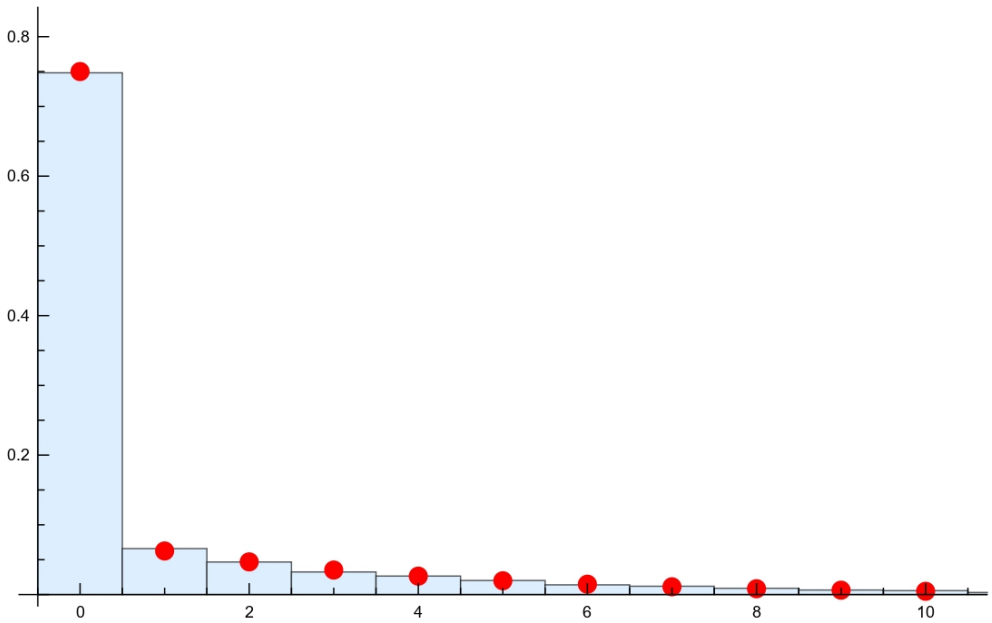
Z této sumy nyní můžeme získat zpět vzorec pro rozdělení počtu potomků v generaci n.

$$P(X=0) = \frac{np}{1-p+np},$$
$$P(X=k) = \frac{n^{k-1}p^{k-1} - 2n^{k-1}p^k + n^{k-1}p^{k+1}}{(1-p+np)^{k+1}}, \quad k = 1,2,\ldots.$$

Ověříme platnost vzorce pro 3. generaci při volbě parametrů $b = 0{,}25$ a $p = 0{,}5$. Provedeme simulaci 10 000 procesů a simulované pravděpodobnosti nejprve porovnáme s těmi napočtenými v tabulce pro $k \leq 10$ a následně také na grafu.

k	$P(X=k)$	**Simulované** $P(X=k)$
0	0,75	0,7448
1	0,0625	0,0640
2	0,0469	0,0462
3	0,0352	0,0340
4	0,0264	0,0302
5	0,0198	0,0181
6	0,0148	0,0168
7	0,0111	0,0105
8	0,00834	0,0085
9	0,00626	0,0076
10	0,00469	0,0052

Tabulka 1.2: Teoretické porovnání teoretické a simulované pravděpodobnosti velikosti populace ve 3. generaci při $p = 0{,}5$



Obrázek 1.2: Porovnání simulovaného rozdělení s teoretickým. Červené tečky znázorňují vypočítané pravděpodobnosti, modré obdélníky odpovídají odhadům příslušných pravděpodobností z 10 000 simulací.

Vidíme, že červené tečky značící teoretické pravděpodobnosti jsou téměř shodné s těmi naměřenými pomocí simulací.

1.2 Vyhynutí

V této kapitole je třeba si zavést tři nové termíny, které se budou odvíjet od střední hodnoty počtu potomků jedné buňky m.

Definice 2. *O větvícím procesu řekneme, že je subkritický, pokud $m < 1$.*
O větvícím procesu řekneme, že je kritický, pokud $m = 1$.
O větvícím procesu řekneme, že je superkritický, pokud $m > 1$.

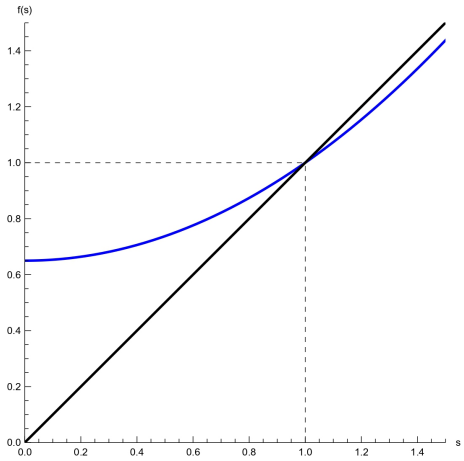
Pro výpočet vyhynutí procesu musíme nejdřív určit matematické vlastnosti pravděpodobnostní vytvořující funkce $f(s)$.

Věta 1. *Nechť s je reálné číslo a $m = \mathbb{E}[Z_1]$, pak $f(s)$ je dle definice 1 mocninná řada s nezápornými koeficienty p_k splňujícími $p_0 + p_1 < 1$ a platí*
(i) $f(s)$ je rostoucí ryze konvexní a rostoucí na $[0,1]$,
(ii) $f(0) = p_0; f(1) = 1$,
(iii) pokud je $m \leq 1$, pak $f(s) > s$ pro $s \in [0,1)$,
(iv) pokud $m > 1$, pak $f(s) = s$ má jednoznačný kořen na $[0,1)$.

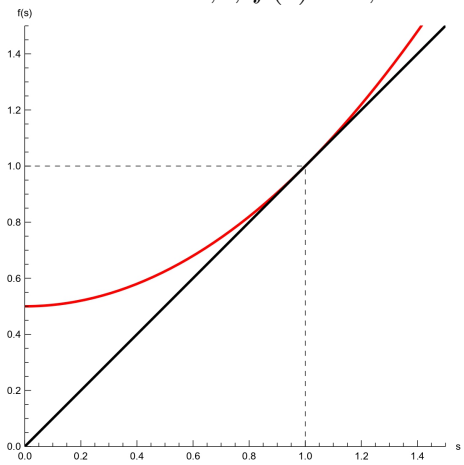
Tato věta nám říká, že existuje nejmenší kořen rovnice $f(s) = s$ pro $s \in [0,1]$. Ten jsme si označili si jako e a níže ukážeme, že se jedná o již zavedenou pravděpodobnost vymření procesu.

Věta 2. *Pokud je $m \leq 1$, pak $e = 1$.*
Pokud je $m > 1$, pak $e < 1$.

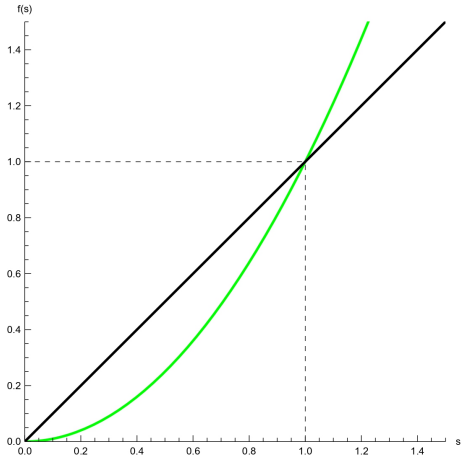
Tvrzení obou vět jsou snadno pochopitelná, pokud nakreslíme grafy pravděpodobnostních vytvořujících funkcí.



Obrázek 1.3: $m = 0{,}7,\ f(s) = 0{,}65 + 0{,}35s^2$



Obrázek 1.4: $m = 1,\ f(s) = 0{,}5 + 0{,}5s^2$



Obrázek 1.5: $m = 2,\ f(s) = s^2$

Obrázek 1.6: Porovnání pravděpodobnostních vytvořujících funkcí subkritického, kritického a superkritického procesu $g(s) = s$.

Věta 3. *Pokud $s \in [0,e]$, pak $f_n(s) \uparrow e$ pro $n \to \infty$.*
Pokud $s \in (e,1)$, pak $f_n(s) \downarrow e$ pro $n \to \infty$.
Pokud $s = e$ nebo 1, pak $f_n(s) = s$ pro všechna n.

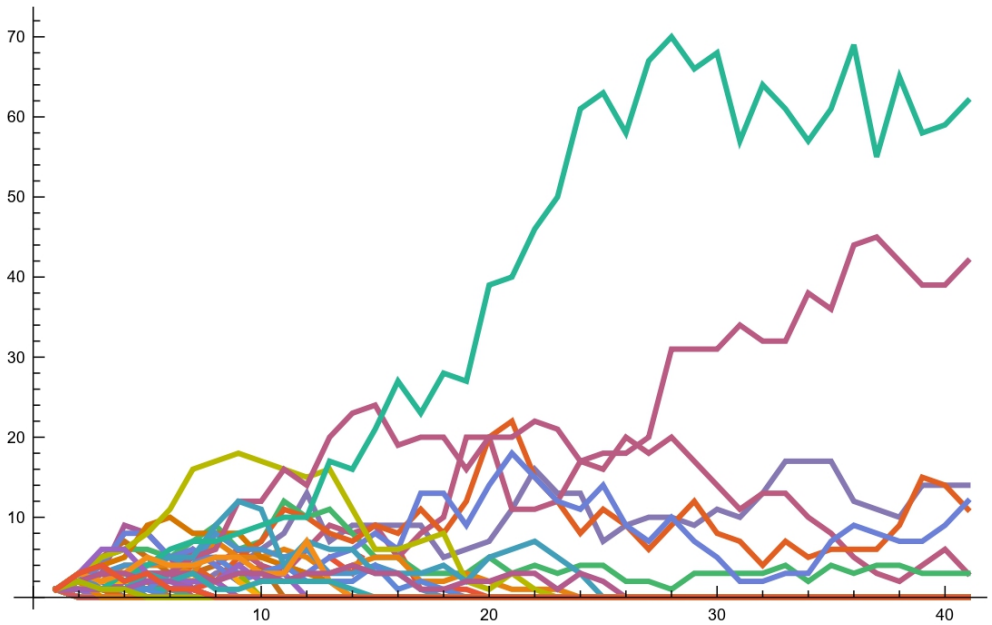
Speciální případ této věty, kdy $f_n(0) \uparrow q$, nám odhalí více. Získáme

$$\begin{aligned}\lim_{n\to\infty} f_n(0) &= \lim_{n\to\infty} P[Z_n = 0] \\ &= \lim_{n\to\infty} P[Z_i = 0 \text{ pro nějaké } 1 \leq i \leq n] \\ &= P[\exists i \geq 1 : Z_i = 0] \\ &= P[\lim_{n\to\infty} Z_n = 0],\end{aligned}$$

což je dle definice pravděpodobnost, že proces Z_n ve svém průběhu vyhyne. Všechny tyto informace nám nyní dávají možnost vytvořit větu o pravděpodobnosti vyhynutí procesu.

Věta 4. *Pravděpodobnost vyhynutí procesu Z_n je nejmenší nezáporný kořen rovnice $f(s) = s$. Je rovna 1, pokud $m \leq 1$ a je menší než jedna, pokud $m > 1$.*

Celý důkaz je zpracován v Grimmet a Stirzaker (2001) 5.4.5. Ukazuje se, že i kritické procesy s jistotou vyhynou, což lze pozorovat i na grafu zobrazujícím 100 kritických procesů a počet jedinců v prvních 40 generacích. Vidíme, že pouze 7 z nich po tomto milníku nevyhynulo. To platí samozřejmě i pro procesy subkritické.



Obrázek 1.7: Simulace kritických procesů

Superkritické procesy pak s pravděpodobností e vyhynou a pravděpodobností $1-e$ se budou rozmnožovat neomezeně.

1.2.1 Integrovaný Galtonův-Watsonův proces

Definujme si Y_n jako součet buněk ze všech generací do n, tedy

$$Y_n = \sum_{i=0}^{n} Z_i$$

a k tomu i její pravděpodobnostní vytvořující funkci $F_n(s)$. Y_n tedy značí počet všech buněk, které byly v procesu do generace n.

Podíváme se, jaký tvar bude mít funkce $F_n(s)$. Opět předpokládáme, že proces začal z jediné buňky,

$$F_0(s) = f_0(s) = s.$$

Pro získání funkce další generace ji musíme vložit do pravděpodobnostní funkce $f(s)$ a ještě ji přenásobit pomocí s. Platí totiž

$$\begin{aligned}
F_{n+1}(s) &= \mathbb{E}[s^{Y_{n+1}}] \\
&= \mathbb{E}\left[s^{1+\sum_{j=1}^{Z_1} Y_n^{(j)}}\right] \\
&= s \cdot \mathbb{E}\left[s^{\sum_{j=1}^{Z_1} Y_n^{(j)}}\right] \\
&= s \cdot \mathbb{E}\left[\mathbb{E}\left[s^{\sum_{j=1}^{Z_1} Y_n^{(j)}} \,\middle|\, Z_1\right]\right] \\
&= s \cdot \mathbb{E}\left[\left(\mathbb{E}\left[s^{Y_n^{(1)}}\right]\right)^{Z_1}\right] \\
&= s \cdot \mathbb{E}\left[(F_n(s))^{Z_1}\right] \\
&= s \cdot f(F_n(s))
\end{aligned}$$

$$F_{n+1}(s) = sf(F_n(s)).$$

Podrobnější důkaz je zpracován v Feller (1968) 12.1.4.

$F(s)$ pak bude značit pravděpodobnostní vytvořující funkci celkového počtu buněk celého procesu. Spočítá se jako

$$\lim_{n\to\infty} F_n(s),$$

která bude konečná, pokud nebude proces superkritický a platí pro ni

$$F(s) = sf(F(s)).$$

Pokusíme se ji explicitně spočítat v případě lineárních zlomků,

$$F(s) = s\left(1 - \frac{b}{1-p} + \frac{bF(s)}{1-pF(s)}\right)$$

to vede na kvadratickou rovnici se dvěma řešeními. Vybereme to, pro které platí $F(1) = 1$. Získáme

$$F(s) = \frac{mps - \sqrt{(-mps + ms - ps - 1)^2 - 4p(mps - ms + s)} - ms + ps + 1}{2p}.$$

1.3 Geometrické větvení

Všechny dosud získané znalosti nyní zkusíme aplikovat na příkladu. Řekněme, že se počet potomků jedné buňky řídí geometrickým rozdělením $P[Z = k] = qp^k$, přičemž platí $p = 1 - q$. Pravděpodobnostní vytvořující funkce pak bude mít tvar $f(s) = q(1-ps)^{-1}$. Pomocí indukce pak můžeme snadno zjistit i její n-tou iteraci.

$$f_n(s) = \begin{cases} \frac{n-(n-1)s}{n+1-ns}, & \text{platí-li } p = q = \frac{1}{2}, \\ \frac{q[p^n - q^n - ps(p^{n-1} - q^{n-1})]}{p^{n+1} - q^{n+1} - ps(p^n - q^n)}, & \text{platí-li } p \neq q. \end{cases} \tag{1.3}$$

Pro tento proces nyní můžeme použít věty o vyhynutí. Pravděpodobnost, že proces bude vyhynulý v n-té generaci činí:

$$P[Z_n = 0] = \begin{cases} \frac{n}{n+1}, & \text{platí-li } p = q = \frac{1}{2}, \\ \frac{q(p^n - q^n)}{p^{n+1} - q^{n+1}}, & \text{platí-li } p \neq q. \end{cases}$$

Pokud půjde n limitně k nekonečnu, získáme kýženou pravděpodobnost vymření procesu.

$$P[\text{vymření}] = \begin{cases} 1, & \text{platí-li } p \leq q, \\ \frac{q}{p}, & \text{platí-li } p > q. \end{cases}$$

Nyní můžeme spočítat, jaká bude pravděpodobnost toho, že proces vyhyne právě v n-té generaci. Definujme si náhodnou veličinu T=min$\{n : Z_n = 0\}$, což bude značit číslo první generace, kdy v procesu již nebyl žádný jedinec. Zajímá nás tedy jaká bude pravděpodobnost $P[T = n]$. Ta bude stejná, jako kdyby byl počet jedinců v generaci n roven nule, ale počet jedinců v generaci o jednu menší bude nenulový.

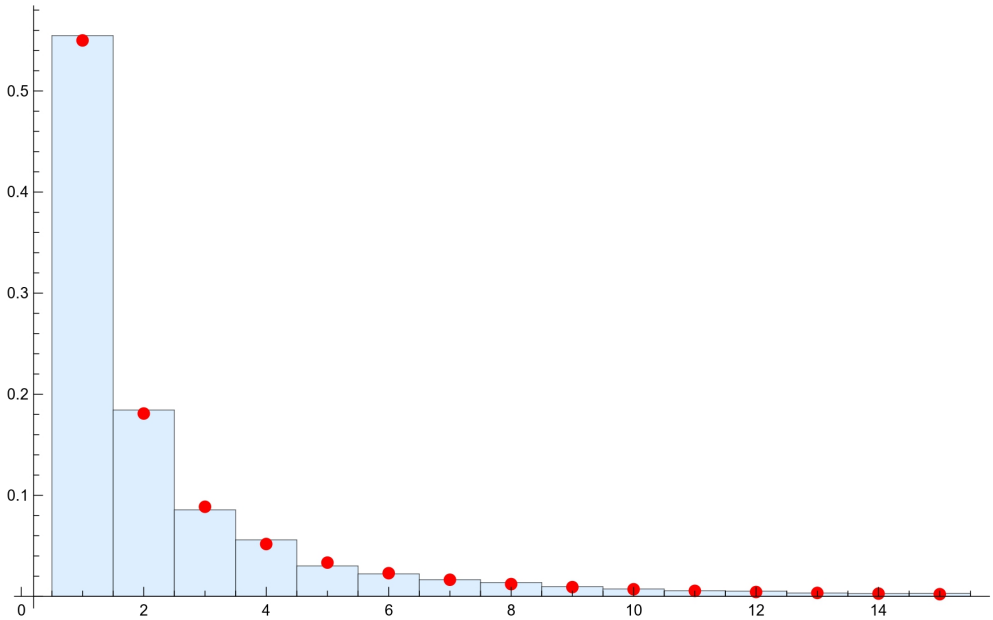
Čili s použitím výše uvedeného vzorce získáme:

$$P[T = n] = P[Z_n = 0] - P[Z_{n-1} = 0] = \begin{cases} \frac{1}{n(n+1)}, & \text{platí-li } p = q, \\ \frac{p^{n-1} q^n (p-q)^2}{(p^n - q^n)(p^{n+1} - q^{n+1})}, & \text{platí-li } p \neq q. \end{cases}$$

Nyní aplikujme tyto vzorce na jeden specifický příklad. Nastavme si parametr p na hodnotu 0,45. Parametr q poté bude samozřejmě 0,55. Pokusíme se ověřit vzorec pro výpočet pravděpodobnosti délky procesu. Vygenerujeme si deset tisíc simulací a jejich délky zaznamenáme do tabulky a porovnáme s hodnotami napočtenými podle vzorce v grafu. Vzhledem k tomu, že je parametr p menší než parametr q, budou všechny procesy konečné.

n	$P(T = n)$	**Simulované** $P(T = n)$
1	0,5500	0,5672
2	0,1809	0,1823
3	0,0887	0,0875
4	0,0518	0,0531
5	0,0334	0,0322
6	0,0229	0,0231
7	0,0164	0,0167
8	0,0121	0,0128
9	0,0092	0,0095
10	0,0071	0,0075
11	0,0055	0,0053
12	0,0043	0,0045
13	0,0034	0,0034
14	0,0027	0,0026

Tabulka 1.3: Teoretické a simulované pravděpodobnosti doby přežití populace při $p = 0{,}45$



Obrázek 1.8: Porovnání simulované pravděpodobnosti délky procesu (modré sloupce) s teoretickou (červené tečky).

Vidíme, že teoretické pravděpodobnosti korespondují s těmi naměřenými. Pokud platí předpoklad, že má rozdělení počtu potomků jednotlivé buňky geometrickou pravděpodobnost, můžeme provést odhad parametru q pomocí momentové metody. Ten označíme $\hat{q}_n$ a výběrový průměr náhodného výběru Z_n z geometric-

kého rozdělení označíme $\overline{Z}_n$.

$$\begin{aligned}
\mathbb{E}[Z] &= \frac{1-q}{q} \\
q &= \frac{1}{1+\mathbb{E}[Z]} \\
\hat{q}_n &= \frac{1}{1+\overline{Z}_n}
\end{aligned}$$

Pokud bychom provedli odhad pomocí metody maximální věrohodnosti, vyšel by totožně.

Nyní se pokusíme ověřit pravděpodobnost vyhynutí. Aby nebyla pravděpodobnost triviální, musíme si zvolit parametr p větší, než parametr q. V našem případě zvolíme 2/3 a 1/3. Teoretická pravděpodobnost bude podílem q a p, čili 1/2. Ověřit toto číslo je ovšem nemožné, jelikož jsme schopni generovat pouze procesy do konečné generace. Je však možné stanovit si jakousi dolní mez pro tuto hodnotu, protože ať je proces jakkoliv velký, může stále vyhynout. Vygenerujme si 10 000 procesů a porovnejme, jaká část dříve převýší hranici 10 000 jedinců v jedné generaci a jaká část vyhyne. Zjistíme, že 49,2 % procesů vyhynulo před dosažením tohoto milníku, což je velice blízko naší teoretické pravděpodobnosti.

S geometrickým větvením se pojí ještě jedna zajímavá vlastnost. Definujme si nejprve charakteristickou funkci náhodné veličiny Z.

Definice 3. *Nechť $i = \sqrt{-1}$, Charakteristickou funkcí náhodné veličiny Z $\phi_Z(t) : \mathbb{R} \to \mathbb{C}$ definujeme jako $\phi_Z(t) = \mathbb{E}(e^{itZ})$.*

Pomocí této funkce ukážeme, že podmíněné rozdělení náhodné veličiny Z_n/m^n za podmínky $Z_n > 0$ konverguje při $n \to \infty$ k exponenciálnímu rozdělení s parametrem $1 - m^{-1}$, kde m opět značí střední hodnotu počtu potomků jedné buňky

$$m = \mathbb{E}[Z_1] = \frac{p}{q} > 1.$$

Podmíněná charakteristická funkce procesu Z_n/m^n je dle definice 3

$$\mathbb{E}[e^{itZ_n/m^n}|Z_n > 0] = \frac{f_n(e^{it/m^n}) - f_n(0)}{1 - f_n(0)}. \tag{1.4}$$

Nyní můžeme přeformulovat výraz 1.3 pomocí m abychom získali rovnici

$$f_n(s) = \frac{m^n - 1 - ms(m^{n-1} - 1)}{m^{n+1} - 1 - ms(m^n - 1)},$$

kterou nyní dosadíme do 1.4 a získáme

$$\mathbb{E}[e^{itZ_n/m^n}|Z_n > 0] \to \frac{m-1}{m-1-mit},$$

což je charakteristická funkce exponenciálního rozdělení s parametrem $1 - m^{-1}$.

2. Asymptotické vlastnosti

V této kapitole probereme asymptotické vlastnosti Galtonových Watsonových procesů. Ty se budou lišit podle toho, zda se bude jednat o subkritické, kritické, nebo superkritické. U těch, které s jistotou vyhynou, budeme zkoumat jejich střední hodnotu za podmínky, že je proces stále aktivní, což dává u těchto procesů daleko lepší infomaci o jejich chování než samotná střední hodnota. U těch, které vyhynout nemusí, budeme naopak zkoumat, zda se limitně blíží k nějakému násobku jejich střední hodnoty.

2.1 Subkritické procesy

Víme, že pravděpodobnost, že proces bude vyhynulý v generaci n je $f_n(0)$ čili pravděpodobnost, že proces v této generaci bude stále aktivní, bude $1 - f_n(0)$. Vypočítáme pak podmíněnou střední hodnotu jako

$$\mathbb{E}(Z_n \mid Z_n > 0) = \frac{\mathbb{E}(Z_n)}{1 - f_n(0)} = \frac{m^n}{1 - f_n(0)}.$$

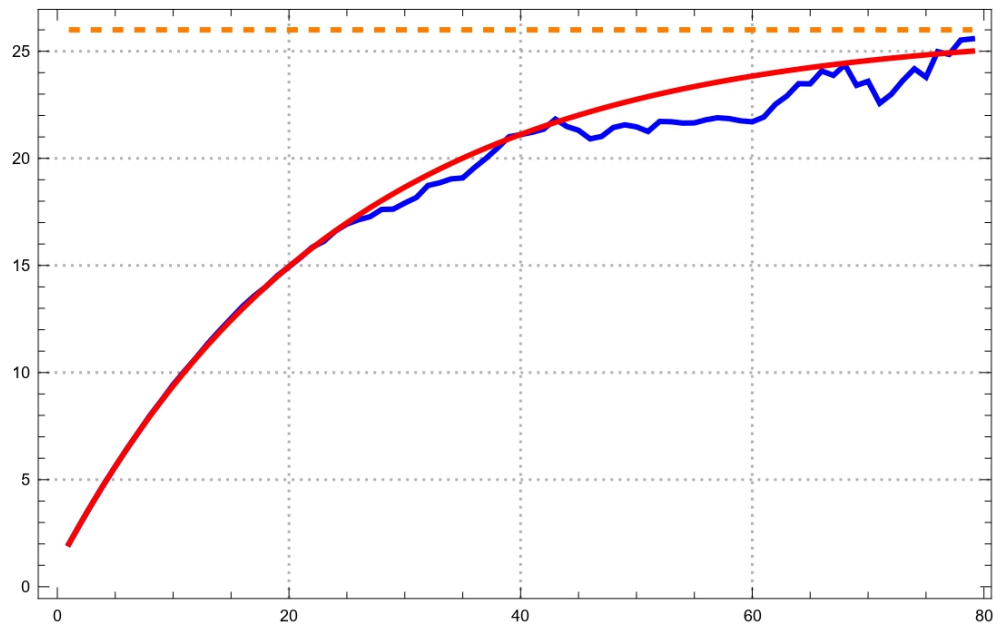
Ukážeme to na opět na příkladu lineárně-lomeného rozdělení a vypočteme i limitu pro n jdoucí k nekonečnu,

$$1 - f_n(0) = m^n \left(\frac{1 - e^*}{m^n - e^*} \right),$$

a tedy

$$\mathbb{E}(Z_n \mid Z_n > 0) = \frac{m^n - e^*}{1 - e^*} \to \frac{e^*}{e^* - 1}, \quad n \to \infty.$$

Ověříme tento vzorec na simulaci. Zvolíme parametr $b = 0{,}24$ a parametr $p = 0{,}5$, čímž získáme $m = 0{,}96$ a $e^* = 1{,}04$, což nám dá limitu podmíněné střední hodnoty 26. Vytvoříme simulaci 100 000 procesů a podíváme se, zda naměřené hodnoty korespondují s těmi napočítanými.



Obrázek 2.1: Porovnání simulované střední hodnoty s teoretickou. Červená čára znázorňuje teoretickou střední hodnotu pro danou generaci, modrá čára odpovídá podmíněné výběrové střední hodnotě z 100 000 procesů, oranžová čára odpovídá teoretické limitě podmíněné středné hodnoty.

Vidíme, že zprvu naměřené hodnoty korespondují s těmi teoretickými téměř dokonale, později se ale odchylky zvětšují. Příčina tohoto jevu je, že i přesto, že máme nadstandardní množství simulací, jen málo procesů přežije až do osmdesáté generace a výběrové střední hodnoty jsou vypočteny z malého množství dat.

2.2 Kritické procesy

Kritické procesy se mohou zdát velmi podobné těm subkritickým, ale vzorce popisující jejich chování jsou naprosto odlišné. Zároveň jsou ale více specifické, což využijeme v následujícím lemmatu.

Lemma 2.1 *Pokud* $m = \mathbb{E}(Z_1) = 1$ *a* $\sigma^2 = \mathrm{Var}(Z_1) < \infty$, *pak*

$$\lim_{n\to\infty} \frac{1}{n}\left[\frac{1}{1-f_n(t)} - \frac{1}{1-t}\right] = \frac{\sigma^2}{2} \tag{2.1}$$

rovnoměrně pro $0 \leq t < 1$. Důkaz tohoto lemmatu je zpracován v Athreya a Ney (2004) 1.9.1. Díky němu můžeme vypočítat pravděpodobnost, že je proces v generaci n stále aktivní. Stačí dosadit $t = 0$. Získáme

$$P(Z_n > 0) = 1 - f_n(0) \sim \frac{2}{n\sigma^2}; \quad n \to \infty,$$

díky čemuž můžeme vypočítat opět i podmíněnou střední hodnotu

$$\mathbb{E}(Z_n \mid Z_n > 0) = \frac{m}{P(Z_n > 0)} = \frac{1}{P(Z_n > 0)} \sim \frac{n\sigma^2}{2}; \quad n \to \infty$$

a následně vytvořit limitní větu pro normalizovaný proces.

Věta 2.2 Pokud $m = 1$ a $\sigma^2 < \infty$, pak

$$\lim_{n\to\infty} P\left(\frac{Z_n}{n} > z \middle| Z_n > 0\right) = \exp\left(-\frac{2z}{\sigma^2}\right), \quad z \geq 0. \tag{2.2}$$

Důkaz věty 2.2 Ukážeme, že

$$\lim_{n\to\infty} \mathbb{E}\left[e^{-\alpha\frac{Z_n}{n}} \middle| Z_n > 0\right] = \frac{1}{1+\frac{\alpha\sigma^2}{2}}, \tag{2.3}$$

což je postačující, protože pravá strana (2.3) je Laplaceova transformace pravé strany (2.2). Podmíněná střední hodnota ve vztahu (2.3) se rovná

$$\frac{f_n\left(e^{-\frac{\alpha}{n}}\right) - f_n(0)}{1 - f_n(0)} = 1 - \frac{\{n[1 - f_n(0)]\}^{-1}}{\left\{n[1 - f_n(e^{-\frac{\alpha}{n}})]\right\}^{-1}}. \tag{2.4}$$

Podle lemmatu (2.1) platí

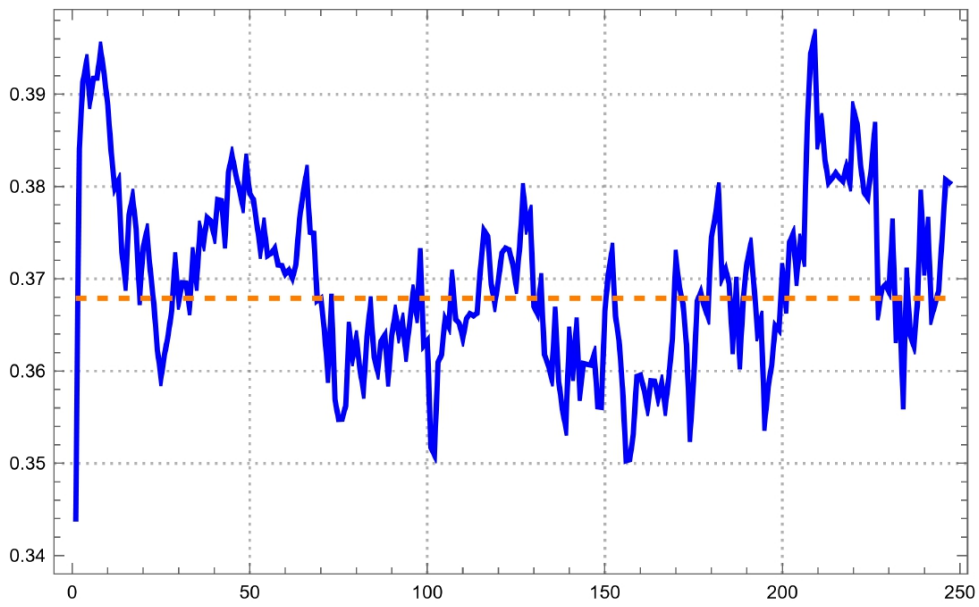
$$\lim_{n\to\infty} \frac{1}{n[1 - f_n(0)]} = \frac{\sigma^2}{2}, \tag{2.5}$$

a (použitím stejnoměrné konvergence)

$$\lim_{n\to\infty} \frac{1}{n[1 - f_n(e^{-\frac{\alpha}{n}})]} = \frac{\sigma^2}{2} + \lim_{n\to\infty} \frac{1}{n(1 - e^{-\frac{\alpha}{n}})} = \frac{\sigma^2}{2} + \frac{1}{\alpha}, \tag{2.6}$$

což spolu s (2.4) implikuje (2.3) a tím i tvrzení věty.

Ukážeme si simulaci na základě této věty. Máme opět 100 000 simulací kritického procesu s lineárně-lomeným rozdělením s parametry $b = 0{,}25$ a $p = 0{,}5$ při zvoleném $z = 1$. Rozptyl je tedy $\sigma^2 = 2$ a podmíněná pravděpodobnost by měla konvergovat k $\exp(-1)$.



Obrázek 2.2: Porovnání simulovaných podmíněných pravděpodobností(modrá čára) s jejich teoretickou limitou (oranžová čára).

Vidíme, že naměřená podmíněná pravděpodobnost kolísá kolem své teoretické limity.

2.3 Superkritické procesy

Superkritické procesy se liší od zbylých tím, že nemusí s jistotou vůbec vyhynout, čili se bude lišit celý proces výpočtu jejich limitních vět. Začneme definicí nového procesu

$$W_n = \frac{Z_n}{m^n},$$

u kterého nás bude zajímat jeho limita, pokud nějakou má. Začneme tím, že se přesvědčíme, že se jedná o martingal.

Definice 2.3 *Posloupnost náhodných veličin* $\{X_n, n \geq 0\}$ *se nazývá martingal, pokud* $\mathbb{E}(|X_0|) < \infty$ *a platí*

$$\mathbb{E}(X_{n+1} \mid X_n, X_{n-1}, \ldots, X_1, X_0) = X_n.$$

Víme, že $\mathbb{E}(W_n) = 1$ a také, že

$$\mathbb{E}(W_{n+1} \mid W_n) = m^{-(n+1)}\mathbb{E}(Z_{n+1} \mid Z_n) = m^{-(n+1)} m Z_n = W_n.$$

a vzhledem k tomu, že Galtonův-Watsonův proces je Markovův řetězec, pak

$$\mathbb{E}(Z_{n+1} \mid Z_n, Z_{n-1}, \ldots, Z_1, Z_0) = \mathbb{E}(Z_{n+1} \mid Z_n).$$

což analogicky platí i pro proces $\{W_n\}$

$$\mathbb{E}(W_{n+1} \mid W_n, W_{n-1}, \ldots, W_1, W_0) = \mathbb{E}(W_{n+1} \mid W_n).$$

Díky tomu můžeme použít část (i) z následující věty:

Věta 2.4 Pokud $\{X_n, n \geq 0\}$ je nekladný martingal takový, že $E(X_n) < \infty$ pro všechna n, pak existuje náhodná proměnná X s konečnou střední hodnotou taková, že:

(i)
$$\lim_{n\to\infty} X_n = X, \quad \text{skoro jistě.}$$

(ii) Pokud je martingal omezen v L^2, tj. pokud $\sup_n E(X_n^2) < \infty$, pak konvergence nastává také ve smyslu L^2. Potom platí $\mathrm{Var}(X) = \lim_{n\to\infty} \mathrm{Var}(X_n)$.

Pokud navíc budeme mít konečný rozptyl, platí

Věta 2.5 Pokud $m > 1$, $\sigma^2 < \infty$ a $Z_0 \equiv 1$, pak platí:

(i)
$$\lim_{n\to\infty} \mathbb{E}(W_n - W)^2 = 0.$$

(ii)
$$\mathbb{E}(W) = 1, \quad \mathrm{Var}(W) = \frac{\sigma^2}{m^2 - m}.$$

(iii)
$$P\{W = 0\} = q = P\{Z_n = 0 \text{ pro nějaké } n\}.$$

Obě věty jsou dokázány v Neveu (1975). K tomu, abychom se dostali přímo k limitním vzorcům, musíme použít Laplaceovu transformaci $\phi(v) = \mathbb{E}(e^{-vW}), v \in \mathbb{R}$, rozdělení W, která může být vyjádřena funkční rovnicí

$$\phi(v) = f\left[\phi\left(\frac{v}{m}\right)\right], \tag{2.7}$$

která je známá jako Abelova rovnice. Vztah (1.1) lze přepsat pomocí Laplaceových transformací jako $\varphi_n(u) = \mathbb{E}[\exp(-uZ_n)] = f_n[\exp(-u)]$, kde $u \geq 0$ je symbolický argument

$$\varphi_{n+1}(u) = f[\varphi_n(u)]. \tag{2.8}$$

Protože Laplaceova transformace rozdělení W_n je dána jako

$$\phi_n(v) = \mathbb{E}[\exp(-uW_n)] = \mathbb{E}\left[\exp\left(-\frac{u}{m^n}Z_n\right)\right] = \varphi_n(u/m^n), \tag{2.9}$$

platí i opačný vztah $\varphi_n(u) = \phi_n(um^n)$. Dosazením do rovnice (2.8) získáme

$$\phi_{n+1}(um^{n+1}) = f[\phi_n(um^n)]. \tag{2.10}$$

Po změně proměnné $v = um^{n+1}$ dostáváme

$$\phi_{n+1}(v) = f\left[\phi_n\left(\frac{v}{m}\right)\right]. \tag{2.11}$$

Protože platí $W_n \xrightarrow{D} W$, platí také $\phi_n(v) \to \phi(v)$. Limita (která odpovídá Laplaceově transformaci rozdělení W) pak splňuje Abelovu rovnici (2.7).

Pokusíme se vypočítat hustotu rozdělení W pro lineárně-lomenné rozdělení. Nejprve musíme vyřešit funkční rovnici 2.9

$$\phi(v) = f\left[\phi\left(\frac{v}{m}\right)\right] = 1 - \frac{b}{1-p} + \frac{b \cdot \phi(v/m)}{1 - p \cdot \phi(v/m)}.$$

Tím získáme

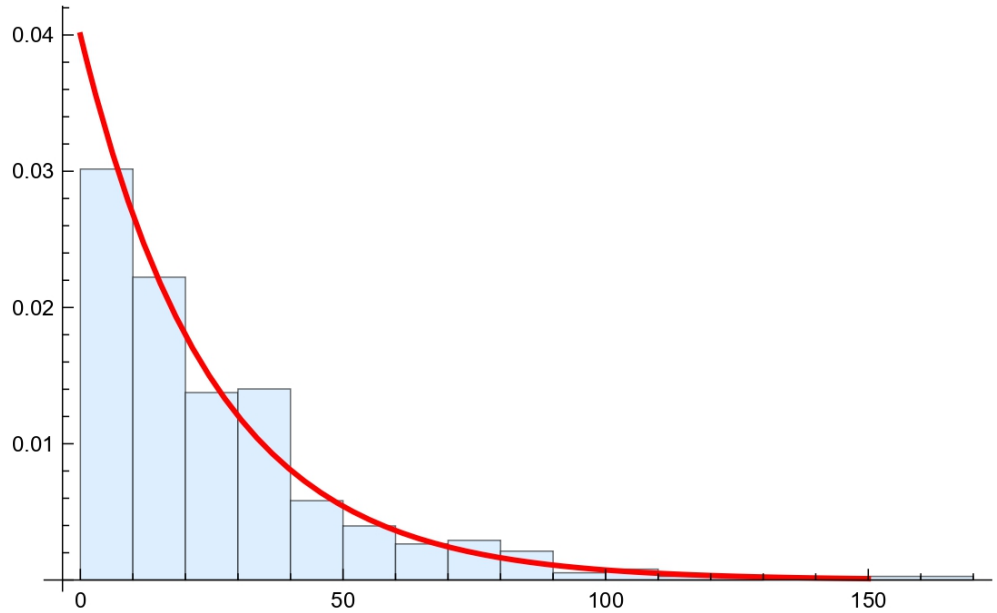
$$\phi(v) = e^* + \frac{(1-e^*)^2}{v + (1-e^*)}.$$

Nyní už zbývá jen použít zpětnou Laplaceovu transformaci, čímž dojdeme k výsledku

$$f_W(w) = e^*\delta(0) + (1-e^*)^2 e^{-(1-e^*)w}, \quad w \geq 0,$$

kde $\delta(w)$ je Diracova delta, funkce reprezentující pravděpodobnostní hmotu e^* v bodě $w = 0$. To ukazuje, že W má smíšené rozdělení s atomem v nule a exponenciálním rozdělením pro $w > 0$ podmíněno tím, že je proces stále aktivní.

Ověříme si tuto hustotu na simulaci 10 000 procesů s parametry $b = 0{,}26$, $p = 0{,}5$. Tím pádem bude parametr e^* roven 0,96. Ukážeme pouze rozdělení podmíněné tím, že proces nevyhynul, čili rozdělení, které je v našem případě tvořené jen 376 procesy. Je také nutné zmínit, že tato hustota platí pouze pokud $n \to \infty$. Pro naše účely vezmeme pouze stou generaci, tedy $n = 100$.



Obrázek 2.3: Porovnání simulované hustoty náhodné veličiny W s teoretickou na 10 000 simulacích. Modré slopce značí pozorovanou hustotu a červená čára teoretickou.

Opět vidíme, že teoretická pravděpodobnost koresponduje s tou naměřenou.

Závěr

Tato bakalářská práce se věnovala studiu základního typu větvících procesů, a to Galtonových-Watsonových procesů. V úvodní části byly definovány základní pojmy a představeny pravděpodobnostní vytvořující funkce jako klíčový nástroj pro jejich popis, z nichž pak bylo odvozeno rozdělení počtu jedinců generace n.

Práce se podrobně zabývala problematikou vyhynutí procesu. Bylo ukázáno, jak střední hodnota počtu potomků jedné částice zásadně ovlivňuje dlouhodobé chování procesu. Dále byl krátce zmíněn koncept integrovaného G.-W. procesu.

Jako konkrétní příklad byl analyzován proces s geometrickým rozdělením počtu potomků, pro který byly explicitně odvozeny vytvořující funkce a pravděpodobnosti vyhynutí.

V poslední kapitole byly diskutovány asymptotické vlastnosti procesů. Pro subkritické a kritické procesy byla zkoumána podmíněná střední hodnota za předpokladu, že proces ještě nevyhynul. Pro superkritické procesy byl definován normalizovaný proces a byla zmíněna jeho konvergence.

Teoretické poznatky a odvozené vzorce byly v průběhu práce ověřovány pomocí simulací. Bylo například demonstrováno rozdělení počtu jedinců v dané generaci, doba do vyhynutí procesu nebo asymptotické chování podmíněné střední hodnoty a podmíněné pravděpodobnosti. Tyto simulace potvrdily platnost teoretických výsledků a poskytly názornější představu o chování studovaných procesů.

Práce tak splnila svůj cíl: představila teorii G.-W. procesů, vyřešila několik souvisejících problémů a demonstrovala platnost teoretických vzorců na simulacích.

Seznam použité literatury

Athreya, K. B. a Ney, P. E. (2004). *Branching Processes.* Dover Books on Mathematics. Dover Publications, New York. ISBN 978-8888852065.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications, Volume I.* John Wiley & Sons, New York, third edition. ISBN 978-0471257080.

Grimmet, G. R. a Stirzaker, D. R. (2001). *Probability and Random Processes.* Third edition. Oxford University Press Inc., New York. ISBN 0-19-857222-0.

Kimmel, M. a Axelrod, D. E. (2015). *Branching Processes in Biology.* Second Edition. Springer, New York. ISBN 978-1-4939-1559-0.

Neveu, J. (1975). *Discrete-Parameter Martingales.* Elsevier Science, Amsterdam. ISBN 978-0720428100.

Seznam obrázků

1.1 Schéma zpětného vzorce 4
1.2 Porovnání simulovaného rozdělení s teoretickým. Červené tečky znázorňují vypočítané pravděpodobnosti, modré obdélníky odpovídají odhadům příslušných pravděpodobností z 10 000 simulací. . 7
1.3 $m = 0{,}7,\ f(s) = 0{,}65 + 0{,}35s^2$ 9
1.4 $m = 1,\ f(s) = 0{,}5 + 0{,}5s^2$ 9
1.5 $m = 2,\ f(s) = s^2$ 9
1.6 Porovnání pravděpodobnostních vytvořujících funkcí subkritického, kritického a superkritického procesu $g(s) = s$. 9
1.7 Simulace kritických procesů 10
1.8 Porovnání simulované pravděpodobnosti délky procesu (modré sloupce) s teoretickou (červené tečky). 13

2.1 Porovnání simulované střední hodnoty s teoretickou. Červená čára znázorňuje teoretickou střední hodnotu pro danou generaci, modrá čára odpovídá podmíněné výběrové střední hodnotě z 100 000 procesů, oranžová čára odpovídá teoretické limitě podmíněné středné hodnoty. 16
2.2 Porovnání simulovaných podmíněných pravděpodobností(modrá čára) s jejich teoretickou limitou (oranžová čára). 18
2.3 Porovnání simulované hustoty náhodné veličiny W s teoretickou na 10 000 simulacích. Modré slopce značí pozorovanou hustotu a červená čára teoretickou. 20

Seznam tabulek

1.1 Tabulka nejpoužívanějších proměnných 3
1.2 Teoretické porovnání teoretické a simulované pravděpodobnosti velikosti populace ve 3. generaci při $p = 0{,}5$ 7
1.3 Teoretické a simulované pravděpodobnosti doby přežití populace při $p = 0{,}45$ 13

Seznam použitých zkratek

A. Přílohy

A.1 První příloha