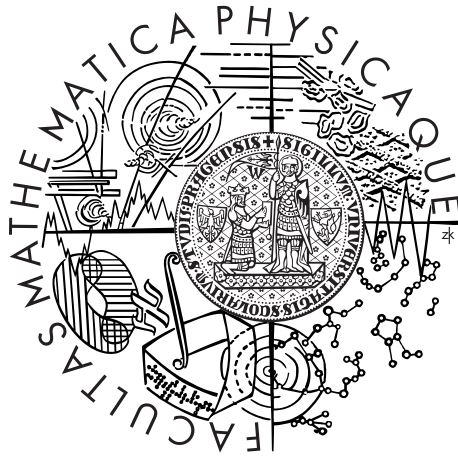


Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE



Barbora Vícenová

Prostorový bodový proces s interakcemi

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: prof. RNDr. Viktor Beneš, DrSc.

Studijní program: Matematika

Studijní obor: Finanční a pojistná matematika

Praha 2015

Na tomto mieste by som sa rada poďakovala vedúcemu prof. RNDr. Viktorovi Benešovi, DrSc., za odborné vedenie a cenné rady pri písaní práce, za zapožičanie literatúry a najmä za čas a trpezlivosť pri početných konzultáciách. Zároveň by som chcela poďakovať prof. Stephanovi Huckemannovi z Univerzity v Göttingene za poskytnutie reálnych bioštatistických dát.

Prohlašuji, že jsem tuto diplomovou práci vypracovala samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

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 v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V Praze dne 6.5.2015

Podpis autora

Název práce: Prostorový bodový proces s interakcemi

Autor: Barbora Vícenová

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: prof. RNDr. Viktor Beneš, DrSc., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Předložená práce se zabývá odhadem parametrů modelu procesu úseček s interakcemi v rovině. Motivací je aplikace na systém svalových vláken v lidských kmenových buňkách, zobrazených fluorescenční mikroskopií. Zavedeme model procesu úseček jako prostorový Gibbsův bodový proces s příznakem a definujeme dvě metody na odhad parametrů: momentovou metodu a metodu Takacs-Fiksel. Dále implementujeme algoritmus pro odhady těmito metodami v programu Mathematica. Modelovou strukturu jsme též schopni simulovat pomocí Markov chain Monte Carlo, užitím procesu rození a zániku. Jsou prezentovány numerické výsledky pro reálná i simulovaná data, shoda modelu s daty se posuzuje pomocí popisných statistik.

Klíčová slova: bodový proces, metoda MCMC, odhady parametrů

Title: Spatial point process with interactions

Author: Barbora Vícenová

Department: Department of Probability and Mathematical Statistics

Supervisor: prof. RNDr. Viktor Beneš, DrSc., Department of Probability and Mathematical Statistics

Abstract: This thesis deals with the estimation of model parameters of the interacting segments process in plane. The motivation is application on the system of stress fibers in human mesenchymal stem cells, which are detected by fluorescent microscopy. The model of segments is defined as a spatial Gibbs point process with marks. We use two methods for parameter estimation: moment method and Takacs-Fiksel method. Further, we implement algorithm for these estimation methods in software Mathematica. Also we are able to simulate the model structure by Markov Chain Monte Carlo, using birth-death process. Numerical results are presented for real and simulated data. Match of model and data is considered by descriptive statistics.

Keywords: point process, Markov chain Monte Carlo, parameter estimation

Názov práce: Priestorový bodový proces s interakciami

Autor: Barbora Vícenová

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedúci diplomovej práce: prof. RNDr. Viktor Beneš, DrSc., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Predložená práca sa zaoberá odhadom parametrov modelu procesu úsečiek s interakciami v rovine. Motiváciou je aplikácia na systém svalových vlákien v ľudských kmeňových bunkách zobrazených fluorescentnou mikroskopiou. Zavedieme model procesu úsečiek ako priestorový Gibbsov bodový proces s príznakom. Definujeme dve metódy na odhad parametrov: momentovú metódu a metódu Takacs-Fiksel. Ďalej v programe Mathematica implementujeme algoritmus pre výpočet odhadov parametrov týmito metódami. Modelovú štruktúru sme schopní simulovať pomocou Markov chain Monte Carlo, použitím procesu rodenia a zániku. Prezentujeme numerické výsledky pre reálne aj simulované dáta. Zhoda modelu s dátami sa posudzuje pomocou popisných štatistík.

Kľúčové slová: bodový proces, metóda MCMC, odhad parametrov

Obsah

1	Základné pojmy	3
1.1	Bodový proces	3
1.2	Poissonov proces	5
1.3	Priestor výstupov	7
1.4	Konečné bodové procesy s hustotou	9
1.5	U -štatistiky	13
2	Model úsečiek s interakciami	16
2.1	Simulácia modelu úsečiek	17
3	Štatistika procesov úsečiek	19
3.1	Momentová metóda odhadu parametrov (MM)	19
3.2	Metóda Takacs-Fiksel (T-F)	20
3.3	Porovnávacie štatistiky	23
4	Dáta a popis kódu	25
4.1	Reálne dáta	25
4.2	Popis kódu	27
5	Numerické výsledky	30
5.1	Reálne dáta (bunka1)	30
5.2	Simulované dáta (bunkaSI)	37
5.3	Nehomogénna reálna bunka (bunka2)	44
	Zoznam použitej literatúry	47
	Zoznam obrázkov	49
	Zoznam tabuliek	51
	Zoznam použitých skratiek	52
A	Obrázkové prílohy	53
B	Obsah priloženého CD	61

Úvod

Priestorové bodové procesy sú dobre známe objekty v teórii pravdepodobnosti a sú silným nástrojom v štatistike pre modelovanie a analyzovanie priestorových dát. My sa v práci budeme zaoberať bodovým procesom v rovine, pričom pre body uvažujeme ešte vlasnosť dĺžka a orientácia. Stred - bod v rovine spolu s jeho dvomi dodatočnými vlasnosťami vytvorí geometrický útvar a to úsečku.

Motiváciou pre skúmanie procesu úsečiek je téma z praxe, z oblasti bioštatistiky. V súčasnosti je skúmanie ľudských kmeňových buniek veľmi populárna téma a len nedávno sa objavili efektívne metódy ako detekovať stresové vlákna v týchto bunkách za rôznych podmienok a tým predpovedať ich ďalšie správanie. Stresové vlákna sú približne lineárne a preto je vhodné ich modelovať ako úsečky.

Cieľom tejto práce je aplikovať na reálne aj simulované dáta procesu úsečiek dve metódy odhadu parametrov modelu, vhodne upravené pre proces úsečiek. Pomocou týchto odhadov simulujeme procesom rodenia a zániku proces úsečiek a vyhodnocujeme, ako dobre model korešponduje s podkladovými dátami. Pri simulácii využívame úsečky generované metódou MCMC. Za účelom porovnania výsledkov zavádzame vhodné popisné štatistiky.

V prvej kapitole sa zaoberáme všeobecným popisom bodových procesov ako aj rôznymi vlasnosťami platnými pre priestorový bodový proces s príznakom. Zavádzame nevyhnutné pojmy a značenie, ktoré sú východiskom pre zvyšok práce. Osobitne v druhej kapitole je zadefinovaný samotný model úsečiek s interakciami a taktiež popis simulácie modelu procesom rodenia a zániku. V tretej kapitole sú detailne popísané jednotlivé metódy odhadu parametrov modelu - momentová metóda a metóda Takacs-Fiksel, ktorú sme pre účely tejto diplomovej práce modifikovali z procesu kruhov na proces úsečiek. Na záver kapitoly sme zaradili popisné štatistiky, ktoré budeme používať na zhodnotenie výsledkov. Stručný popis pôvodu a formátu reálnych dát uvádzame v kapitole štyri. Bližšie sme rozobrali aj samotnú implementáciu metód odhadu parametrov a simuláciu procesu rodenia a zániku v softvéri Mathematica 10.0 (Wolfram Research, Inc., 2014). V poslednej kapitole sme sa zamerali na numerické výsledky, ktoré vyšli použitím algoritmu na reálne aj simulované dáta.

Kapitola 1

Základné pojmy

1.1 Bodový proces

Bodový proces je náhodná vzorka bodov v d -rozmernom priestore. Jednorozmerný proces môže byť použitý ako model postupnosti náhodných časov výskytu nejakej špecifickej udalosti. Priestorový bodový proces sa využíva pri modelovaní náhodného usporiadania bodov v d -rozmernom priestore pre $d \geq 2$. Príkladom dvojrozmerného procesu je mapa výskytu konkrétneho druhu stromu v lese, pričom poloha aj počet stromov je náhodný.

V teórii k jednorozmerným a priestorovým bodovým procesom sú rozdiely spôsobené mimo iné absenciou prirodzeného usporiadania bodov vo vyšších dimenziách. Na jednorozmerný bodový proces sa môžeme pozeráť viacerými spôsobmi, napr. ako na časy výskytu udalosti $T_1 < T_2 < \dots$, kedy ale jednotlivé časy T_i sú na sebe závislé. Iná možnosť je zaoberať sa časmi medzi jednotlivými výskytmi $S_i = T_{i+1} - T_i$, ktorá je výhodná najmä v modeloch, v ktorých sú S_1, S_2, \dots nezávislé. Ďalším možným prístupom je charakterizovať bodový proces ako kumulatívny sčítací, kde

$$N_t = \sum_{i=1}^{\infty} \mathbf{1}\{T_i \leq t\},$$

pre všetky $t \geq 0$, pričom $\mathbf{1}\{\dots\}$ označuje funkciu indikátor, ktorá je rovná 1, ak výraz "...“ je pravdivý a 0 inak. Tento prístup je výhodný v tom, že umožňuje proces previesť na náhodnú funkciu spojitého času t , ale nevýhodou je opäť závislosť jednotlivých veličín N_t . Posledný z pohľadov na jednorozmerný bodový proces, ktorý uvedieme, sú intervalové počty

$$N(a, b] = N_b - N_a$$

pre $0 \leq a \leq b$. Pre Poissonov proces sú intervalové počty pre disjunktné intervaly stochasticky nezávislé.

Vo vyšších dimenziách nie je možné nájsť prirodzenú analógiu k metóde určovania časov medzi jednotlivými výskytmi S_i ani k sčítaciemu procesu N_t . Užitočným spôsobom ako narábať s priestorovým procesom je zovšeobecniť intervalové počty $N(a, b]$ na množinové počty $N(A)$, ktoré pre každú ohraničenú a uzavretú množinu $A \subset \mathbb{R}^d$ určujú počet bodov v množine A . Hodnoty sčítacej premennej $N(A)$ pre všetky podmnožiny A nám poskytujú dostatočnú informáciu k tomu, aby sme vedeli zrekonštruovať pozíciu všetkých bodov procesu. Preto môžeme bodový proces definovať aj ako súbor náhodných veličín $N(A)$ indexovaných množinami A . Sčítacie premenné $N(A)$ pre rozličné množiny A spĺňajú určité vzťahy, ako vlastnosť aditivity

$$N(A \cup B) = N(A) + N(B),$$

vždy keď A a B sú disjunktné množiny ($A \cap B = \emptyset$) a samozrejme

$$N(\emptyset) = 0,$$

kde \emptyset označuje prázdnu množinu. Pre $N(A)$ taktiež platí, že sú spojité v tom zmysle, že ak A_n je klesajúca postupnosť uzavretých a ohraničených množín s limitou $\bigcap_n A_n = A$, potom musíme mať

$$N(A_n) \rightarrow N(A).$$

Tieto vlastnosti musia byť splnené pre každú realizáciu bodového procesu alebo aspoň s pravdepodobnosťou 1. Ony zabezpečujú, že N je miera (alebo prinajmenšom s pravdepodobnosťou 1, že hodnoty $N(A)$ môžu byť rozšírené na mieru). Toto je koncept náhodnej miery (Kallenberg (1983), Stoyan a kol. (1995)). Formálne teda bodový proces môže byť definovaný ako náhodná miera, ktorej hodnoty $N(A)$ sú nezáporné prirodzené čísla (Daley a Vere-Jones (1988), Stoyan a kol. (1995)). Zvyčajne tiež predpokladáme, že bodový proces je lokálne konečný, teda

$$N(A) < \infty$$

s pravdepodobnosťou 1 pre všetky ohraničené $A \subset \mathbb{R}^d$. Taktiež predpokladáme, že bodový proces je jednoduchý:

$$N(\{z\}) \leq 1 \quad \text{pre všetky } z \in \mathbb{R}^d$$

s pravdepodobnosťou 1. Inak povedané, požadujeme s pravdepodobnosťou 1, aby

žiadne dva body procesu neboli totožné. K jednoduchému procesu môžeme pristupovať ako k bodovému procesu \mathbf{Z} , považovanému za náhodnú množinu, alebo ako k sčítacej premennej N pre ten istý jednoduchý bodový proces.

Indikátor prázdnoty $V(A)$ množiny A , ktorý je rovný jednej v prípade, že množina A neobsahuje žiadne body, nesie kompletnú informáciu o jednoduchom bodovom procese. Ak poznáme hodnoty $V(A)$ pre všetky množiny A , tak vieme určiť presnú polohu každého bodu z jednoduchého bodového procesu \mathbf{Z} . Aby sme to vedeli urobiť potrebujeme množinu D , ktorá je zjednotením všetkých otvorených množín A takých, že $V(A) = 1$. Doplnok D je lokálne konečná množina bodov a určuje náhodnú množinu bodov \mathbf{Z} . Indikátory prázdnoty musia spĺňať

$$V(A \cup B) = \min\{V(A), V(B)\}$$

pre všetky množiny A, B a majú ďalšie analogické vlastnosti, ako tomu bolo pre sčítaciu premennú $N(A)$. Preto môžeme alternatívne definovať jednoduchý bodový proces ako náhodnú funkciu V , ktorá spĺňa požadované vlastnosti skoro iste.

1.2 Poissonov proces

Najskôr uvedieme popis pre jednorozmerný Poissonov proces. Uvažujeme nasledujúce predpoklady:

1. Počet prichádzajúcich bodov v danom časovom intervale má strednú hodnotu proporcionálnu k dĺžke intervalu

$$\mathbb{E}N(a, b] = \lambda(b - a),$$

kde λ je intenzita procesu;

2. príchody v disjunktných časových intervaloch sú nezávislé: ak $a_1 < b_1 < a_2 < b_2 < \dots < a_m < b_m$, potom $N(a_1, b_1], \dots, N(a_m, b_m]$ sú nezávislé;
3. pravdepodobnosť dvoch alebo viacerých príchodov v časovom intervale je asymptoticky nižšieho rádu ako dĺžka intervalu:

$$\mathbb{P}(N(a, a + h] \geq 2) = o(h), \quad h \downarrow 0.$$

Z uvedených predpokladov vyplýva, že počet bodov prichádzajúcich v danom časovom intervale má Poissonovo rozdelenie:

$$N(a, b] \sim \text{Poisson}(\lambda(b - a)).$$

Tento záver vychádza z rozdelenia intervalu $(a, b]$ na vysoký počet n malých intervalov. Počet príchodov v malých intervaloch dĺžky h je 0 alebo 1, až na udalosti s pravdepodobnosťou $o(h)$ pre $h \downarrow 0$. Keďže $N(a, b]$ je súčtom týchto počtov, má približne binomické rozdelenie. Pre $n \rightarrow \infty$ dostávame, že $N(a, b]$ musí mať Poissonovo rozdelenie.

Definícia 1. *Jednorozmerný Poissonov proces s konštantnou intenzitou $\lambda > 0$ definujeme ako bodový proces v \mathbb{R} , ktorý spĺňa podmienky:*

1. *pre každý interval $(a, b]$ má počet $N(a, b]$ Poissonovo rozdelenie so strednou hodnotou $\lambda(b - a)$;*
2. *ak $(a_1, b_1], \dots, (a_m, b_m]$ sú disjunktné ohraničené intervaly, potom súčty $N(a_1, b_1], \dots, N(a_m, b_m]$ sú nezávislé náhodné veličiny.*

Definícia 2. *Priestorový (dvojrozmerný) Poissonov proces s konštantnou intenzitou $\lambda > 0$ je bodový proces v \mathbb{R}^2 spĺňajúci:*

1. *pre každú ohraničenú uzavretú množinu A má súčet $N(A)$ Poissonovo rozdelenie so strednou hodnotou $\lambda|A|$;*
2. *ak A_1, \dots, A_m sú disjunktné množiny, potom $N(A_1), \dots, N(A_m)$ sú nezávislé.*

Výraz $|A|$ označuje Lebesgueovu mieru množiny A , čo v \mathbb{R}^2 znamená jej obsah.

Konštantná intenzita λ pre priestorový proces vyjadruje očakávaný počet bodov na jednotku plochy. Dôležitú vlastnosť Poissonovho procesu popisuje nasledujúca lemma.

Lemma 1 (Baddeley, 2007, str. 10). *Uvažujeme priestorový Poissonov proces v \mathbb{R}^2 s konštantnou intenzitou $\lambda > 0$. Nech $B \subset \mathbb{R}^2$ je ľubovoľná oblasť s $0 < |B| < \infty$. Ak máme dané $N(B) = n$, tak podmienené rozdelenie $N(A)$ pre $A \subseteq B$ je binomické*

$$\mathbb{P}(N(A) = k \mid N(B) = n) = \binom{n}{k} p^k (1 - p)^{n-k},$$

kde $p = |A|/|B|$. Navyše podmienené združené rozdelenie $N(A_1), \dots, N(A_m)$ pre ľubovoľné $A_1, \dots, A_m \subseteq B$ je rovnaké ako združené rozdelenie týchto premenných v binomickom procese. Inak povedané, ak máme daných n bodov Poissonovho procesu v B , tak tieto body sú podmienené nezávislé a rovnomerne rozdelené v B .

Dôkaz. Jednotlivé kroky dôkazu sú popísané v článku Baddeley (2007, str. 11). □

Podmienená vlastnosť nám priamo poskytuje spôsob, ako simulovať priestorový Poissonov proces. Realizáciu Poissonovho procesu s intenzitou λ v B vygenerujeme tak, že najskôr vygenerujeme náhodnú premennú M z Poissonovho rozdelenia s parametrom $\lambda|B|$. Potom pre dané $M = m$ generujeme m nezávislých rovnomerne rozdelených bodov v B .

1.3 Priestor výstupov

Bodový proces môžeme popísať aj definovaním priestoru možných výstupov a potom špecifikovať pravdepodobnosti jednotlivých udalostí. Priestor realizácií bodového procesu v \mathbb{R}^d je \mathbf{N} , množina všetkých sčítacích mier na \mathbb{R}^d , kde sčítacia miera je nezáporné prirodzené číslo, ktoré má konečnú hodnotu pre každú kompaktnú množinu. Základná udalosť pre bodový proces je, že v množine A je presne k bodov:

$$E_{A,k} = \{N \in \mathbf{N} : N(A) = k\}$$

pre kompaktnú množinu $A \subset \mathbb{R}^d$ a $k = 0, 1, 2, \dots$

Definícia 3. *Nech \mathbf{N} je množina všetkých sčítacích mier na \mathbb{R}^d a nech \mathcal{N} je σ -algebra podmnožín \mathbf{N} generovaných všetkými udalosťami formy $E_{A,k}$. Priestor \mathbf{N} so σ -algebrou \mathcal{N} sa nazýva priestor výstupov pre bodový proces v \mathbb{R}^d .*

V skutočnosti \mathcal{N} je najmenšia σ -algebra na \mathbf{N} , ktorá garantuje, že ak N je bodový proces na pravdepodobnostnom priestore $(\Omega, \mathcal{A}, \mathbb{P})$, potom premenné $N(A)$ pre každú kompaktnú množinu A sú náhodné premenné na tom istom priestore. Merateľnosť je požiadavka na každú udalosť $E \in \mathcal{N}$, že udalosť

$$\{N \in E\} = \{\omega \in \Omega : N^\omega \in E\}$$

patrí do \mathcal{A} . Bodový proces \mathbf{Z} teraz môžeme formálne definovať ako náhodnú sčítaciu mieru, tj. ako merateľné zobrazenie $\mathbf{Z} : \Omega \rightarrow \mathbf{N}$ z pravdepodobnostného priestoru $(\Omega, \mathcal{A}, \mathbb{P})$ do priestoru výstupov $(\mathbf{N}, \mathcal{N})$.

Definícia 4. *Rozdelenie bodového procesu \mathbf{Z} je pravdepodobnostná miera $\mathbf{P}_{\mathbf{Z}}$ na priestore výstupov $(\mathbf{N}, \mathcal{N})$ definovaná ako*

$$\mathbf{P}_{\mathbf{Z}}(A) = \mathbb{P}(\mathbf{Z} \in A), \quad A \in \mathcal{N}.$$

Jednotlivé body procesu môžu mať k nim prislúchajúcu dodatočnú informáciu, inak povedané aj príznak. Proces s príznakom môžeme vyjadriť ako dvojicu (z, m) , kde z je poloha bodu a m je jeho príznak.

Definícia 5. *Bodový proces s príznakom na priestore \mathbb{R}^d s príznakmi v priestore M je bodový proces \mathbf{Y} na $\mathbb{R}^d \times M$ taký, že $\mathbf{Y}(B \times M) < \infty$ skoro iste pre všetky kompaktné množiny $B \subset \mathbb{R}^d$. Teda korešpondujúci projektovaný proces bez príznakov je lokálne konečný.*

Veta 2 (Kingman, 1993). *Nech \mathbf{Y} je bodový proces s príznakom v \mathbb{R}^d , s príznakmi v M a $\boldsymbol{\eta}$ je projektovaný proces v \mathbb{R}^d bez príznakov. Potom nasledujúce tvrdenia sú ekvivalentné:*

1. *$\boldsymbol{\eta}$ je Poissonov proces v \mathbb{R}^d s intenzitou λ a pre dané $\boldsymbol{\eta}$ príznaky prislúchajúce jednotlivým bodom $\boldsymbol{\eta}$ sú nezávislé a rovnako rozdelené so spoločným rozdelením Q na M ;*
2. *\mathbf{Y} je Poissonov proces v $\mathbb{R}^d \times M$ s mierou intenzity $\lambda \otimes Q$.*

Dôkaz. Jednotlivé kroky dôkazu sú popísané napr. v článku Kingman (1993). □

Definícia 6. *Majme bodový proces \mathbf{Z} v \mathbb{R}^d . Výraz*

$$\nu(B) = \mathbb{E}[\mathbf{Z}(B)], \quad B \subset \mathbb{R}^d,$$

definuje mieru ν v \mathbb{R}^d a nazývame ho miera intenzity \mathbf{Z} za predpokladu, že $\nu(B) < \infty$ pre všetky kompaktné B .

Nech \mathbf{Y} je bodový proces s príznakom v zmysle definície 5. Rozšírime definíciu miery intenzity aj na proces \mathbf{Y} . Miera intenzity pre \mathbf{Y} je miera ν na $\mathbb{R}^d \times M$ definovaná ako

$$\nu(U) = \mathbb{E}\mathbf{Y}(U), \quad U \subset \mathbb{R}^d \times M.$$

Je úplne určená hodnotami

$$\begin{aligned} \nu(B \times C) &= \mathbb{E}\mathbf{Y}(B \times C) \\ &= \mathbb{E} \sum_{(z, m) \in \mathbf{Y}} \mathbf{1}\{z \in B\} \mathbf{1}\{m \in C\} \end{aligned}$$

pre všetky kompaktné $B \subset \mathbb{R}^d$ a merateľné $C \subset M$.

Definícia 7. Bodový proces s príznakom na \mathbb{R}^d s príznakmi v M je stacionárny, ak jeho rozdelenie je invariantné voči posunom len v \mathbb{R}^d

$$(v, m) \mapsto (v + w, m)$$

pre všetky $w \in \mathbb{R}^d$.

Veta 3 (Baddeley, 2007, str. 31). Nech \mathbf{Y} je stacionárny bodový proces s príznakom v \mathbb{R}^d . Predpokladáme, že korešpondujúci proces bodov bez príznakov má konečnú intenzitu (teda $\mathbb{E}\mathbf{Y}(B \times M) < \infty$ pre všetky kompaktné $B \subset \mathbb{R}^d$). Potom miera intenzity ν procesu \mathbf{Y} má tvar

$$\nu(B \times K) = \alpha|B|Q(K)$$

pre všetky $B \subset \mathbb{R}^d$, $K \subset M$, kde $\alpha \geq 0$ je očakávaný počet bodov na jednotku objemu a Q je pravdepodobnostná miera na M , ktorá sa nazýva rozdelenie typického príznaku.

Dôkaz. Jednotlivé kroky dôkazu sú podrobne popísané v článku Baddeley (2007, str. 31).



1.4 Konečné bodové procesy s hustotou

Definícia 8. Konečným bodovým procesom nazveme bodový proces \mathbf{X} na $B \subset \mathbb{R}^d$, pre ktorý je $N(B) < \infty$ skoro iste.

Uvažujme Poissonov proces na $B \subset \mathbb{R}^d$ s mierou intenzity λ , ktorý je konečný. Táto voľba je ekvivalentná výberu náhodného čísla $K \sim \text{Poisson}(\lambda(B))$ a potom pre dané $K = k$ generujeme k i.i.d náhodných bodov s rozdelením $W(A) = \lambda(A)/\lambda(B)$. Realizácie konečného bodového procesu \mathbf{X} patria do priestoru

$$\mathbf{N}^p = \{N \in \mathbf{N} : N(B) < \infty\}$$

konečných, jednoduchých sčítacích mier na B . Tento priestor môžeme rozložiť do podpriestorov podľa celkového počtu bodov:

$$\mathbf{N}^p = \mathbf{N}_0 \cup \mathbf{N}_1 \cup \mathbf{N}_2 \cup \dots,$$

kde pre každé $k = 0, 1, 2, \dots$

$$\mathbf{N}_k = \{N \in \mathbf{N} : N(B) = k\}$$

je množina všetkých sčítacích mier s celkovým množstvom k , ktoré sú množinou všetkých konfigurácií k bodov. Toto sa dá viac explicitne vyjadriť zavedením priestoru usporiadaných k -tíc

$$S^{!k} = \{(x_1, \dots, x_k) : x_i \in B, x_i \neq x_j \text{ pre všetky } i \neq j\}.$$

Definujeme mapovanie $I_k : S^{!k} \rightarrow \mathbf{N}_k$ ako

$$I_k(x_1, \dots, x_k) = \delta_{x_1} + \dots + \delta_{x_k},$$

kde δ_{x_i} je Diracova miera sústredená v bode x_i . To nám dáva

$$\mathbf{N}_k \equiv S^{!k} / \sim,$$

kde \sim je relácia ekvivalencie podľa permutácie, teda

$$(x_1, \dots, x_k) \sim (y_1, \dots, y_k) \Leftrightarrow \{x_1, \dots, x_k\} = \{y_1, \dots, y_k\}.$$

Typicky, λ je Lebesgueova miera na ohraničenej množine v \mathbb{R}^d . Označme symbolom P_η rozdelenie Poissonovho procesu η s mierou intenzity λ .

Definícia 9. *Nech $p : \mathbf{N}^p \rightarrow \mathbb{R}_+$ je merateľná funkcia spĺňajúca $\int_{\mathbf{N}} p(\mathbf{x}) dP_\eta(\mathbf{x}) = 1$. Definujeme*

$$\mathbf{P}(A) = \int_A p(\mathbf{x}) dP_\eta(\mathbf{x})$$

pre ľubovoľné $A \in \mathcal{N}$. Potom \mathbf{P} je rozdelenie bodového procesu v zmysle definície 4. Funkcia p je hustota bodového procesu s rozdelením \mathbf{P} (vzhľadom k Poissonovmu procesu η).

Lemma 4 (Baddeley, 2007, str. 62). *Pre bodový proces μ s hustotou p máme*

$$\mathbb{P}(\mu \in A) = e^{-\lambda(B)} \sum_{n=0}^{\infty} \frac{1}{n!} \int_B \dots \int_B \mathbf{1}\{I_n(x_1, \dots, x_n) \in A\} p(I_n(x_1, \dots, x_n)) \lambda(dx_1) \dots \lambda(dx_n) \quad (1.1)$$

pre každé $A \in \mathcal{N}$ a

$$\mathbb{E}[F(\boldsymbol{\mu})] = e^{-\lambda(B)} \sum_{n=0}^{\infty} \frac{1}{n!} \int_B \dots \int_B F \{I_n(x_1, \dots, x_n) \in A\} p(I_n(x_1, \dots, x_n)) \lambda(dx_1) \dots \lambda(dx_n) \quad (1.2)$$

pre ľubovoľnú integrovateľnú funkciu $F : \mathbf{N} \rightarrow \mathbb{R}_+$. Uvedené identity je možné prepísať ako

$$\mathbb{P}(\boldsymbol{\mu} \in A) = \mathbb{E}[p(\boldsymbol{\eta}) \mathbf{1}_A(\boldsymbol{\eta})], \quad (1.3)$$

$$\mathbb{E}[F(\boldsymbol{\mu})] = \mathbb{E}[F(\boldsymbol{\eta}) p(\boldsymbol{\eta})], \quad (1.4)$$

kde $\boldsymbol{\eta}$ je Poissonov proces s mierou intenzity λ .

Definícia 10. Nech $\boldsymbol{\mu}$ je bodový proces na \mathbb{R}^d , redukovaná Campbellova miera $\boldsymbol{\mu}$ je miera $C^!$ na $\mathbb{R}^d \times \mathbf{N}$ taká, že

$$C^![B \times A] = \mathbb{E} \left[\sum_{u \in \boldsymbol{\mu}} \mathbf{1}\{u \in B\} \mathbf{1}\{\boldsymbol{\mu} \setminus u \in A\} \right]$$

pre $B \subset \mathbb{R}^d$ a $A \in \mathcal{N}$. Predpokladajme, že redukovaná Campbellova miera $C^!$ procesu $\boldsymbol{\mu}$ je absolútne spojitá vzhľadom na $\lambda_d \otimes \mathbf{P}$ (λ_d je d -rozmerná Lebesgueova miera a \mathbf{P} rozdelenie $\boldsymbol{\mu}$). Potom Radon-Nikodýmova derivácia $\lambda^* : \mathbb{R}^d \times \mathcal{N} \rightarrow \mathbb{R}_+$ z $C^!$ vzhľadom na $\lambda_d \otimes \mathbf{P}$ sa nazýva podmienená intenzita $\boldsymbol{\mu}$. Je definovaná tak, aby platilo

$$C^![B \times A] = \int_B \mathbb{E}[\lambda^*(u, \boldsymbol{\mu}) \mathbf{1}\{\boldsymbol{\mu} \in A\}] du$$

pre $B \subset \mathbb{R}^d$ a $A \in \mathcal{N}$.

Uvažujme konečný bodový proces $\boldsymbol{\mu}$ na kompaktnej množine $B \subset \mathbb{R}^d$. Ak existuje podmienená intenzita $\lambda^*(y, \boldsymbol{\mu})$, tak spĺňa

$$\mathbb{E} \sum_{u \in \boldsymbol{\mu}} F(u, \boldsymbol{\mu} \setminus u) = \int_{\mathbb{R}^d} \mathbb{E}[\lambda^*(y, \boldsymbol{\mu}) F(y, \boldsymbol{\mu})] dy \quad (1.5)$$

pre ľubovoľnú integrovateľnú F .

Nech $\boldsymbol{\mu}$ má hustotu pravdepodobnosti $p(\boldsymbol{\mu})$ vzhľadom na konštantný Poissonov proces $\boldsymbol{\eta}$ s intenzitou jedna na B . Potom stredná hodnota hocijakej integrovateľnej funkcie $h(\boldsymbol{\mu})$ sa dá zapísať vo forme (1.2) alebo (1.3). Použitím tohto na obe strany (1.5) dostaneme

$$\mathbb{E} \left[p(\boldsymbol{\eta}) \sum_{u \in \boldsymbol{\eta}} F(u, \boldsymbol{\eta} \setminus u) \right] = \int_{\mathbb{R}^d} \mathbb{E}[\lambda^*(y, \boldsymbol{\eta}) p(\boldsymbol{\eta}) F(y, \boldsymbol{\eta})] dy. \quad (1.6)$$

Ak položíme

$$h(y, \boldsymbol{\mu}) = p(\boldsymbol{\mu} \cup \{y\}) F(y, \boldsymbol{\mu}),$$

tak ľavá strana rovnice (1.6) sa dá prepísať na

$$\mathbb{E} \left[p(\boldsymbol{\eta}) \sum_{u \in \boldsymbol{\eta}} F(u, \boldsymbol{\eta} \setminus u) \right] = \mathbb{E} \left[\sum_{u \in \boldsymbol{\eta}} h(u, \boldsymbol{\eta} \setminus u) \right] = \int_B \mathbb{E} [h(y, \boldsymbol{\mu})] dy,$$

kde posledná rovnosť vznikla použitím rovnice (1.6) na proces $\boldsymbol{\eta}$, keďže podmienená intenzita je na B identicky rovná jednej. Preto dostaneme

$$\int_{\mathbb{R}^d} \mathbb{E} [\lambda^*(y, \boldsymbol{\eta}) p(\boldsymbol{\eta}) F(y, \boldsymbol{\eta})] dy = \int_B \mathbb{E} [p(\boldsymbol{\eta} \cup \{y\}) F(y, \boldsymbol{\eta})] dy$$

pre ľubovoľnú funkciu F . Z toho vyplýva, že

$$\lambda^*(y, \boldsymbol{\eta}) p(\boldsymbol{\eta}) = p(\boldsymbol{\eta} \cup y)$$

skoro iste, pre skoro všetky $y \in B$. Tento výsledok formuluje nasledujúca veta.

Veta 5 (Baddeley, 2007, str. 65). *Nech p je hustota pravdepodobnosti konečného bodového procesu $\boldsymbol{\mu}$ na ohraničenej oblasti $B \subset \mathbb{R}^d$. Predpokladáme, že*

$$p(\mathbf{x}) > 0 \quad \Rightarrow \quad p(\tilde{\mathbf{x}}) > 0, \quad \text{pre všetky } \tilde{\mathbf{x}} \subset \mathbf{x}.$$

Potom podmienená intenzita $\boldsymbol{\mu}$ existuje a je rovná

$$\lambda^*(y, \mathbf{x}) = \frac{p(\mathbf{x} \cup y)}{p(\mathbf{x})}, \quad \mathbf{x} \in \mathbf{N}, y \in B, y \notin \mathbf{x}, \quad (1.7)$$

skoro všade.

Lemma 6 (Baddeley, 2007, str. 66). *Uvažujme konečný bodový proces $\boldsymbol{\mu}$ na ohraničenej oblasti B v \mathbb{R}^d . Predpokladáme, že $\boldsymbol{\mu}$ má hustotu p a podmienenú intenzitu λ^* . Potom p je úplne určená λ^* .*

Dôkaz. Jednotlivé kroky dôkazu sú popísané v článku Baddeley (2007, str. 66). □

Poznámka. Bodový proces $\boldsymbol{\mu}$ patrí do triedy Gibbsových procesov.

1.5 U -štatistiky

U -štatistika rádu $k \in \mathbb{N}$ konečného bodového procesu $\boldsymbol{\mu}$ je funkcionál definovaný ako

$$F(\boldsymbol{\mu}) = \sum_{(x_1, \dots, x_k) \in \mu_{\neq}^k} f(x_1, \dots, x_k), \quad (1.8)$$

kde $f : B^k \rightarrow \mathbb{R}$ je symetrická funkcia vzhľadom na permutácie jej premenných a $f \in L_1(\lambda^k)$. V tomto značení je μ_{\neq}^k množina k -tíc rozličných bodov procesu $\boldsymbol{\mu}$. Hovoríme, že F je vytvorená f . Zo Slivnyak-Mecke vety (Schneider a Weil, 2008) máme

$$\mathbb{E}F(\boldsymbol{\eta}) = \int_B \dots \int_B f(x_1, \dots, x_k) \lambda(d(x_1, \dots, x_k)),$$

pričom označením $\lambda(d(x_1, \dots, x_k))$ myslíme $\lambda(dx_1) \dots \lambda(dx_k)$.

Pre konečný bodový proces $\boldsymbol{\mu}$ uvažujeme podmienenú intenzitu z (1.7). Pravdepodobnosť $\mathbb{P}(y \in \boldsymbol{\mu}) = 0$ a pre $p(\mathbf{x}) = 0$ položíme $\lambda^*(y, \mathbf{x}) = 0$. Ak $n > 1$ použijeme analogicky vzťah platný skoro iste

$$\lambda_n^*(y_1, \dots, y_n, \mathbf{x}) = \frac{p(\mathbf{x} \cup \{y_1, \dots, y_n\})}{p(\mathbf{x})}$$

pre $y_1, \dots, y_n \in B$. Potom λ_n^* je jednoznačná podmienená intenzita n -tého rádu procesu $\boldsymbol{\mu}$, $\lambda_0^* \equiv 1$. Všimnime si, že λ_n^* je symetrická v premenných y_1, \dots, y_n .

Definícia 11. *Stredná hodnota podmienenej intenzity λ^* bodového procesu $\boldsymbol{\mu}$ je korelačná funkcia*

$$\rho(y, \boldsymbol{\mu}) = \mathbb{E}\lambda^*(y, \boldsymbol{\mu}). \quad (1.9)$$

Zároveň definujeme korelačnú funkciu n -tého rádu ako

$$\rho_n(y_1, \dots, y_n, \boldsymbol{\mu}) = \mathbb{E}(\lambda_n^*(y_1, \dots, y_n, \boldsymbol{\mu})).$$

Veta 7 (Beneš a Zikmundová, 2014, str. 900). *Pre U -štatistiku $F \in L_2(P_{\boldsymbol{\eta}})$ rádu k a hustotu $p \in L_2(P_{\boldsymbol{\eta}})$ máme*

$$\mathbb{E}F(\boldsymbol{\mu}) = \int_{B^k} f(x_1, \dots, x_k) \mathbb{E}[\lambda_k^*(x_1, \dots, x_k, \boldsymbol{\mu})] \lambda(d(x_1, \dots, x_k)). \quad (1.10)$$

Dôkaz. Jednotlivé kroky dôkazu sú podrobne popísané v článku Beneš a Zikmundová (2014, str. 900). □

Nech $B \subset \mathbb{R}^d$, $d \in \mathbb{N}$ je ohraničená Borelovská množina s kladnou Lebesgueovou mierou, \mathbf{X} je konečný bodový proces na B s príznakom, pričom príznak má

geometrický význam. Dvojica (v, m) , kde v je bod a m je príznak, tvorí geometrický útvar (napr. v je stred a m dĺžka úsečky alebo v je stred a m polomer kruhu). Symbolom U_x označíme zjednotenie všetkých útvarov realizácie procesu x . Uvažujeme hustotu pravdepodobnosti (Møller a Helisová, 2008)

$$p(x) = c_\theta^{-1} \exp(\theta G(U_x)), \quad (1.11)$$

procesu \mathbf{X} vzhľadom na Poissonov bodový proces $\boldsymbol{\eta}$. Máme $\theta = (\theta_1, \dots, \theta_l)$ je vektor reálnych parametrov, c_θ je normalizačná konštanta a $G(U_x) \in \mathbb{R}^l$ je vektor geometrických charakteristík U_x . Exponent v rovnici (1.11) obsahuje skalárny súčin v \mathbb{R}^l . Najväčšia množina θ takých, aby bola rodina exponenciálnych hustôt dobre definovaná je $\{\theta \in \mathbb{R}^l : \mathbb{E}[\exp(\theta G(U_\eta))] < \infty\}$, viď (Møller a Helisová, 2008).

Definícia 12. Operátor diferencie $D_y F$ pre funkcionál F bodového procesu $\boldsymbol{\mu}$ definujeme ako náhodnú premennú

$$D_y F(\boldsymbol{\mu}) = F(\boldsymbol{\mu} + \delta_y) - F(\boldsymbol{\mu})$$

pre $y \in B$, kde δ_y je Diracova miera sústredená v bode y . Indukciou pre $m \geq 2$ a $(y_1, \dots, y_m) \in B^m$ máme funkciu

$$D_{y_1, \dots, y_m}^m F = D_{y_1}^1 D_{y_2, \dots, y_m}^{m-1} F,$$

kde $D_y^1 = D_y$ a $D^0 F = F$.

Pre vektor geometrických charakteristík $G(U_x) = (G_1(U_x), \dots, G_t(U_x))$, $t \in \mathbb{N}$ označme vektor m -tých diferencií nasledovne

$$D_{y_1, \dots, y_m}^m G(U_x) = (D_{y_1, \dots, y_m}^m G_1(U_x), \dots, D_{y_1, \dots, y_m}^m G_t(U_x))^T.$$

Veta 8 (Beneš a Zikmundová, 2014, str. 906). Uvažujeme hustotu pravdepodobnosti z (1.11). Potom zodpovedajúca podmienená intenzita λ_m^* rádu $m \in \mathbb{N}$ pre $x \in \mathbf{N}$ splňa

$$\lambda_m^*(y_m, \dots, y_1, x) = \exp(\theta Q_m G(U_x)) \quad \text{skoro iste,}$$

kde

$$Q_m G(U_x) = D_{y_1, \dots, y_m}^m G(U_x) + \sum_{i_1, \dots, i_{m-1} \in \{1, \dots, m\}} D_{y_1, \dots, y_{i_{m-1}}}^{m-1} G(U_x) + \dots + \sum_{1 \leq i \leq m} D_{y_i} G(U_x),$$

a pričom indexy v sumách musia byť rozdielne.

Dôkaz. Jednotlivé kroky dôkazu sú podrobne popísané v článku Beneš a Zikmundová (2014, str. 906).



Kapitola 2

Model úsečiek s interakciami

V tejto kapitole sa budeme zaoberať procesom úsečiek s interakciami v \mathbb{R}^2 . Uvažujme najskôr ohraničenú množinu $B \subset \mathbb{R}^2$,

$$Y = B \times (0, b] \times [-\pi/2, \pi/2), \quad (2.1)$$

kde $(0, b]$ je interval dĺžok úsečiek. Poissonov proces η na Y má mieru intenzity λ ,

$$\lambda(d(v, r, \phi)) = \chi(v) dv Q(dr) V(d\phi), \quad (2.2)$$

pričom v označuje polohu stredu úsečky, r je dĺžka úsečky a ϕ orientácia vzhľadom na horizontálnu os. Ďalej Q, V sú pravdepodobnostné miery, V nedegenerovaná a χ je ohraničená funkcia intenzity bodov na B . Proces úsečiek μ má hustotu (1.11) s $\theta = (\theta_1, \theta_2)$ a predpokladáme $\theta_2 \leq 0$, aby sme zabezpečili, že p je hustota. Okrem toho

$$G(U_x) = (L(U_x), N(U_x)), \quad (2.3)$$

kde $L(U_x)$ je celková dĺžka všetkých úsečiek realizácie x a $N(U_x)$ je celkový počet priesečníkov medzi úsečkami v x . Preto ak L je dĺžka jednej úsečky, tak

$$L(U_\mu) = \sum_{u \in \mu} L(u) \quad (2.4)$$

je U -štatistika prvého rádu a

$$N(U_\mu) = \frac{1}{2} \sum_{(u, v) \in \mu_\neq^2} \mathbf{1}_{[u \cap v \neq \emptyset]} \quad (2.5)$$

je U -štatistika druhého rádu.

Uvažujme proces úsečiek μ z \mathbb{R}^2 s U -štatistikami v (2.3). Chceme vyjadriť

podmienenú hustotu procesu. Pre $m \in \mathbb{N}$ je

$$Q_m = \sum_{i=1}^m D_{y_i} G(U_\mu) + \sum_{1 \leq i < j \leq m} D_{y_i, y_j}^2 G(U_\mu),$$

kde jednotlivé diferencie majú pre $y, y_i \in Y$ (2.1), $y, y_i \notin x, x \in \mathbf{N}$ nasledujúci tvar

$$D_y G(U_x) = \begin{pmatrix} L(y) \\ \sum_{u \in x} \mathbf{1}_{[u \cap y \neq \emptyset]} \end{pmatrix}, D_{y_1, y_2}^2 G(U_x) = \begin{pmatrix} 0 \\ \mathbf{1}_{[y_1 \cap y_2 \neq \emptyset]} \end{pmatrix}. \quad (2.6)$$

Z vety 8 dostávame

$$\lambda^*(y, x) = \frac{p(x \cup \{y\})}{p(x)} = \exp(\theta_1 L(y) + \theta_2 N_x(y)),$$

$$\lambda_2^*(y_1, y_2, x) = \exp(\theta_1 L(y_1 \cup y_2) + \theta_2 (N_x(y_1) + N_x(y_2) + N(y_1 \cap y_2))),$$

kde pod označením $L(y_i \cup y_j)$ rozumieme súčet dĺžok segmentov y_i a y_j , ďalej $N_x(y_i)$ vyjadruje počet priesečníkov úsečky y_i so systémom úsečiek x a $N(y_i \cap y_j)$ má hodnotu jedna, ak úsečky y_i, y_j majú prienik, nula inak.

Dôsledok 1 (Beneš a Zikmundová, 2014, str. 910). *Nech μ je proces úsečiek na Y (2.1) s hustotou (1.11), $\theta_2 \leq 0$, potom pre U -štatistiky (2.4) a (2.5) máme*

$$\begin{aligned} \mathbb{E}L(U_\mu) &= \int_Y \rho(y, \mu) L(y) \lambda(dy), \\ \mathbb{E}N(U_\mu) &= \frac{1}{2} \int_{Y^2} \rho_2(y_1, y_2, \mu) \mathbf{1}_{[y_1 \cap y_2 \neq \emptyset]} \lambda(d(y_1, y_2)), \\ \mathbb{E}[L(U_\mu)^2] &= \int_Y \rho(y, \mu) L(y)^2 \lambda(dy) \\ &\quad + \int_{Y^2} \rho_2(y_1, y_2, \mu) L(y_1) L(y_2) \lambda(d(y_1, y_2)). \end{aligned} \quad (2.7)$$

Predpoklad $\theta_2 < 0$ zodpovedá odpudivým interakciám medzi úsečkami, ak $\theta_2 = 0$ jedná sa o Poissonov proces.

2.1 Simulácia modelu úsečiek

Realizáciu procesu úsečiek s interakciami môžeme simulovať podľa Metropolis-Hastings algoritmu rodenia a zániku. Uvažujeme proces úsečiek μ zadaný v úvode tejto kapitoly. Nech U_x je zjednotenie úsečiek realizácie x . Pre úsečku $y = (v, r, \phi)$ a konfiguráciu bodov x definujeme Hastingsov pomer z (Geyer a Møller, 1994) ako

$$H(x, y) = \lambda_1^*(y, x) \frac{\lambda(Y)}{(n(x) + 1)},$$

kde $n(x)$ je počet úsečiek v realizácii x . Ďalej v (2.2) položíme $\chi(v) \equiv z$, pričom $z > 0$ je parameter. Potom máme $\lambda = zLeb \otimes Q \otimes V$. Po dosadení dostávame vzťah

$$H(x, y) = \frac{z|B|}{n(x) + 1} \exp(\theta_1 L(y) + \theta_2 N_x(y)), \quad (2.8)$$

kde $|B| = Leb(B)$.

Ak $x^{(iter)}$ je stav v iterácii $iter$, tak generujeme návrh, ktorým je buď "narodenie" $x^{(iter)} \cup \{y\}$ novej úsečky y s polohou stredy c , dĺžkou r a orientáciou ϕ , alebo "zánik" $x^{(iter)} \setminus \{x_i\}$ existujúcej úsečky $x_i = (c_i, r_i, \phi_i) \in x^{(iter)}$. Keď navrhujeme narodenie c, r, ϕ sú nezávislé, c je rovnomerne náhodne vybraný bod z B , r , resp. ϕ sú generované z rozdelenia Q , resp. V . V prípade, že navrhujeme zánik, x_i je rovnomerne náhodne vybraná úsečka z $x^{(iter)}$. Narodenie aj zánik nastávajú pre každý bod s rovnakou pravdepodobnosťou a to $1/2$. Prijatie návrhu závisí na Hastingsových pomeroch $H(x^{(iter)}, y)$ a $H(x^{(iter)} \setminus \{x_i\}, x_i)$.

Naznačme teraz jednoduchú schému algoritmu:

1. Uvažujeme $x^{(0)}$ ako východziu konfiguráciu o 1 úsečke z Y (2.1).
2. nech $x^{(iter)}$ je konfigurácia bodov, s ktorou pracujeme v iterácii $iter$.
3. S pravdepodobnosťou $1/2$ navrhujeme prídanie úsečky y a s pravdepodobnosťou

$$\min(1, H(x^{(iter)}, y))$$

tento návrh aj prijmeme, inak nastavíme $x^{(iter+1)} = x^{(iter)}$.

4. Inak navrhujeme zánik jednej úsečky x_i a s pravdepodobnosťou

$$\min(1, 1/H(x^{(iter)} \setminus \{x_i\}, x_i))$$

je návrh prijatý. V opačnom prípade $x^{(iter+1)} = x^{(iter)}$.

5. Po skončení požadovaného množstva iterácií $ITER$ máme nasimulovanú realizáciu $x = x^{(ITER)}$.

Postup pre simuláciu modelu úsečiek popísaný v tejto podkapitole sme prebrali z práce (Zikmundová, 2014).

Kapitola 3

Štatistika procesov úsečiek

V tejto kapitole uvádzame dve metódy odhadu parametrov (θ_1, θ_2, z) zavedeného procesu úsečiek. Tieto metódy budú aplikované na simulované aj reálne dáta. Prvá je momentová metóda navrhnutá na základe výsledkov z článku (Beneš a Zikmundová, 2014). Druhá je známa Takacs-Fixelova metóda, ktorá bola použitá na proces kruhov v (Dereudre a kol., 2014). My ju modifikujeme na proces úsečiek.

3.1 Momentová metóda odhadu parametrov (MM)

Majme jednu konkrétnu realizáciu procesu úsečiek x , ktorú považujeme za známu. To sú pre nás podkladové dáta, pričom úsečky sú dané ako body (stredy úsečiek) s polohou v okne B , dĺžky a orientácie prislúchajúce ku každému stredy, aby sme vedeli skonštruovať úsečku. Na základe známej realizácie x aproximujeme rozdelenie dĺžok a orientácií pomocou pravdepodobnostného histogramu. Nech počet stĺpcov histogramu je S , potom

$$t = \frac{\max(\text{dĺžka}) - \min(\text{dĺžka})}{S}$$

je rozsah každého stĺpca histogramu. Hranice jednotlivých stĺpcov histogramu sú v $(\min(\text{dĺžka}), \min(\text{dĺžka}) + t, \dots, \max(\text{dĺžka}))$. Spočítame pravdepodobnosti, že dĺžky úsečiek realizácie x patria daným intervalom. Výška stĺpcov histogramu je potom určená pravdepodobnosťou, že dĺžka úsečky realizácie x patrí rozsahu stĺpčeka. Obdobne postupujeme pri určení rozdelenia orientácií. Tieto pravdepodobnostné histogramy určujú aproximované rozdelenia Q pre dĺžky a V pre orientácie úsečiek, ktoré tým pádom považujeme za známe.

Pre účely odhadu parametrov z rovníc (2.7) odhadneme ľavé strany z realizácie x a integrály vpravo nahradíme sumou cez náhodný výber úsečiek $\{y_j\}_{j=1}^J = \{c_i, r_i, \phi_i\}_{j=1}^J$ z aproximovaného rozdelenia $Leb \otimes Q \otimes V$ spočítaného z reali-

zácie x . Jedná sa o metódu Monte Carlo založenú na J nezávislých stredoch úsečiek c_1, \dots, c_J rovnomerne rozdelených na B a J nezávislých realizáciách r_1, \dots, r_J , resp. ϕ_1, \dots, ϕ_J z aproximovaných rozdelení Q , resp. V . Dĺžky a orientácie generujeme z pravdepodobnostných histogramov tak, že najskôr náhodne zvolíme jeden zo stĺpcov histogramu, pričom každý z nich má šancu zodpovedajúcu pravdepodobnosti, že sa tam nachádza dĺžka úsečky realizácie x . V rámci stĺpca už dĺžku úsečky volíme ako rovnomerne náhodné číslo na intervale (\lceil dolná hranica stĺpca \rceil, \lfloor horná hranica stĺpca \rfloor).

Teraz pomocou uvedenej aproximácie prepíšeme rovnice (2.7) a zavedieme veličiny C_i^{MM} pre $i = 1, 2, 3$ nasledovne:

$$\begin{aligned} C_1^{MM} &= L(U_x) - \frac{|B|z}{J} \sum_{j=1}^J L(y_j) \exp(\theta_1 L(y_j) + \theta_2 N_x(y_j)), \\ C_2^{MM} &= N(U_x) - \frac{|B|^2 z^2}{J(J-1)} \sum_{j<l}^J \sum_{j<l}^J N(y_j \cap y_l) \\ &\quad \exp(\theta_1 L(y_j \cup y_l) + \theta_2 (N(y_j \cap y_l) + N_x(y_j) + N_x(y_l))), \\ C_3^{MM} &= L(U_x)^2 - \frac{|B|z}{J} \sum_{j=1}^J L(y_j)^2 \exp(\theta_1 L(y_j) + \theta_2 N_x(y_j)) - \\ &\quad - \frac{|B|^2 z^2}{J(J-1)} \sum_{j \neq l}^J \sum_{j \neq l}^J L(y_j) L(y_l) \\ &\quad \exp(\theta_1 L(y_j \cup y_l) + \theta_2 (N(y_j \cap y_l) + N_x(y_j) + N_x(y_l))). \end{aligned}$$

Samotné odhady parametrov momentovou metódou potom dostaneme minimalizáciou

$$(\hat{z}, \hat{\theta}) := \operatorname{argmin}_{z, \theta} \sum_{k=1}^3 (C_k^{MM})^2,$$

za podmienky $z > 0$ a $\theta_2 \leq 0$.

3.2 Metóda Takacs-Fiksel (T-F)

Nech μ je proces segmentov na množine Y (2.1), pričom každý segment je určený jeho stredom $z \in B \subset \mathbb{R}^2$, dĺžkou $r \in (0, b]$ a orientáciou $\phi \in [-\pi/2, \pi/2]$. Pre ľubovoľnú realizáciu x procesu μ je U_x zjednotením všetkých úsečiek danej realizácie x . Ďalej, v úvode tejto kapitoly uvádzame pravdepodobnostné miery Q, V , ktoré určujú rozdelenie dĺžok a orientácií. Proces μ s parametrami θ a $z > 0$ je pravdepodobnostná miera $\mathbf{P}^{z, \theta}$ na množine konfigurácií bodov z Y (2.1), ktorá je absolútne spojitá vzhľadom na Poissonov proces s príznakom η .

Definujeme lokálnu energiu jedného segmentu y vzhľadom na realizáciu x nasledujúcim vzťahom

$$h^\theta(y, x) := \theta D_y G(U_x) = \theta_1 L(y) + \theta_2 N_x(y),$$

kde $D_y G(U_x)$ je zo vzorca (2.6).

Nasledujúca veta charakterizuje proces úsečiek prostredníctvom tzv. GNZ rovnice.

Veta 9 (Georgii, 1976, Nguyen a Zessin, 1979). *Pre ľubovoľnú ohraničenú množinu $B \subset \mathbb{R}^2$, pravdepodobnostná miera \mathbf{P} na množine realizácií \mathbf{N} je bodový proces na Y z (2.1) s parametrom θ , intenzitou $z > 0$ a rozdelením Q, V dĺžok a orientácií práve vtedy, keď pre hocikáku nezápornú funkciu $f : Y \times \mathbf{N} \rightarrow \mathbb{R}$*

$$\mathbb{E} \left(\sum_{u \in \mu} f(u, \mu \setminus u) \right) = \mathbb{E} \left(\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_0^b \int_B z \exp(h^\theta(y, \mu)) f(y, \mu) dv Q(dr) V(d\phi) \right), \quad (3.1)$$

kde $y = (v, r, \phi)$, \mathbb{E} je stredná hodnota vzhľadom na \mathbf{P} a μ má rozdelenie \mathbf{P} .

Rovnica (3.1) je zároveň východiskom pre metódu Takacs-Fiksel na odhad parametrov (θ, z) .

Uvedieme metódu odhadu Takacs-Fiksel, ktorá závisí na voľbe testových funkcií. Majme pre nejakú nezápornú funkciu $f : Y \times \mathbf{N} \rightarrow \mathbb{R}$ a ľubovoľné $z > 0$, $\theta = (\theta_1 \in \mathbb{R}, \theta_2 \leq 0)$ náhodnú premennú

$$C_B^{z, \theta}(\mu; f) = \sum_{u \in \mu} f(u, \mu \setminus u) - \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_0^b \int_B z \exp(h^\theta(y, \mu)) f(y, \mu) dv Q(dr) V(d\phi), \quad (3.2)$$

kde μ má rozdelenie $\mathbf{P}^{z^*, \theta^*}$. Pod označením z^*, θ^* myslíme skutočné parametre procesu μ . Z ergodickej vety, ak B je dostatočne veľká, tak $C_B^{z, \theta}(\mu; f) / |B|$ je približne rovné $\mathbb{E}_{z^*, \theta^*} \left(C_{[0, 1]^2}^{z, \theta}(\mu; f) \right)$, kde $\mathbb{E}_{z^*, \theta^*}$ je stredná hodnota vzhľadom na $\mathbf{P}^{z^*, \theta^*}$. Navyše, z GNZ rovnice (3.1) sa dá ukázať, že

$\mathbb{E}_{z^*, \theta^*} \left(C_{[0, 1]^2}^{z, \theta}(\mu; f) \right) = 0$. Preto pre nejakú funkciu f by náhodná premenná $C_B^{z, \theta}(\mu; f)$ mala byť blízko k nule, keď $z = z^*$ a $\theta = \theta^*$. Táto skutočnosť tvorí základ pre metódu odhadu T-F.

Pre daných K funkcií $(f_k)_{1 \leq k \leq K}$ je T-F odhad definovaný jednoducho ako

$$\left(\hat{z}, \hat{\theta} \right) := \operatorname{argmin}_{z, \theta} \sum_{k=1}^K \left(C_B^{z, \theta}(x; f_k) \right)^2, \quad (3.3)$$

kde x je realizácia μ a K stačí pre proces úsečiek rovné trom. Použijeme tzv.

pseudovierohodný odhad, ktorý je osobitným prípadom T-F procedúry, kde

$$f_k(y, \mathbf{x}) = \frac{\partial h^\theta(y, \mathbf{x})}{\partial \theta_k} \quad k = 1, 2,$$

$$f_1(y, \mathbf{x}) = L(y)$$

$$f_2(y, \mathbf{x}) = N_{\mathbf{x}}(y)$$

a posledná testová funkcia je

$$f_3(y, \mathbf{x}) = \frac{1}{|B|z}.$$

Aby sme získali odhady jednotlivých parametrov podľa (3.3) potrebujeme napočítať veličiny z (3.2) pre všetky tri testové funkcie f_i .

Integrálovú časť v rovnici (3.2) môžeme aproximovať metódou Monte Carlo rovnako, ako tomu bolo pri momentovej metóde v podkapitole 3.1. Tým získame aproximáciu

$$C_B^{z, \theta}(\mathbf{x}; f) \approx \sum_{u \in \mathbf{x}} f(u, \mathbf{x} \setminus u) - \frac{|B|z}{J} \sum_{i=1}^J \exp(h^\theta(y_i, \mathbf{x})) f(y_i, \mathbf{x}). \quad (3.4)$$

Prepísaním pre jednotlivé f_i dostávame nasledujúce vzťahy:

$$\begin{aligned} C_B^{z, \theta}(\mathbf{x}, f_1) &= \sum_{u \in \mathbf{x}} f_1(u, \mathbf{x} \setminus u) - \\ &- \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_0^b \int_B z \exp(h^\theta(y, \mathbf{x})) f_1(y, \mathbf{x}) dy Q(dr) V(d\phi) \doteq \\ &\doteq \sum_{u \in \mathbf{x}} L(u) - \frac{|B|z}{J} \sum_{i=1}^J \exp(h^\theta(y_i, \mathbf{x})) f_1(y_i, \mathbf{x}) = \\ &= \sum_{u \in \mathbf{x}} L(u) - \frac{|B|z}{J} \sum_{i=1}^J \exp(\theta_1 L(y_i) + \theta_2 N_{\mathbf{x}}(y_i)) L(y_i), \end{aligned}$$

$$\begin{aligned} C_B^{z, \theta}(\mathbf{x}, f_2) &= \sum_{u \in \mathbf{x}} f_2(u, \mathbf{x} \setminus u) - \\ &- \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_0^b \int_B z \exp(h^\theta(y, \mathbf{x})) f_2(y, \mathbf{x}) dy Q(dr) V(d\phi) \doteq \\ &\doteq \sum_{u \in \mathbf{x}} N_{\mathbf{x} \setminus u}(u) - \frac{|B|z}{J} \sum_{i=1}^J \exp(h^\theta(y_i, \mathbf{x})) f_2(y_i, \mathbf{x}) = \\ &= \sum_{u \in \mathbf{x}} N_{\mathbf{x} \setminus u}(u) - \frac{|B|z}{J} \sum_{i=1}^J \exp(\theta_1 L(y_i) + \theta_2 N_{\mathbf{x}}(y_i)) N_{\mathbf{x}}(y_i), \end{aligned}$$

$$\begin{aligned}
C_B^{z,\theta}(\mathbf{x}, f_3) &= \sum_{u \in \mathbf{x}} f_3(u, \mathbf{x} \setminus u) - \\
&- \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_0^b \int_B z \exp(h^\theta(y, \mathbf{x})) f_3(y, \mathbf{x}) dy Q(dr) V(d\phi) \doteq \\
&\doteq \sum_{u \in \mathbf{x}} \frac{1}{|B|z} - \frac{|B|z}{J} \sum_{i=1}^J \exp(h^\theta(y_i, \mathbf{x})) f_3(y_i, \mathbf{x}) = \\
&= \frac{n(\mathbf{x})}{|B|z} - \frac{1}{J} \sum_{i=1}^J \exp(\theta_1 L(y_i) + \theta_2 N_{\mathbf{x}}(y_i)).
\end{aligned}$$

3.3 Porovnávacie štatistiky

Po zedefinovaní metód na odhad parametrov procesu úsečiek s interakciami vzniká otázka, ako jednotlivé metódy efektívne porovnať, resp. ako preukázať zhodu modelu s dátami. Všeobecne v priestorovej štatistike vzniká pri testovaní zhody viacero problémov. Takmer výhradne sa používajú testy Monte Carlo založené na simulovaných dátach s odhadnutými parametrami. V tejto práci si nekladíme za cieľ skonštruovať presný štatistický test dobrej zhody ale obmedzíme sa na popisnú štatistiku realizácií procesu úsečiek s prihliadnutím na použitý model.

Momentovú a T-F metódu môžeme porovnať v zmysle, ako dobre sa podobajú jednotlivé odhadnuté parametre, čo ale nemusí byť najlepším ukazovateľom zhody, keďže minimalizujeme funkciu o troch premenných a tie sa navzájom ovplyvňujú. Iný pohľad, a tým sa budeme zaoberať najviac je, že porovnáme vlastnosti výsledných realizácií na simulovaných dátach. K tomu využívame medzivýsledky z jednotlivých iterácií pri procese rodenia a zániku definovanom v podkapitole (2.1).

Teraz uvedieme niekoľko špecifických porovnaní:

1. grafické zobrazenie výslednej realizácie úsečiek po poslednej iterácii, respektíve východzí systém úsečiek
2. graf vývoja celkového počtu, celkovej dĺžky a celkového počtu priesečníkov úsečiek v priebehu simulácie
3. číselné charakteristiky z bodu 2. v podobe priemerov a smerodajných odchýlok spočítaných zo stavov z každej tisícej iterácie
4. histogramy z grafov vývoja z bodu 2., teda zastúpenie konkrétnych hodnôt pre celkové počty úsečiek počas simulácie (rovnako pre celkovú dĺžku a celkový počet priesečníkov)

5. histogramy rozdelenia dĺžok a orientácií konfigurácie po 60-tisíc, 80-tisíc a 100-tisíc iteráciách
6. histogram počtu priesečníkov jednotlivých úsečiek s konfiguráciou vo výslednom systéme simulácie.

Kapitola 4

Dáta a popis kódu

4.1 Reálne dáta

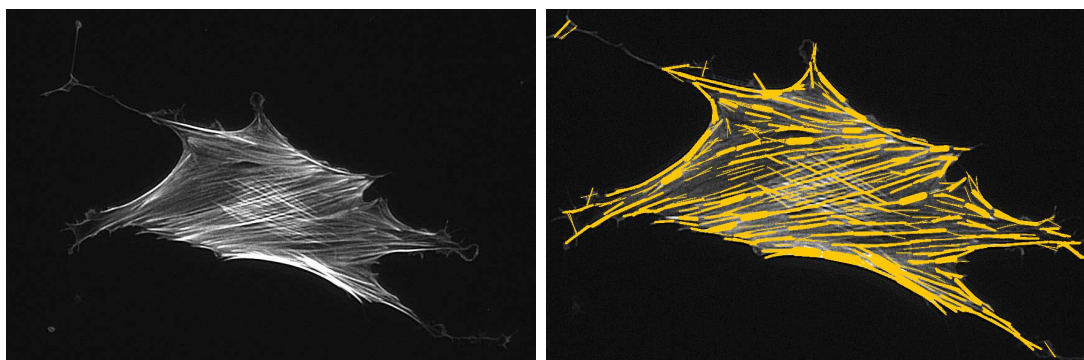
V tejto podkapitole popíšeme charakter reálnych dát, čo znamenajú a ako boli získané. Nasledujúci popis podkladových dát aj motiváciu pre ich skúmanie preberáme z článku (Eltzner a kol., 2014). Realizáciou procesu úsečiek sú svalové vlákna v ľudskej kmeňovej bunke. V posledných rokoch sa ukazuje, že pre bunkovú dynamiku je okrem tradične skúmaných biochemických podnetov dôležité aj mechanické prostredie. Obzvlášť výrazná je mechanicky stimulovaná zmena ľudských kmeňových buniek (KB) kultivovaných na pružných substrátoch rozličnej tuhosti. Počas ranného štádia tohto mechanicky stimulovaného zmenového procesu v KB hrá štruktúra a polarizácia aktínmyozínových stresových vlákien kľúčovú úlohu a môže byť použitá ako včasný morfológický ukazateľ a analogicky myoblastového rozlíšenia. Stresové vlákna sú približne lineárne s rozličnou dĺžkou, šírkou a orientáciou. Použitím fluorescentnej mikroskopie je možné zobraziť rozloženie vlákien a spoľahlivá extrakcia polohy, dĺžky, šírky a orientácie vlákien je nevyhnutnosťou pre ďalšie spracovanie a základné porozumenie celkovej bunkovej mechaniky. Najmä štúdie dynamiky stresových vlákien, kde sú zaznamenávané mnohopočetné časové rady živých buniek paralelne počas 24-48 hodín vyžadujú rýchly a spoľahlivý algoritmus rozpoznávania stresových vlákien. Uvedený článok popisuje práve algoritmus, ktorý vedie k výstupu množiny stresových vlákien reprezentovaných úsečkami s rozličnou polohou, dĺžkou, šírkou a orientáciou. Dostatočne kvalitná extrakcia týchto vlastností vlákien nebola možná použitím doteraz známych metód. V článku (Eltzner a kol., 2014) je podrobne popísaný algoritmus Filament Sensor(voľne preložené "snímač vlákien"), ktorý ďalej v texte označíme FS. Hlavné požiadavky sú, aby FS extraktoval:

1. rýchlo a bez dohľadu
2. robustne (odhalenie tmavých čiar pretínajúcich svetlejšie, riešenie nehomo-

genity a rozmazanosti obrázku)

3. všetky vlastnosti vlákien (poloha, dĺžka, ...).

Pre ilustráciu uvádzame dva obrázky (viď obrázok 4.1), ktoré sme spolu s .csv súborom obsahujúcim číselné parametre realizácie získali od pána profesora S. Huckemanna z Univerzity v Göttingene (jedného zo spoluautorov článku (Eltzner a kol., 2014)). Prvý zobrazuje, ako vyzerá bunka fluorescentne nasvietená a zväčšená, druhý má zvýraznené žltou farbou, ktoré úsečky (vlákna) zachytil FS algoritmus.



Obr. 4.1: Fluorescentne nasvietená kmeňová bunka (vľavo) a tá istá bunka so zvýraznením úsečiek detekovaných FS algoritmom (vpravo).

Vyššie v texte uvádzame, že pre jednotlivé úsečky sa eviduje aj šírka jednotlivých vlákien, pričom my pri aplikácii metód túto vedomosť nevyužívame, vlákna považujeme za úsečky rovnakej hrúbky. Pre účely načítania dát Mathematica notebookom je .csv súbor upravený v MS Exceli a uložený ako .xlsx súbor v zložke, kde sa nachádza aj samotný notebook. V pôvodnom .csv súbore sú jednotlivé údaje pre každú úsečku v samostatnom riadku, ale v spoločnom stĺpci pričom hodnoty sú oddelené čiarkou. Preto treba najskôr rozdeliť čísla do samostatných stĺpcov, stĺpec pre šírku úsečky rovno vymažeme. Do stĺpca za uhol pridáme jednoduchým prepočtom (ak je uhol $\phi < 90^\circ$, tak je to ϕ , inak $\phi - 180^\circ$) uhol v stupňoch v rozmedzí $(-90^\circ, 90^\circ)$.

Ako je z podstaty dát zjavné, vyskytujú sa tam nekonvexné oblasti a koncentrácia vlákien po okraji bunky je vyššia ako v jej strede. Pre použitie metód na odhad parametrov modelu úsečiek je nevyhnutné, aby sme zobrali do úvahy konvexnú homogénnu oblasť. Preto vyberieme len tie stredy úsečiek, ktoré spadnú do vybraného okna. Okno volíme odhadom, podľa toho, koľko bodov chceme skúmať a tak, aby sme vybrali vlákna zo stredu bunky, nie z jej okrajov.

4.2 Popis kódu

Program v Mathematice je rozložený do sekcií a zároveň stručne okomentovaný pri väčšine príkazoch a funkciách, aby bolo zjavné o čo v konkrétnych častiach kódu ide, prípadne prečo sme zvolili určité východzie parametre a spôsob výpočtu.

V prílohe na CD je niekoľko notebookov, pre každú reálnu bunku jeden a tiež program pre simulované dáta. Program popíšeme všeobecne, pre jednotlivé bunky resp. simulované dáta sú funkcionality rovnaké, menia sa medzivýsledky a prípadné závery z nich.

Program sa začína načítaním dát zo správne upraveného .xlsx súboru. Načítavame len presné stĺpce a riadky, v ktorých sú hodnoty. To znamená, že je nutné nastaviť počet riadkov, začíname riadkom dva a končíme počtom úsečiek plus jedna (vynecháva sa prvý riadok s legendou tabuľky). Máme načítané polohy stredov úsečiek, dĺžky a uhly, a preto nasleduje základný rozbor týchto hodnôt, aby sme získali počiatočnú informáciu o charaktere dát. Určí sa veľkosť okna - ako maximum a minimum súradníc stredov úsečiek, Lebesgueova miera (obsah) okna a vykreslíme histogramy pravdepodobnostného rozdelenia dĺžok a uhlov. Histogramy zobrazujeme tri s rôznym počtom stĺpcov, aby sa dalo ľahšie rozhodnúť, ktorý z nich bude najlepšie použiť ako východzie rozdelenie pre generovanie úsečiek.

Ďalej definujeme dve funkcie - jedna dáva výstup true/false podľa toho, či majú dve úsečky priesečník (priesečník v zmysle, že sa pretínajú v práve jednom bode, funkcia neodhalí úsečky totožné, ani z časti totožné), druhá nájde konkrétny bod, kde sa pretínajú. Vstupom do funkcií je začiatkový a koncový bod, preto je nevyhnutné vytvoriť premennú, v ktorej evidujeme všetky počiatočné a koncové body úsečiek. Použili sme štandardné goniometrické vzorce využívajúce polohu stredu, dĺžku a uhol. V ďalších výpočtoch nie je podstatné, ktorý bod úsečky zvolíme ako počiatočný.

V neposlednom rade pri vstupných dátach napočítame súčet dĺžok systému a celkový počet priesečníkov. Potom skontrolujeme, či sa nejaké tri alebo viac úsečiek nepretína v jednom bode, aby sme včas odhalili nevhodné vstupné dáta.

Ako sme popisovali v 3. kapitole pri metódach odhadu parametrov budeme potrebovať vygenerovať také body $y_i = (v_i, r_i, \phi_i)$, že v_i sú rovnomerne rozmiestnené stredy úsečiek v okne B , r_i je generované z rozdelenia dĺžok podľa histogramu o počte stĺpcov, ktorý sme nastavili (parameter *binsDlzkky*) a ϕ je z rozdelenia histogramu orientácií s počtom stĺpcov *binsUhly*. Samotný postup generovania spočíva v tom, že najskôr na základe pravdepodobností dĺžok z daných rozsahov histogramu vyberieme, z ktorého intervalu chceme dĺžku vybrať. Na tomto

intervale potom už rovnomerne náhodne vyberieme celočíselnú dĺžku a takto pokračujeme, kým nevytvoríme požadovaný počet dĺžok. Obdobne postupujeme pri uhloch. Počet takto vytvorených úsečiek uvažujeme 10000, pričom tento počet je možné upravovať podľa potreby s tým, že samozrejme s vyšším počtom rastie aj výpočtová náročnosť oboch metód odhadu parametrov modelu.

V tomto momente už je všetko pripravené na samotný odhad parametrov (θ_1, θ_2, z) . Pre urýchlenie výpočtu pri metóde T-F najskôr parametricky napočítame aproximované premenné $C_B^{z, \theta}(x, f_i)$ z (3.4) a minimalizujeme súčet ich kvadrátov cez všetky tri parametre. Samozrejme s podmienkami $\theta_2 \leq 0$ a $z > 0$. Obdobne postupujeme aj pri momentovej metóde, kde sme naprogramovali veľičiny z podkapitoly 3.1. Tentokrát ale využívame len 1000 z nagenovaných úsečiek.

Tým pádom už máme odhady parametrov modelu pre obe metódy spočítané a môžeme pripraviť úsečky a funkcie potrebné k simuláciám a následne porovnávacie štatistiky (grafické aj číselné). Príprava úsečiek, ktoré použijeme pri procese rodenia a zániku sa významne nelíši od spôsobu, akým sme generovali úsečky pre účely odhadu parametrov. Jediným rozdielom je, že používame inú *SeedRandom* - vstavajúcu funkciu, ktorá nastaví náhodný generátor na nejakú hodnotu, aby bolo možné výpočty zhrnúť s rovnakým výsledkom. Počet úsečiek volíme ako *iter* a môže sa líšiť od počtu iterácií (*kolko*). Je tomu tak najmä z praktických dôvodov, aby bolo možné vyskúšať aj nižší počet iterácií v procese rodenia a zániku v prípade, že by napr. nastala situácia, kedy by boli úsečky len prijímané a ich počet by sa neustálil.

Vo všetkých našich výpočtoch je počet iterácií v procese rodenia a zániku rovný stotisíc (premenná *kolko*). Ako *kolko* je možné voliť aj iné číslo, ideálne by však malo ísť o násobok desaťtisíc a nesmie byť väčšie ako *iter*. Proces rodenia a zániku spočíva vo voľbe počiatkovej úsečky a v dvoch funkciách. Počiatkovú úsečku volíme ako nejakú "strednú" úsečku, teda takú, čo má približne priemerné súradnice stredu, dĺžku a uhol, ako mali úsečky v podkladových dátach. Prvá funkcia potom spočíta Hastingsov pomer z (2.8) pre nagenovanú úsečku a realizáciu aktuálnu pre iteráciu, v ktorej sa nachádzame. Druhá funkcia rozhodne, podľa algoritmu popísaného v podkapitole 2.1, či daná úsečka vznikne, nejaká existujúca zanikne alebo sa nestane nič. Výstup z tejto funkcie je v takom formáte, aby opäť vedel vstúpiť do funkcie a teda systém po poslednej iterácii je výstupom zo zloženej funkcie, kedy *kolko*-krát do seba vnoríme funkciu *fciaRozhodni3* pre parametre odhadnuté T-F a *fciaRozhodni3MM* pre odhady momentovou metódou. Pri simulácii teda pracujeme pre obe metódy odhadu s rovnakými náhodne generovanými úsečkami, jediným rozdielom simulácie sú parametre (θ_1, θ_2, z) , pomocou ktorých sa spočíta Hastingsov pomer. Všetky realizácie po každej iterácii

pozostávajú zo súradníc začiatkových a koncových bodov úsečiek a z vektora dĺžok všetkých úsečiek. Sú uložené v premenných *foldlistTF*, respektíve *foldlistMM*. To nám umožňuje pozeráť sa na vývoj viacerých hodnôt už počas simulovania, nie len na jeho konci.

Záverečná časť programu obsahuje všetky grafické aj číselné porovnania metód medzi sebou, kde to má zmysel uvažovať, tam samozrejme porovnávame s podkladovými dátami. Najjednoduchším porovnaním je vykreslenie systémov úsečiek po skončení simulácie. Ďalšie porovnania, ktoré sa v práci vyskytujú sú stručne popísané v podkapitole 3.3. Na tomto mieste by sme ich rozšírili o nejaké špecifiká aplikované pri programovaní.

Grafy vývoja celkového počtu úsečiek aj ich celkovej dĺžky v priebehu simulácie obsahuje stavy vo všetkých iteráciách. Tomu tak ale nie je v prípade priesečníkov - začíname po iterácii $od = 10000$ a zobrazujeme potom stav po každej ďalšej stej iterácii ($krok2 = 100$). Počítanie priesečníkov má najvyššiu výpočtovú náročnosť spomedzi ostatných charakteristík, preto sme to takto nastavili. Informácia je aj tak dostatočná a významne sa urýchli výpočet. Keďže máme posunutý počiatok do polohy *od* nastavujeme adekvátne aj čiarky na súradnicovej osi. Ich výpočet je naviazaný na *krok2* a veličinu *od* preto je vhodné *krok2* nemeniť, prípadne meniť po zvážení vplyvu na označenie čiarok na osi. Rovnako berieme do úvahy len tieto vybrané body pri zostrojovaní histogramu zastúpenia konkrétnych hodnôt pre celkový počet priesečníkov. Histogram je teda v prípade celkového počtu priesečníkov konštruovaný z $(kolko - od) / krok2$ čísel, ale pri počte úsečiek a celkovej dĺžke počas simulovania je to *kolko* čísel.

Výpočet priemeru a smerodajnej odchýlky už uvažujeme všade rovnako - počítame ich po každej tisícej iterácii ($krok1 = 1000$). Premennú *krok1* môžeme zmeniť, je však potrebné, aby podiel $(kolko - od) / krok1$ bol prirodzené číslo.

Podklady histogramov pre stav dĺžok po iteráciách 60tisíc, 80tisíc a 100tisíc máme priamo uložené v týchto pozíciách v premenných *foldlistTF*, *foldlistMM*. Čo sa týka orientácií, tie vypočítame cez arkus tangens ako

$$\phi = \arctan((y_2 - y_1) / (x_2 - x_1)),$$

kde x_1, y_1 sú súradnice začiatku úsečky a x_2, y_2 súradnice jej konca.

Celý tento popis je zhodný pre analyzovanie reálnej bunky ako aj simulovaných dát. Mierne odlišný je na začiatku program pre simulované dáta, keďže sme ich najskôr potrebovali nasimulovať a potom aplikovať celý horeuvedený postup. Vychádzame z reálnej bunky a simulujeme pre pevne stanovené hodnoty parametrov (θ_1, θ_2, z) , ktoré sme zvolili ako $(-0,19; -0,12; 1,1)$. Výsledná realizácia po procese rodenia a zániku pri počte iterácií 100-tisíc sú naše podkladové simulované dáta.

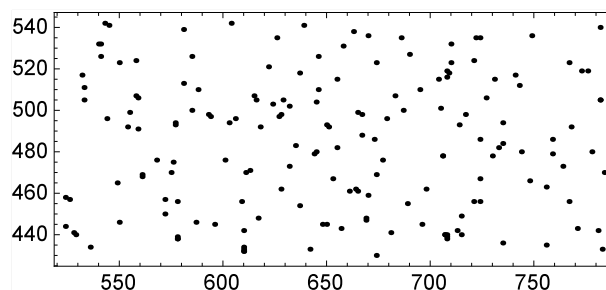
Kapitola 5

Numerické výsledky

V záverečnej kapitole práce budeme interpretovať výsledky a porovnáme ich pre reálnu a simulovanú bunku. V podkapitole 5.1 vychádzame z reálnych dát jednej kmeňovej bunky. Odhadneme parametre modelu a simulujeme model s odhadnutými parametrami. Porovnáваме popisné štatistiky reálnych a simulovaných dát a súčasne porovnáваме obidve metódy odhadu. V podkapitole 5.2 pracujeme predovšetkým so simulovanými dátami. Postupujeme rovnako ako v podkapitole 5.1 s tým rozdielom, že východzie dáta sú simulované a teda vieme, že zodpovedajú modelu (potom už porovnávanie dáva lepšie výsledky). Všetky grafické aj číselné výstupy sú aj v Mathematica notebookoch priložených na CD.

5.1 Reálne dáta (bunka1)

V duchu popisu kódu uvedieme jednotlivé výsledky pre prvú bunku (označme ju bunka1). Na obrázku 5.1 je vykreslenie stredov úsečiek bunky1. Tieto body sme zvolili ako vzorku z bunky tak, aby bolo rozmiestnenie bodov približne homogénne a nebrali sme do úvahy body z okrajových oblastí bunky.



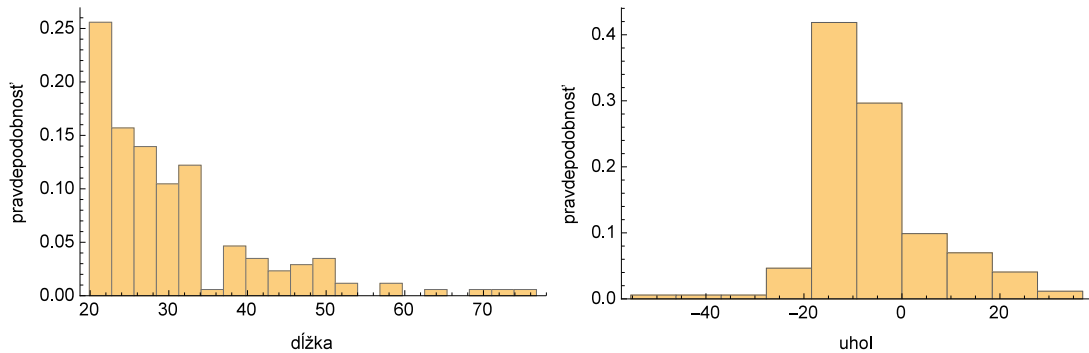
Obr. 5.1: Bodový graf stredov úsečiek bunky1

Pod obrázkom je umiestnená tabuľka 5.1, ktorá obsahuje základné rozmery okna, maximum dĺžky a Lebesgueovu mieru množiny. Okno sme zvolili obdĺžnikového tvaru, s rozmerom 260×112 . Samotná bunka1 pozostáva zo 172 úsečiek.

Tabuľka 5.1: Základné vlastnosti popisujúce bunku1

Vlastnosť	Hodnota
Max x súradnica	784
Min x súradnica	524
Max y súradnica	542
Min y súradnica	430
Max dĺžka	74
Lebesgueova miera $ B $	29120

Po načítaní súradníc stredov, dĺžok a uhlov sme skúšaním histogramov s rôznym počtom stĺpcov rozhodli, že za rozdelenie dĺžok budeme považovať histogram s dvadsiatimi stĺpcami ($binsDlzkoy = 20$). Pre orientácie je postačujúce, aby stĺpcov bolo len desať ($binsUhly = 10$). Príslušné histogramy vidíme na obrázku 5.2. Viac ako 2/3 dĺžok sú v rozsahu 20 – 30, podobne takmer 70% orientácií je v rozmedzí -20° až 0° .



Obr. 5.2: Histogramy rozdelenia pravdepodobnosti dĺžok (vľavo) a orientácií (vpravo), z ktorých vychádzame pri generovaní úsečiek - bunka1

Pre účely odhadovej metódy T-F využijeme $mcmc = 10000$ úsečiek so stredmi rovnomerne rozmiestnenými v okne ohraničenom hodnotami z tabuľky 5.1, dĺžky a orientácie sú celočíselné a pochádzajú z rozdelení, ktoré určujú pravdepodobnostné histogramy z 5.2. Odhady parametrov momentovou metódou sú počítané s využitím rovnakých náhodných úsečiek ako u T-F metódy, ale počítame len s prvými tisíc bodmi ($mcmcMM = 1000$).

V nasledujúcej tabuľke 5.2 sú zapísané parametre odhadnuté obidvomi metódami pre bunku1.

Za najmenej odlišné by sme mohli považovať odhady pre θ_1 , avšak $\hat{\theta}_2$ vyšlo pre metódu T-F nulové. Znamená to, že sa jedná o Poissonov proces a čo do celkového počtu priesečníkov môžeme očakávať, že po skončení simulácie procesu rodenia a zániku sa jednotlivé metódy budú odlišovať vo vývoji celkového počtu priesečníkov oveľa viac, ako sa budú odlišovať v celkovej dĺžke. Paramater z

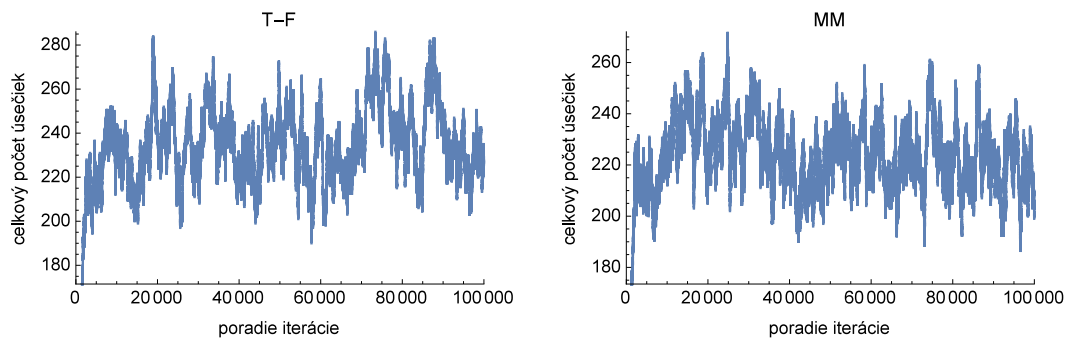
Tabuľka 5.2: Odhady parametrov (θ_1, θ_2, z) obidvomi metódami pre bunku1.

Parameter	T-F	MM
$\hat{\theta}_1$	-0,255606	-0,1936
$\hat{\theta}_2$	0	-0,322196
\hat{z}	4,52689	1,31085

odhadnutý T-F metódou vyšiel viac ako trojnásobný oproti MM odhadu.

Teraz použijeme MCMC k simulácii procesov úsečiek s využitím parametrov z tabuľky 5.2 odhadnutých príslušnou metódou na reálnych dátach bunky1. Vo výsledných grafoch pre priebeh simulácie a jej záverečnú realizáciu úsečiek neskôr uvidíme, aký bude mať rozdiel v odhadoch dopad na celkový počet úsečiek, keďže ten výrazne závisí aj od toho, ako dobre parameter intenzity z doplnia pri výpočte Hastingsovho pomeru aj zvyšné dva parametre. Prvá úsečka pre simuláciu procesu rodenia a zániku má počiatočný bod $\{673, 471\}$ a koncový $\{640, 475\}$. Jej dĺžka je 33.

Všetky vstupy do simulácie máme známe a preto teraz uvedieme výstupy z priebehu simulácie, ako aj vlastnosti záverečnej realizácie po skončení simulácie s využitím parametrov z tabuľky 5.2. Veľmi dobre ilustračný je obrázok A.1, na ktorom vidíme výsledné systémy úsečiek po skončení simulácií v porovnaní s podkladovým systémom úsečiek bunky1. Vyzerá to tak, že simulované dáta obsahujú úsečiek o niečo viac a skôr kratších, orientácie nevykazujú žiadnu zjavnú zmenu.



Obr. 5.3: Graf vývoja celkového počtu úsečiek pre systémy úsečiek po každej iterácii procesu rodenia a zániku (vľavo pre T-F, vpravo pre MM) - bunka1

Na obrázku 5.3 sa môžeme pozrieť na vývoj celkového počtu úsečiek počas simulácie. Tento údaj pre obe metódy kolíše, ale v oboch prípadoch môžeme povedať, že kolíše okolo nejakej priemernej hodnoty a nedochádza k neobmedzenému rastu ani poklesu počtu úsečiek do nuly. Číselné vyjadrenie priemerného počtu a smerodajnej odchýlky nájdeme v tabuľke 5.3, pričom tieto štatistiky počítame zo

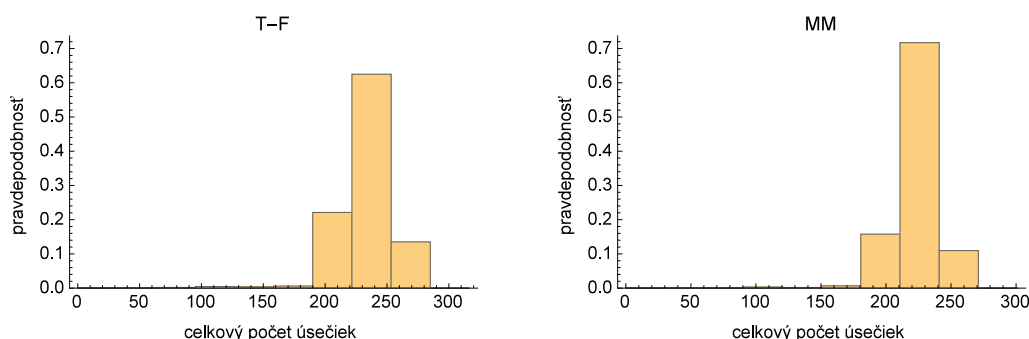
stavov realizácií po každej tisícej iterácii (*krok1*) medzi realizáciou $od = 10000$ a poslednou iteráciou. Iterácie s nižším poradovým číslom neberieme do úvahy, lebo simulácia začína na jednej úsečke a tým by sa znižoval priemer a zväčšovala smerodajná odchýlka.

Tabuľka 5.3: Priemer a smerodajná odchýlka celkového počtu úsečiek zo simulácií pre bunku1.

Štatistika	T-F	MM	Podkladové dáta
Priemer	235,8	225,3	172
Smer. odchýlka	16,9	12,9	-

Ako sme predpokladali, ukázalo sa, že čo do počtu úsečiek ich bude viac v simulácii T-F ako v MM. Rozdiel v priemeroch však nie je veľký, je to len 10 úsečiek. U oboch metód ale máme zhruba o viac ako štvrtinu vyšší počet úsečiek ako bolo v pôvodnom systéme bunky1. Smerodajná odchýlka vyšla nižšia pre parametre metódy MM. Posledným porovnaním, ktoré uvedieme pre celkový počet úsečiek počas simulácie je histogram zkonštruovaný z hodnôt grafu vývoja počtu úsečiek. Teda zistíme, s akou pravdepodobnosťou sa vyskytujú jednotlivé celkové počty úsečiek (viď obr. 5.4). Histogram zodpovedajúci T-F metóde je mierne posunutý vpravo, viac ako 60% realizácií v priebehu simulácie obsahuje 230 až 260 úsečiek. Pre momentovú metódu však vidíme, že výkyvy nie sú také veľké a pri viac ako 70% realizáciách máme celkový počet úsečiek v rozsahu 210 až 240.

Stručným záverom pre správanie sa počtu úsečiek je, že u oboch metód zo simulácie vyšlo viac úsečiek, ako mal pôvodný systém a zároveň bližšie (v zmysle priemeru) aj stálejšie výsledky (v zmysle smerodajnej odchýlky aj histogramu) boli pre parametre odhadnuté momentovou metódou.



Obr. 5.4: Histogramy počtu úsečiek v priebehu simulácie pre parametre odhadnuté obidvomi metódami - bunka1.

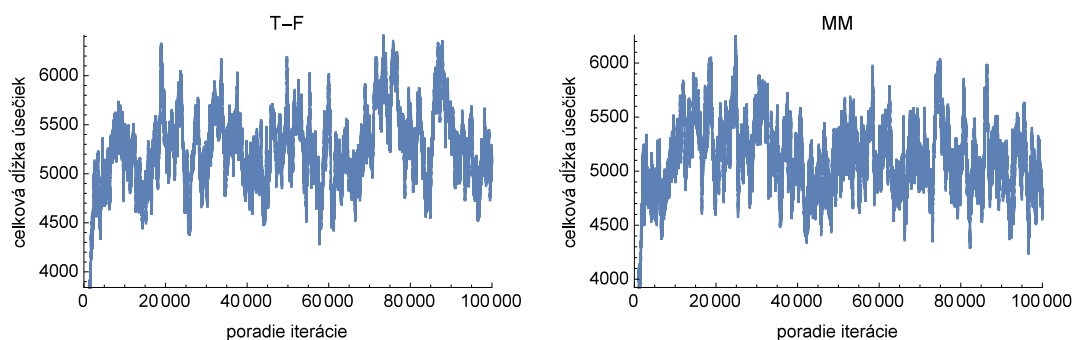
Podobne ako pre celkový počet úsečiek teraz popíšeme výsledky týkajúce sa

celkovej dĺžky úsečiek v priebehu simulácií. Najkôr sa pozrime, ako vyšla priemerná dĺžka a jej smerodajná odchýlka pre rovnaké realizácie v priebehu simulácie ako tomu bolo pri sledovaní celkového počtu úsečiek. Číselné hodnoty obsahuje tabuľka 5.4.

Tabuľka 5.4: Priemer a smerodajná odchýlka celkovej dĺžky úsečiek pre bunku1.

Štatistika	T-F	MM	Podkladové dáta
Priemer	5292,6	5160,3	5243
Smer. odchýlka	377,9	301,8	-

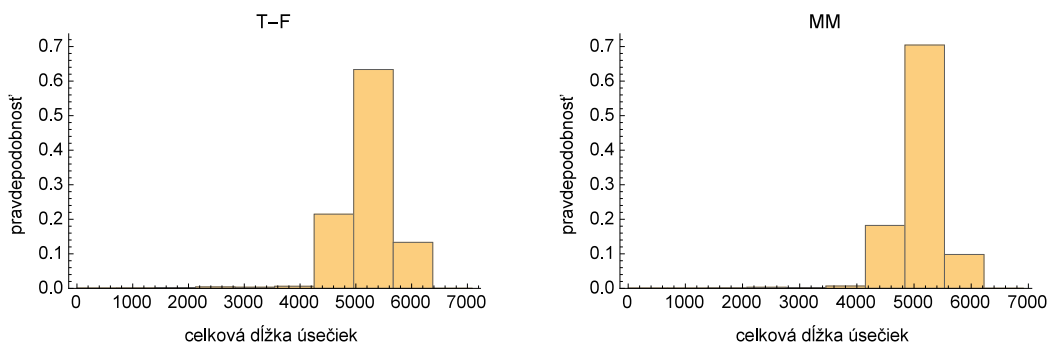
Ukazuje sa, že obe metódy veľmi dobre zachovávajú celkovú dĺžku. Priemerné hodnoty sa od skutočného súčtu dĺžok podkladovej realizácie úsečiek odlišujú o menej ako 3%. Smerodajná odchýlka je opäť vyššia pre údaj zodpovedajúci metóde T-F. Ako sa v priebehu simulácie vyvíjal súčet dĺžok úsečiek vidíme na grafe 5.5. Celková dĺžka najskôr rýchlo rastie prijímaním úsečiek a potom sa od priemernej hodnoty vychýľuje nahor k 6000 aj nadol po hodnotu 4300, ale neobjavujú sa extrémnejšie výkyvy.



Obr. 5.5: Graf vývoja celkovej dĺžky úsečiek pre systémy úsečiek po každej iterácii procesu rodenia a zániku (vľavo pre T-F, vpravo pre MM) - bunka1

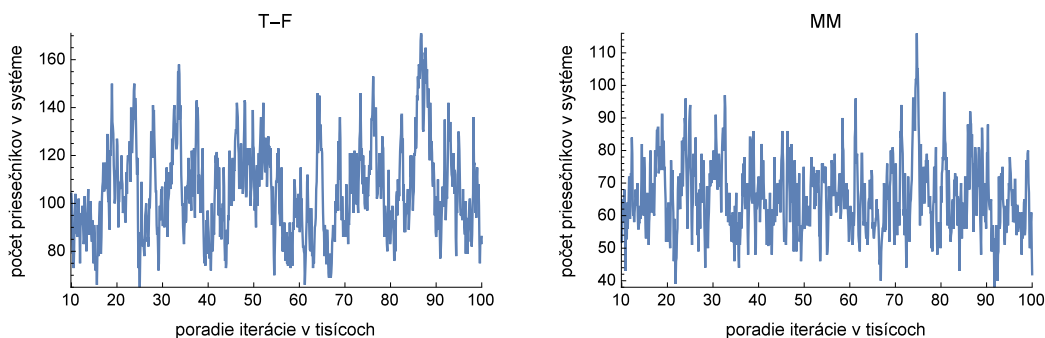
Zostáva zamerať sa na početnosť jednotlivých celkových dĺžok počas simulácie, ktoré zobrazíme pomocou histogramu. Na obr. 5.6 si môžeme všimnúť, že obe metódy sú veľmi podobné až na drobný rozdiel, ktorý indikoval už výsledok v smerodajnej odchýlke a to, že pre momentovú metódu evidujeme menšie výkyvy, keďže viac ako 70% hodnôt spadá do jedného stĺpca histogramu. Ten zodpovedá rozsahu celkovej dĺžky približne 4800 až 5570.

Dosiaľ uvedené výsledky tejto kapitoly naznačujú, že keďže oba modely veľmi dobre zachovávajú štatistiku celková dĺžka všetkých úsečiek systému, ale počet úsečiek sa významnejšie zvýšil, odrazí sa to zrejme v zmene rozdelenia dĺžok. Toto ešte upresníme neskôr, venujme sa teraz výstupom týkajúcim sa súčtu priesečníkov medzi úsečkami jednotlivých realizácií počas simulácie. Graf vývoja počtu



Obr. 5.6: Histogramy celkovej dĺžky úsečiek v priebehu simulácie pre parametre odhadnuté obidvomi metódami - bunka1.

priesečníkov nie je spočítaný po každej iterácii, ale po každej stej a začína od iterácie 10000, podľa čoho je aj upravená os na obrázku 5.7



Obr. 5.7: Graf vývoja celkového počtu priesečníkov pre systémy úsečiek po každej stej iterácii procesu rodenia a zániku (vľavo pre T-F, vpravo pre MM) - bunka1

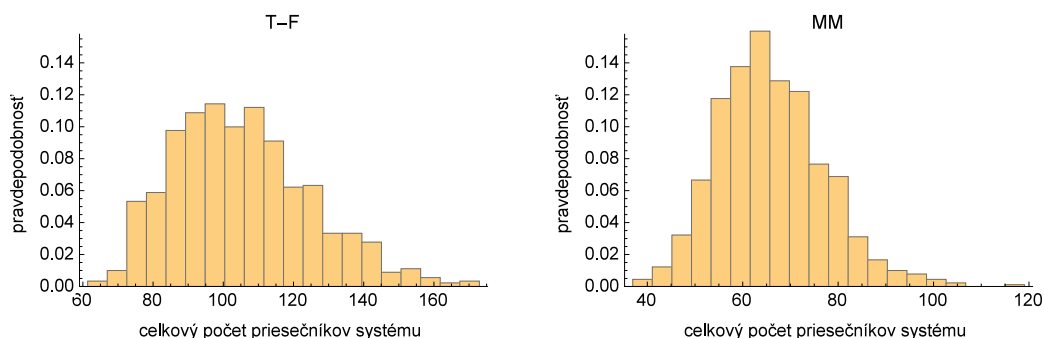
V počte priesečníkov sa pohybujú hodnoty pre metódu T-F okolo čísla 100, klesajú až k 70-tke a najvyššie vystúpajú k 160-tim priesečníkom. Čo je ale zaujímavé, narážame na prvý významnejší rozdiel medzi metódami, pretože počet priesečníkov pre simuláciu s parametrami odhadnutými momentovou metódou sa krivka pohybuje zhruba okolo počtu 65. Priemerné hodnoty aj smerodajné odchýlky zhrnieme v nasledujúcej tabuľke.

Tabuľka 5.5: Priemer a smerodajná odchýlka celkového počtu priesečníkov v realizácii úsečiek počas simulácie pre bunku1.

Štatistika	T-F	MM	Podkladové dáta
Priemer	105	65,6	102
Smer. odchýlka	20,5	11,9	-

Výsledky z tabuľky 5.5 nám hovoria, že priemerná hodnota počtu priesečníkov v T-F metóde vychádza 105, čo veľmi dobre korešponduje s podkladovými

dátami. Naopak metóda MM počet priesečníkov spoľahlivo nezachovala. Keďže priemerné hodnoty sú v oboch sledovaných modeloch odlišné, zastúpenie jednotlivých počtov priesečníkov počas simulácie z obrázku 5.8 zvyrazňuje tento fakt.



Obr. 5.8: Histogramy celkových počtov priesečníkov v priebehu simulácie pre parametre odhadnuté obidvomi metódami - bunka1.

Histogramy pozostávajú z viacerých stĺpcov aby sme poukázali na odlišnosť sledovaných modelov. Väčšina hodnôt z ľavého histogramu spadá do rozmedzia 80 – 140, z pravého histogramu je to rozsah 50 – 90.

Aj pre priesečníky je možné zostaviť charakteristiku pre jednotlivé úsečky. Zameriame sa na posledné realizácie po skončení simulácie v porovnaní s podkladovými dátami bunky1. Spočítame koľko priesečníkov má každá úsečka so zvyškom systému a z týchto hodnôt zostavíme histogram. Výsledok môžeme vidieť na obrázku A.2, ktorý je v obrazovej prílohe. V bunke1 sa počet priesečníkov pohyboval od nuly do 7, pri metóde T-F to bolo 6 a najmenšie maximum je pri MM a to 4. Aj samotné rozdelenie hodnôt je iné, čo sa ale očakávalo, keďže MM model generoval menej priesečníkov a preto aj takmer 66% úsečiek nemá priesečník so systémom. U podkladových dát bunky1 je pravdepodobnosť úsečky bez priesečníku so systémom rovná len asi 37%. Potom pravdepodobnosť jedného priesečníku je u T-F a bunky1 podobná, ako zastúpenie tých nepretínajúcich systém, kdežto v MM modeli je úsečiek pretínajúcich systém raz asi polovica oproti bezpriesečníkovým.

Ako posledné výstupy týkajúce sa bunky1 uvidíme histogramy s rozdeleniami dĺžok a uhlov po 60., 80. a 100. tisícej iterácii. V predchádzajúcom texte sme načrtli, že očakávame istý posun v rozdelení dĺžok. Dôvodom je vyšší priemerný počet úsečiek, pričom celková dĺžka korešponduje s bunkou1. Čo sa týka orientácií dosiaľ uvedené výsledky neindikovali významnú zmenu. Pozrime sa teraz na príslušné histogramy. Uvádzame ich v obrázkovej prílohe, aby bolo možné čo najlepšie porovnať výstupy príslušné jednotlivým metódam v porovnaní s rozdeleniami dĺžok a orientácií bunky1, ktoré boli zároveň použité ako východzie rozdelenia pre generovanie náhodných úsečiek.

Všimnime si na obr. A.3 najskôr stav v realizácii 60000. Rozdelenie dĺžok pre T-F je posunuté vľavo, do menších hodnôt. Zároveň najkratších úsečiek dĺžky 20 – 22 je len o málo viac ako v reálnej bunke1. Výraznejší rozdiel je v MM histograme, kde vidíme, že úsečky dĺžky 20 – 21 tvoria až 45%. Pre obe metódy už je len jedna dĺžka vyššia ako 36, pričom v pôvodnom systéme sme takto dlhých úsečiek mali viac ako 15%. Podobný efekt posunu rozdelenia dĺžok smerom ku kratším úsečkám vidíme aj pre realizácie po 80tisíc iteráciách. Teda trend „skracovania„ úsečiek sa objavuje už u prvej trojice histogramov a zostáva zachovaný aj pri posledných realizáciách simulácie. Obrázok A.3 teda ukazuje, že v priebehu simulácie sa rozdelenie dĺžok výraznejšie mení v zmysle zastúpenia kratších úsečiek 20 – 25, ale dlhšie úsečky už sa neobjavujú.

Pozrime sa teraz na histogramy orientácií na obr. A.4. Prvá trojica obrázkov, ktorá ukazuje stav po iterácii 60000 vs. podkladové dáta vyšla veľmi dobre, rozdelenia sú podobné, mierny rozdiel je, že v bunke1 je o niečo viac orientácií $-9^\circ - 0^\circ$ a to 30% oproti 25%, respektíve 22%.

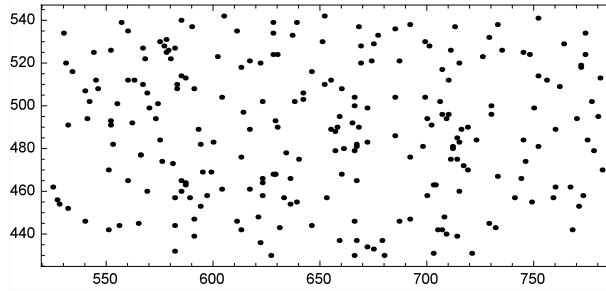
Z ďalšej sady histogramov vidíme, že rozdelenie orientácií zodpovedajúce T-F je takmer úplne rovnaké ako má samotná bunka1. Zmena nastala v MM obrázku, kde sa zvýšilo zastúpenie zápornejších uhlov. Dominujú orientácie v rozsahu -15° až -9° so zastúpením 33%. Keď si všimneme poslednú trojicu histogramov vidíme, že rozdelenia dobre korešpondujú. Dalo by sa preto povedať, že mierne posuny v zastúpení orientácií počas simulácie nastávajú, ale celkovo už aj po iterácii 60-tisíc dobre korešpondujú s rozdelením orientácií bunky1.

5.2 Simulované dáta (bunkaSI)

Výsledky pre simulované dáta uvedieme v rovnakom formáte, ako tomu bolo pre bunku1, tiež vyhodnotíme všetky štatistiky uvádzané v podkapitole 3.3. Rozdelenia dĺžok, orientácií z obrázku 5.2 a plochu okna bunky1 (viď tabuľku 5.1 sme využili pre získanie simulovaných dát procesom rodenia a zániku s parametrami $(\theta_1, \theta_2, z) = (-0,19; -0,12; 1,1)$. Poslednú realizáciu, teda „novú„ bunku po stotisíc iteráciách procesu vezmeme ako podkladové simulované dáta a ďalej ju budeme označovať bunkaSI.

Pozrime sa najskôr na obrázok 5.9, kde sú formou bodového grafu zobrazené stredy úsečiek bunkySI. V porovnaní s bunkou1 máme stredov viac a sú rovnomernejšie rozmiestnené. To je aj prirodzené, keďže bunka1 sú reálne dáta zodpovedajúce konkrétnej ľudskej kmeňovej bunke a simulovaná bunka by mala lepšie odzrkadľovať požiadavku homogenosti.

Základné parametre okna a Lebesgueova miera $|B|$ sú zapísané v tabuľke 5.6. Nájdeme v nej aj údaj o maximálnej dĺžke úsečky, tou je v tomto prípade

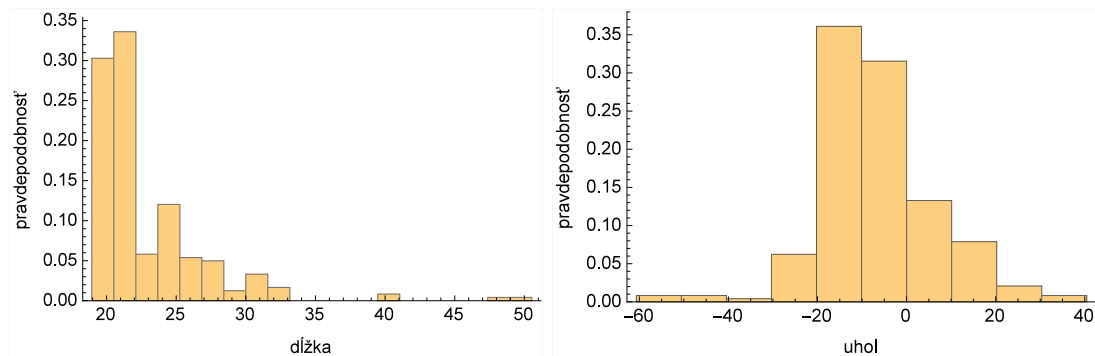


Obr. 5.9: Bodový graf stredov úsečiek bunkuSI

Tabuľka 5.6: Základné vlastnosti popisujúce bunkuSI

Vlastnosť	Hodnota
Max x súradnica	782
Min x súradnica	525
Max y súradnica	542
Min y súradnica	430
Max dĺžka	50
Lebesgueova miera $ B $	28784

dĺžka 50. Lebesgueova miera, teda obsah okna, v ktorom sa vyskytujú stredy všetkých úsečiek systému bunkuSI je o niečo menší ako pre bunku1. Dôvod je zjavný, jednoducho do systému neboli vybrané úsečky s najkrajnejšou súradnicou x bunku1.



Obr. 5.10: Histogramy rozdelenia pravdepodobnosti dĺžok (vľavo) a orientácií (vpravo), z ktorých vychádzame pri generovaní úsečiek pri bunkeSI.

Úvodné hodnoty špecifické pre bunkuSI sú nám už známe a v našom algoritme nasleduje odhad parametrov (θ_1, θ_2, z) . K odhadom nevyhnutne potrebujeme náhodne nagenerovať úsečky, ktorých stredy sú rovnomerne rozmiestnené po celom okne B a ich dĺžky a orientácie pochádzajú z rozdelení zobrazených na obrázku 5.10. Pre dĺžky opäť využívame histogram o 20 stĺpcoch, pričom viac ako 90% dĺžok je menších ako 30, dlhších úsečiek máme len niekoľko. Rozdelenie orientácií máme v rozsahu -60° až 40° , kde viac ako 65% hodnôt leží medzi -20° a 0° .

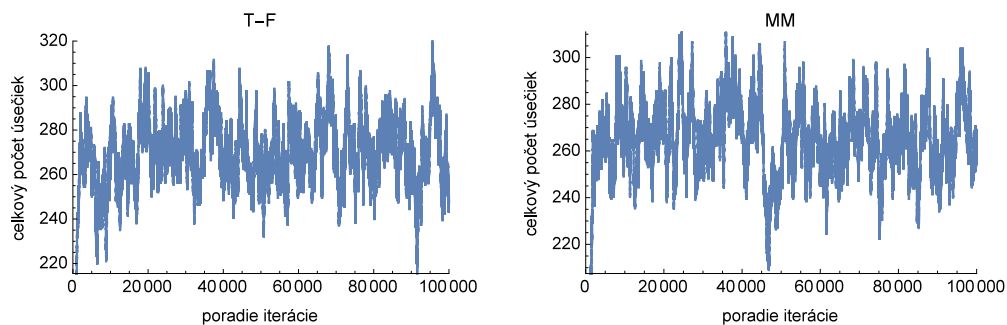
Tabuľka 5.7: Odhady parametrov (θ_1, θ_2, z) obidvomi metódami pre bunkuSI.

Parameter	T-F	MM	Pôvodné(bez $\hat{\cdot}$)
$\hat{\theta}_1$	-0,234887	-0,221707	-0,19
$\hat{\theta}_2$	-0,063435	-0,261809	-0,12
\hat{z}	1,58094	1,38171	1,1

Odhady parametrov sme v Mathematice vypočítali minimalizáciou popísanou v podkapitolách 3.1 a 3.2 a použitím nagerovaných úsečiek. Ako vyšli odhady je zhrnuté v tabuľke 5.7. V prvom rade porovnajme odhadnuté parametre s tými, ktoré boli použité pre vytvorenie bunkySI. Najviac sa odlišujú v parametri θ_2 , kde pri T-F metóde tvorí odhad polovicu a pri MM metóde viac ako dvojnásobok zvoleného parametra. Odhady sa nezhodujú s pôvodnou voľbou ani u zvyšných dvoch parametrov. Tieto rozdiely ale nevyhnutne nemusíme považovať za chybu v odhadoch. Vzhľadom na komplikovanú štruktúru výrazu, ktorý minimalizujeme je možné, že zhruba rovnako dobre model popisujú viaceré kombinácie parametrov. To sa ostatne ukáže pri výsledkoch týkajúcich sa následnej simulácie procesu rodenia a zániku s odhadnutými parametrami z tabuľky 5.7.

Prejdime teraz k popisu správania sledovaných veličín celkový počet úsečiek, ich celková dĺžka a počet priesečníkov systému úsečiek počas simulácie. Počet iterácií procesu rodenia a zániku pre parametre odhadnuté obomi metódami je stotisíc.

Úvodom sa pozrime na vývoj celkového počtu úsečiek. Na obrázku 5.11 si môžeme všimnúť, ako dobre sa počty úsečiek na oboch grafoch zhodujú. Samozrejme majú odlišný priebeh, ale väčšina hodnôt sa drží v páse 240 až 310 úsečiek. Veľmi podobné číselné výsledky preto máme aj v 5.8, kde sú vypísané priemery



Obr. 5.11: Graf vývoja celkového počtu úsečiek pre systémy úsečiek po každej iterácii procesu rodenia a zániku (vľavo pre T-F, vpravo pre MM) - bunkaSI

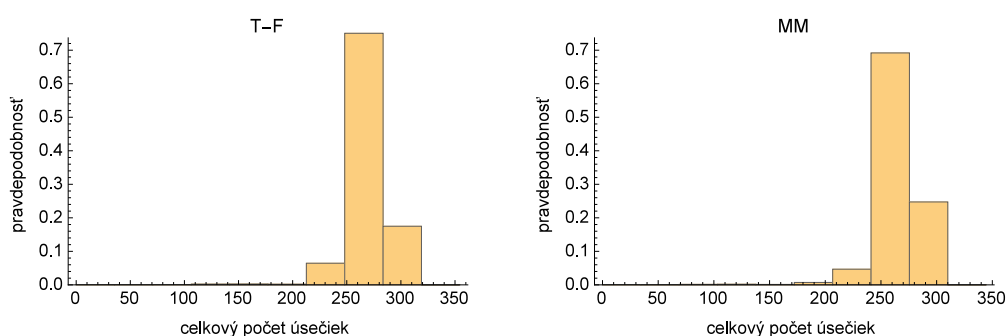
a smerodajné odchýlky celkového počtu úsečiek v priebehu simulácie. Počet úsečiek v bunkeSI je 241 a teda priemery 272 (T-F) a 265,5 (MM) tvoria menej ako

13%ný nárast počtu úsečiek. Smerodajné odchýlky vyšli veľmi podobné pre obe metódy odhadu parametrov modelu.

Tabuľka 5.8: Priemer a smerodajná odchýlka celkového počtu úsečiek pre bunkuSI.

Štatistika	T-F	MM	Podkladové dáta
Priemer	272	265,5	241
Smer. odchýlka	14,5	14,7	-

Čo spôsobili výkyvy v grafoch vývoja počtu úsečiek si pozrieme na dvoch histogramoch na obr. 5.12. Je na ňom zobrazené zastúpenie jednotlivých počtov úsečiek pre všetky systémy úsečiek počas simulácie. V histograme vľavo patrí až 75% hodnôt do intervalu (249, 283). Pri MM je histogram mierne posunutý do menších počtov úsečiek a najbohatšie zastúpenie 69% majú hodnoty v rozsahu 240 až 275. Záverom, výsledky počtu úsečiek v T-F a v MM dobre korešpondujú, ale sú v priemere vyššie, ako bol počet úsečiek v bunkeSI.



Obr. 5.12: Histogramy počtu úsečiek v priebehu simulácie pre parametre odhadnuté obidvomi metódami (bunkaSI).

Zamerajme sa teraz na ďalšiu dôležitú charakteristiku priebehu simulácie a to súčet dĺžok úsečiek jednotlivých systémov. Priemery a smerodajné odchýlky sme spočítali z celkovej dĺžky systému úsečiek prislúchajúcej ku každej tisícej iterácii v rozmedzí 10000 až 100000.

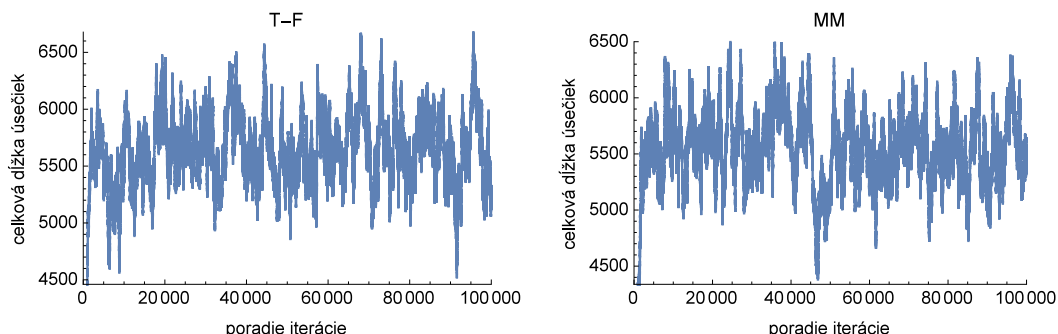
Tabuľka 5.9: Priemer a smerodajná odchýlka celkovej dĺžky úsečiek pre bunkuSI.

Štatistika	T-F	MM	Podkladové dáta
Priemer	5683,3	5555,4	5536
Smer. odchýlka	308,8	304,8	-

Konkrétne hodnoty si môžeme pozrieť v tabuľke 5.9. Priemer celkovej dĺžky úsečiek počas simulácie pre MM je bližší skutočnej celkovej dĺžke úsečiek bunkySI,

ale je dôležité poznamenať, že aj keď je priemer T-F vyšší, jedná sa len o necelé 3% oproti údaju 5536 pre bunkuSI.

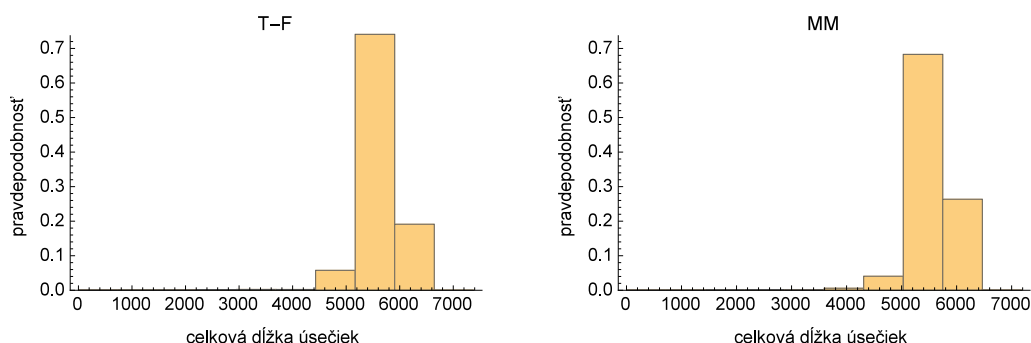
Už sme načrtli výsledky vývoja celkovej dĺžky v absolútnych číslach priemerov a smerodajných odchýlok. Na obrázku 5.13 vidíme graf vývoja celkovej dĺžky pre



Obr. 5.13: Graf vývoja celkovej dĺžky úsečiek pre systémy úsečiek po každej iterácii procesu rodenia a zániku (vľavo pre T-F, vpravo pre MM) v bunkeSI

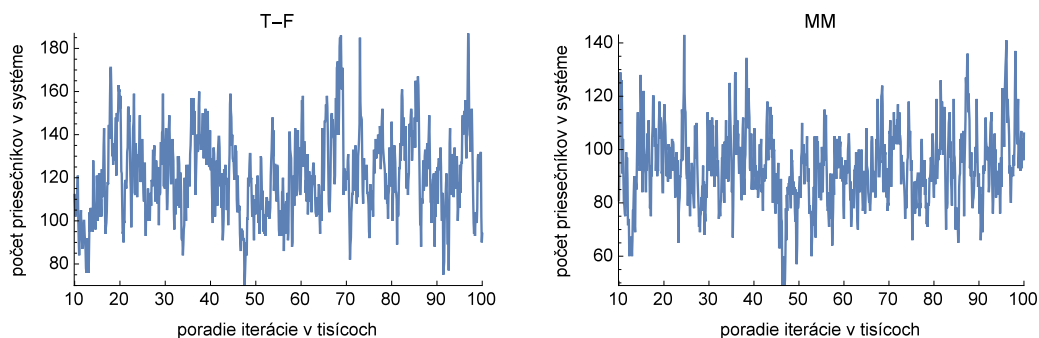
simulácie s využitím parametrov odhadnutých obomi metódami.

Oba grafy majú podobný rozsah osi y , výsledné krivky sa pohybujú v páse (4500, 6500). Ešte nám zostáva pozrieť sa na histogramy výskytu jednotlivých celkových dĺžok počas celej simulácie. Rozdelenia sú veľmi jednoznačné, ako ukazuje obr. 5.14. Pre T-F sú až 3/4 v jednom stĺpci histogramu, ktorý zodpovedá súčtu dĺžok 5200 – 5950. Keď sa obzrieme na všetky výsledky súvisiace s celkovou dĺžkou úsečiek systému je jasné, že pri oboch metódach je táto štatistika výborne zachovaná a korešponduje s bunkouSI.



Obr. 5.14: Histogramy celkovej dĺžky úsečiek v priebehu simulácie pre parametre odhadnuté obidvomi metódami - bunkaSI

Keďže odhady parametra θ_2 , ktorý sa vzťahuje k počtu priesečníkov sa najviac líšili od zvoleného $\theta_2 = -0,12$ zaujíma nás, či aj v konečnom dôsledku počas simulácie uvidíme rozdiely, alebo sa kombinácie parametrov - aj keď odlišné, ukážu ako vhodné pre vytvorenie dát podobných pôvodnej bunkeSI. Na obr. 5.15 si pozrime, ako sa vyvíjal celkový počet priesečníkov v priebehu simulácie.



Obr. 5.15: Graf vývoja celkového počtu priesečníkov pre systémy úsečiek po každej stej iterácii procesu rodenia a zániku (vľavo pre T-F, vpravo pre MM) - bunkaSI

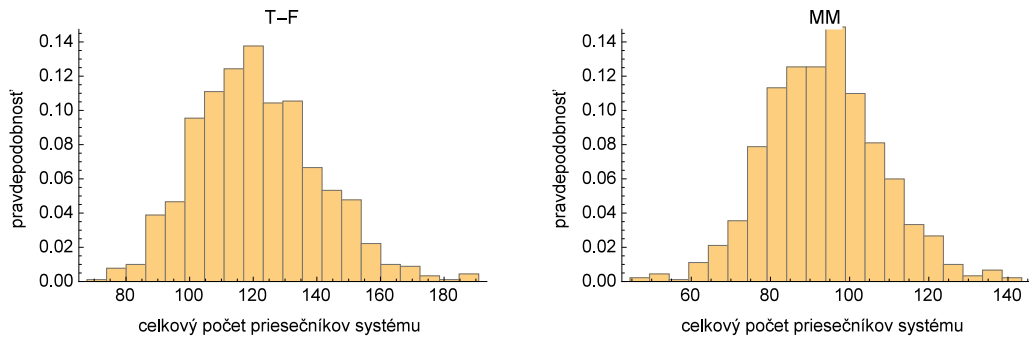
Už na prvý pohľad je očividné, že krivka v grafe T-F sa pohybuje vo vyšších počtoch priesečníkov ako krivka v MM. Pre T-F máme maximálne hodnoty pri 180 kdežto v MM grafe sú najvyššie hodnoty asi 140. Taktiež by sme chceli vedieť, ako sa správajú priemerné hodnoty počtu priesečníkov oproti skutočnému počtu v bunkeSI. Porovnanie si pozrime v tabuľke 5.10. Počet priesečníkov bunkySI je 113 a teda je medzi priemerami T-F a MM. Bližšie k podkladovým dátam má T-F priemer rovný 122,3. V smerodajnej odchýlke je výsledok lepší v časti tabuľky MM, teda počty priesečníkov majú v tomto prípade menšie výkyvy počas simulácie.

Tabuľka 5.10: Priemer a smerodajná odchýlka celkového počtu priesečníkov v realizácii úsečiek počas simulácie pre bunkuSI.

Štatistika	T-F	MM	Podkladové dáta
Priemer	122,3	93,3	113
Smer. odchýlka	21,2	13,7	-

Rozdelenia počtu priesečníkov počas simulácie majú menej rozsiahle intervaly ale pravdepodobnosti sú viac rozložené do jednotlivých stĺpcov. Vyplýva to jednak z toho, že počty priesečníkov sú rádovo iné hodnoty ako napr. súčty dĺžok ale aj z toho, že tieto histogramy nepočítame zo stotisíc hodnôt ale len z $(100000 - 10000)/100 = 900$. V histograme T-F je navýraznejšie zastúpenie hodnôt z intervalu 117 až 123, pričom tvoria necelých 14% počtov priesečníkov z 900 uvažovaných.

Výsledky súvisiace s bunkouSI ešte dokresľuje niekoľko obrázkov umiestnených v obrázkovej prílohe. Nájdeme tam napríklad porovnanie histogramov počtu priesečníkov jednotlivých úsečiek so zvyškom systému pre realizácie úsečiek po skončení simulácie a pre samotnú bunkuSI. Na obrázku A.6 si všimnime, že



Obr. 5.16: Histogramy celkových počtov priesečníkov v priebehu simulácie pre parametre odhadnuté obidvomi metódami- bunkaSI

pre všetky tri histogramy je zastúpenie úsečiek, ktoré systém nepretínajú takmer rovnaký a to približne 46%. Najväčší rozdiel oproti podkladovým dátam je pre pravdepodobnosť jedného priesečníku, kedy na histogramoch prislúchajúcich metódam T-F a MM stĺpce siahajú k 37%, ale v bunkeSI len k 29%.

Na obr. A.7 máme pravdepodobnostné histogramy pre rozdelenie dĺžok v systémoch úsečiek po 60-tisíc, 80-tisíc a 100-tisíc iteráciách. Už pri prvej trojici histogramov si môžeme všimnúť, že len približne 5% dĺžok pri oboch metódach odhadu parametrov je dlhších ako 25. Zároveň sú úsečky do dĺžky 25 rovnomernejšie zastúpené ako v bunkeSI, tam výrazne dominujú dĺžky 20 – 21. V ďalších dvoch pohľadoch na rozdelenie dĺžok vidíme miernu zmenu v zastúpení najkratších úsečiek, so stúpajúcim množstvom iterácií sa rozdelenia pre T-F a MM viac podobajú histogramu bunkySI. Faktom ale zostáva, že zastúpenie dĺžok nad 25 tvorí pre obe metódy necelých 5% oproti bunke SI, kde ich je viac ako 15%.

Podobne, ako sme to spravili pre dĺžky teraz porovnáme aj rozdelenia orientácií. Na obrázku A.8 vidíme, že už po iterácii 60000 rozdelenia orientácií korešpondujú s východzou bunkouSI. Zmeny pri vyššom počte iterácií sú nepatrné, preto by sme mohli považovať rozdelenia orientácií za ustálené už od iterácie 60tisíc s tým, že histogramy sú takmer rovnaké ako pre bunkuSI.

Po vyhodnotení všetkých štatistík pre bunkuSI si na záver môžeme pozrieť vykreslenie systémov úsečiek po skončení simulácie v porovnaní s bunkouSI (viď obr. A.5). Z výsledkov celej tejto podkapitoly očakávame, že nespozorujeme žiadny zásadný rozdiel. Skutočne, z obrázkov sú jednotlivé realizácie veľmi podobné v zmysle sledovaných charakteristík (počet a rozmiestnenie bodov, dĺžky úsečiek, orientácie).

5.3 Nehomogénna reálna bunka (bunka2)

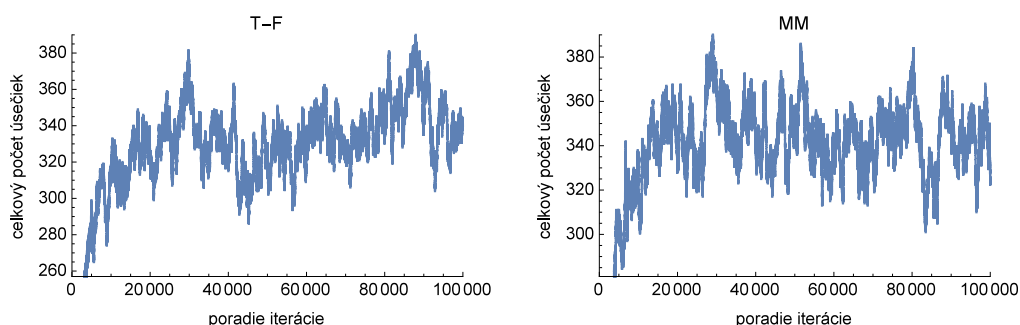
Na záver kapitoly s numerickými výsledkami ešte uvedieme niektoré výsledky pre reálnu bunku s očividne nehomogénnym rozmiestnením úsečiek. Pojmeme to stručnejšie ako pri predchádzajúcich dvoch bunkách a poukážeme najmä na miesta, kde simulácia najviac nekorešponduje s podkladovou realizáciou úsečiek bunky2. Bunka2 pozostáva zo 176 úsečiek a Lebesgueova miera okna je 53917,5 (takmer dvojnásobne väčšie okno ako pri bunke1). V tabuľke 5.11 sú zapísané výsledky pre odhad parametrov obomi metódami. Všimnime si, že odhady sú veľmi podobné a teda sa dá očakávať, že výsledky zo simulácií pre obe metódy sa nebudú výrazne odlišovať.

Tabuľka 5.11: Odhady parametrov (θ_1, θ_2, z) obidvomi metódami pre bunku2.

Parameter	T-F	MM
$\hat{\theta}_1$	-0,20982	-0,207983
$\hat{\theta}_2$	-0,315109	-0,332777
\hat{z}	1,33119	1,31246

S odhadnutými parametrami sme v programe spustili simuláciu a na obrázku A.9 v prílohe si môžeme všimnúť, ako výrazne sa realizácie po skončení simulácie odlišujú od pôvodného systému. Rozdiel nie je len v homogénnejšom rozmiestnení pri T-F a MM, ale najmä v počte úsečiek a v ich dĺžkach.

Zamerajme sa teraz na jednotlivé vlastnosti realizácií počas simulácie procesu rodenia a zániku. Na obrázku 5.17 vidíme vývoj počtu úsečiek počas simulácie. V oboch grafoch sa ich celkový počet pohybuje v páse 290 až 380.



Obr. 5.17: Graf vývoja celkového počtu úsečiek pre systémy úsečiek po každej iterácii procesu rodenia a zániku (vľavo pre T-F, vpravo pre MM) - bunka2

Priemery pre vývoj celkového počtu úsečiek, celkovú dĺžku realizácií a celkový počet priesečníkov v priebehu simulácie sme zhrnuli v tabuľke 5.12. Najvýznamnejší rozdiel je v počte úsečiek, kde nám priemerné hodnoty pre T-F aj MM vyšli

takmer dvojnásobok oproti počtu úsečiek v bunke2. V zvyšných dvoch riadkoch tabuľky by sa zdalo, že hodnoty pomerne dobre korešponujú s východzu realizáciou. Čo ale v tomto prípade znamená, že dve z troch celkových charakteristík, na ktorých je model postavený síce dobre súhlasia, ale pri dvojnásobnom počte úsečiek zo simulácie vyšla úplne iná realizácia, ako bol pôvodný model. Výsledky pre bunku2 nemôžeme považovať za uspokojivé, pre túto vzorku parametre odhadnuté obomi metódami nezachovávajú požadované vlastnosti modelu (obzvlášť počet úsečiek, a s ním súvisiaci celkovo iný zväľad výsledných realizácií simulácie).

Tabuľka 5.12: Priemery celkového počtu úsečiek, ich celkovej dĺžky a počtu priesečníkov počas simulácie pre bunku2.

Štatistika	T-F	MM	Podkladové dáta
Priemerný počet úsečiek	334,8	345,2	176
Priemerná celková dĺžka	6635	6847,8	6655,4
Priemerný celk. počet priesečníkov	74,7	79,7	76

Zmenu v rozdelení dĺžok oproti bunke2 výborne popisuje A.10 v obrázkovej prílohe. Pokým v histograme podkladových dát je ešte približne 30% úsečiek dlhších ako 40, v T-F a MM už takéto úsečky nepozorujeme vôbec.

Záver

V tejto práci sme zadefinovali proces úsečiek ako Gibbsov priestorový proces. Uviedli sme dve metódy odhadu parametrov modelu a to momentovú a Takacs-Fiksel, ktorá je známa pre model pretínajúcich sa kruhov. My sme ju modifikovali pre proces úsečiek a zjednotili značenie s momentovou metódou. S odhadnutými parametrami potom MCMC metódou simulujeme procesom rodenia a zániku model pre reálne aj simulované dáta.

Metódy odhadu a simuláciu sme implementovali do softvéru Mathematica, kde prebehli všetky výpočty a zároveň je v algoritme zahrnutá časť s výpočtom a vykreslením popisných štatistík na porovnanie kvality odhadov aj zhody modelu s podkladovými dátami. Algoritmus je univerzálny, spracuje rôzne podkladové dáta, ak sú dané v požadovanom formáte.

Výsledky máme pre reálne dáta - bunku1, bunku2 a pre simulované dáta (bunkuSI), ktoré vznikli z bunky1 vhodnou voľbou parametrov tak, aby bunkaSI zodpovedala modelu. Najlepšie výsledky, v zmysle parametrov a následnej simulácie modelu nám vyšli pre bunkuSI. V bunke1 sa po simulácii viac prejavili rozdiely oproti podkladovým dátam, či už v počte úsečiek alebo aj v počte priesečníkov systému. Za hlavnú príčinu považujeme nedostatočnú homogenitu dát a fakt, že sa jednalo o reálne dáta, ktoré nemusia úplne vyhovovať danému modelu.

Bunka2 je ukážkou, čo sa stane, keď použijeme algoritmus na nehomogénne dáta. Ukazuje sa, že táto požiadavka je veľmi podstatnou, pretože v bunke2 sa počet úsečiek simuláciou zdvojnásobil. Celková dĺžka úsečiek aj počet priesečníkov bol zachovaný, ale pri dvojnásobnom počte úsečiek. Tým sa úplne zmenil vzhľad výsledných realizácií po skončení simulácie s využitím odhadnutých parametrov. V tomto mieste vidíme priestor pre rozšírenie algoritmu v zmysle parametra z , ktorý by bolo vhodné namiesto konštanty zvoliť ako nejakú funkciu.

Zoznam použitej literatúry

- BADDELEY, A. (2007). Spatial point processes and their applications. In *Baddeley, A. a Barany, I. a Schneider, R. a Weil, W. (eds.): Stochastic Geometry. Lecture Notes in Mathematics, 1892*, pages 1–75. Springer Berlin Heidelberg.
- BENEŠ, V. a ZIKMUNDOVÁ, M. (2014). Functionals of spatial point processes having a density with respect to the poisson process. *Kybernetika*, **50**(6), 896–913.
- DALEY, D. J. a VERE-JONES, D. (1988). *An Introduction to the Theory of Point Processes*. Springer, Berlin Heidelberg New York.
- DEREUDRE, D., LAVANCIER, F. a HELISOVÁ, K. (2014). Estimation of the intensity parameter of the germ-grain Quermass-interaction model when the number of germs is not observed. *Scandinavian Journal of Statistics*, pages 1–21.
- ELTZNER, B., WOLLNIK, C., GOTTSCHLICH, C., HUCKEMANN, S. a REHLFELDT, F. (2014). A Filament Sensor for the Detection of Fiber Structures in Human Mesenchymal Stem Cells. *arXiv:1408.4002v1[cs.CV]*.
- GEORGII, H.-O. (1976). Canonical and grand canonical Gibbs states for continuum systems. *Commun. Math. Phys.*, **48**, 31–51.
- GEYER, C. a MØLLER, J. (1994). Simulation procedures and likelihood inference for spatial point process. *Scandinavian Journal of Statistics*, **21**(4), 359–373.
- KALLENBERG, O. (1983). *Random Measures*. Third edition. Akademie Verlag Academic Press, Berlin New York.
- KINGMAN, J. F. C. (1993). *Poisson Processes*. Oxford University Press, Oxford.
- MØLLER, J. a HELISOVÁ, K. (2008). Power diagrams and interaction processes for unions of disc. *Adv. Appl. Probab.*, **40**, 321–347.
- NGUYEN, X. X. a ZESSIN, H. (1979). Integral and differential characterizations of the Gibbs process. *Math. Nach.*, **88**, 105–115.
- SCHNEIDER, R. a WEIL, W. (2008). *Stochastic and Integral Geometry*. Springer, Berlin.
- STOYAN, D., KENDALL, W. a MECKE, J. (1995). *Stochastic Geometry and its Applications*. Second edition. John Wiley and Sons, Chichester.

WOLFRAM RESEARCH, INC. (2014). *Mathematica*. Version 10.0. Wolfram Research, Inc., Champaign, Illinois.

ZIKMUNDOVÁ, M. (2014). *Interacting spatial particle systems*. Doktorská dizertační práce, MFF UK Praha.

Zoznam obrázkov

4.1	Fluorescentne nasvietená kmeňová bunka (vľavo) a tá istá bunka so zvýraznením úsečiek detekovaných FS algoritmom (vpravo). . .	26
5.1	Bodový graf stredov úsečiek bunky1	30
5.2	Histogramy rozdelenia pravdepodobnosti dĺžok (vľavo) a orientácií (vpravo), z ktorých vychádzame pri generovaní úsečiek - bunka1 .	31
5.3	Graf vývoja celkového počtu úsečiek pre systémy úsečiek po každej iterácii procesu rodenia a zániku (vľavo pre T-F, vpravo pre MM) - bunka1	32
5.4	Histogramy počtu úsečiek v priebehu simulácie pre parametre odhadnuté obidvomi metódami - bunka1.	33
5.5	Graf vývoja celkovej dĺžky úsečiek pre systémy úsečiek po každej iterácii procesu rodenia a zániku (vľavo pre T-F, vpravo pre MM) - bunka1	34
5.6	Histogramy celkovej dĺžky úsečiek v priebehu simulácie pre parametre odhadnuté obidvomi metódami - bunka1.	35
5.7	Graf vývoja celkového počtu priesečníkov pre systémy úsečiek po každej stej iterácii procesu rodenia a zániku (vľavo pre T-F, vpravo pre MM) - bunka1	35
5.8	Histogramy celkových počtov priesečníkov v priebehu simulácie pre parametre odhadnuté obidvomi metódami - bunka1.	36
5.9	Bodový graf stredov úsečiek bunkySI	38
5.10	Histogramy rozdelenia pravdepodobnosti dĺžok (vľavo) a orientácií (vpravo), z ktorých vychádzame pri generovaní úsečiek pri bunkeSI.	38
5.11	Graf vývoja celkového počtu úsečiek pre systémy úsečiek po každej iterácii procesu rodenia a zániku (vľavo pre T-F, vpravo pre MM) - bunkaSI	39
5.12	Histogramy počtu úsečiek v priebehu simulácie pre parametre odhadnuté obidvomi metódami (bunkaSI).	40
5.13	Graf vývoja celkovej dĺžky úsečiek pre systémy úsečiek po každej iterácii procesu rodenia a zániku (vľavo pre T-F, vpravo pre MM) v bunkeSI	41
5.14	Histogramy celkovej dĺžky úsečiek v priebehu simulácie pre parametre odhadnuté obidvomi metódami - bunkaSI	41
5.15	Graf vývoja celkového počtu priesečníkov pre systémy úsečiek po každej stej iterácii procesu rodenia a zániku (vľavo pre T-F, vpravo pre MM) - bunkaSI	42

5.16	Histogramy celkových počtov priesečníkov v priebehu simulácie pre parametre odhadnuté obidvomi metódami- bunkaSI	43
5.17	Graf vývoja celkového počtu úsečiek pre systémy úsečiek po každej iterácii procesu rodenia a zániku (vľavo pre T-F, vpravo pre MM) - bunka2	44
A.1	Vykreslenie systémov úsečiek po poslednej iterácii simulácie v porovnaní s pôvodnou realizáciou bunky1	54
A.2	Histogramy počtu priesečníkov jednotlivých úsečiek so zvyškom systému po skončení simulácií v porovnaní s histogramom pre podkladové dáta bunky1.	54
A.3	Histogramy rozdelení dĺžok zhora postupne pre realizáciu po 60-tisíc iteráciách, 80-tisíc a pre poslednú realizáciu v porovnaní s rozdelením dĺžok úsečiek bunky1	55
A.4	Histogramy rozdelení orientácií zhora postupne pre realizáciu po 60-tisíc iteráciách, 80-tisíc a pre poslednú realizáciu v porovnaní s rozdelením orientácií úsečiek bunky1	56
A.5	Vykreslenie systémov úsečiek po poslednej iterácii simulácie v porovnaní s pôvodnou realizáciou bunkySI	57
A.6	Histogramy počtu priesečníkov jednotlivých úsečiek so zvyškom systému po skončení simulácií v porovnaní s histogramom pre podkladové dáta bunkySI.	57
A.7	Histogramy rozdelení dĺžok zhora postupne pre realizáciu po 60-tisíc iteráciách, 80-tisíc a pre poslednú realizáciu v porovnaní s rozdelením dĺžok úsečiek bunkySI	58
A.8	Histogramy rozdelení orientácií zhora postupne pre realizáciu po 60-tisíc iteráciách, 80-tisíc a pre poslednú realizáciu v porovnaní s rozdelením orientácií úsečiek bunkySI	59
A.9	Vykreslenie systémov úsečiek po poslednej iterácii simulácie v porovnaní s pôvodnou realizáciou bunky2	60
A.10	Histogramy rozdelení dĺžok pre poslednú realizáciu v porovnaní s rozdelením dĺžok úsečiek bunky2	60

Zoznam tabuliek

5.1	Základné vlastnosti popisujúce bunku1	31
5.2	Odhady parametrov (θ_1, θ_2, z) obidvomi metódami pre bunku1. . .	32
5.3	Priemer a smerodajná odchýlka celkového počtu úsečiek zo simulácií pre bunku1.	33
5.4	Priemer a smerodajná odchýlka celkovej dĺžky úsečiek pre bunku1.	34
5.5	Priemer a smerodajná odchýlka celkového počtu priesečníkov v realizácii úsečiek počas simulácie pre bunku1.	35
5.6	Základné vlastnosti popisujúce bunkuSI	38
5.7	Odhady parametrov (θ_1, θ_2, z) obidvomi metódami pre bunkuSI. .	39
5.8	Priemer a smerodajná odchýlka celkového počtu úsečiek pre bunkuSI.	40
5.9	Priemer a smerodajná odchýlka celkovej dĺžky úsečiek pre bunkuSI.	40
5.10	Priemer a smerodajná odchýlka celkového počtu priesečníkov v realizácii úsečiek počas simulácie pre bunkuSI.	42
5.11	Odhady parametrov (θ_1, θ_2, z) obidvomi metódami pre bunku2. . .	44
5.12	Priemery celkového počtu úsečiek, ich celkovej dĺžky a počtu priesečníkov počas simulácie pre bunku2.	45

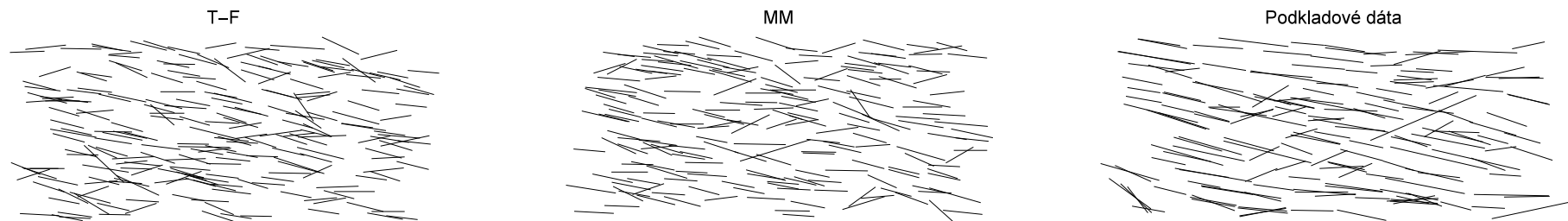
Zoznam použitých skratiek

- T-F - označenie metódy Takacs-Fiksel pre odhad parametrov procesu úsečiek
- MM - skratka pre momentovú metódu odhadu parametrov procesu úsečiek
- bunka1 - realizácia približne homogénne rozmiestnených úsečiek vybraných z reálnej kmeňovej bunky
- bunka2 - realizácia nehomogénne rozmiestnených úsečiek vybraných z inej reálnej kmeňovej bunky ako pri bunke1
- bunkaSI - pracovný názov pre simulované dáta, teda simulovanú „bunku„
- KB - skratka pre ľudskú kmeňovú bunku
- FS - Filament Sensor algoritmus na detekovanie stresových vlákien v kmeňovej bunke

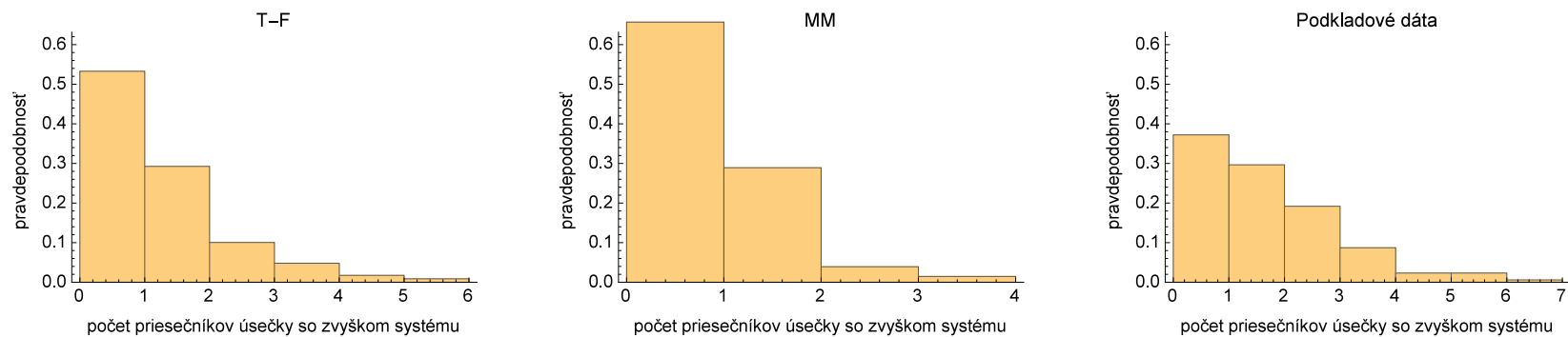
Dodatok A

Obrázkové prílohy

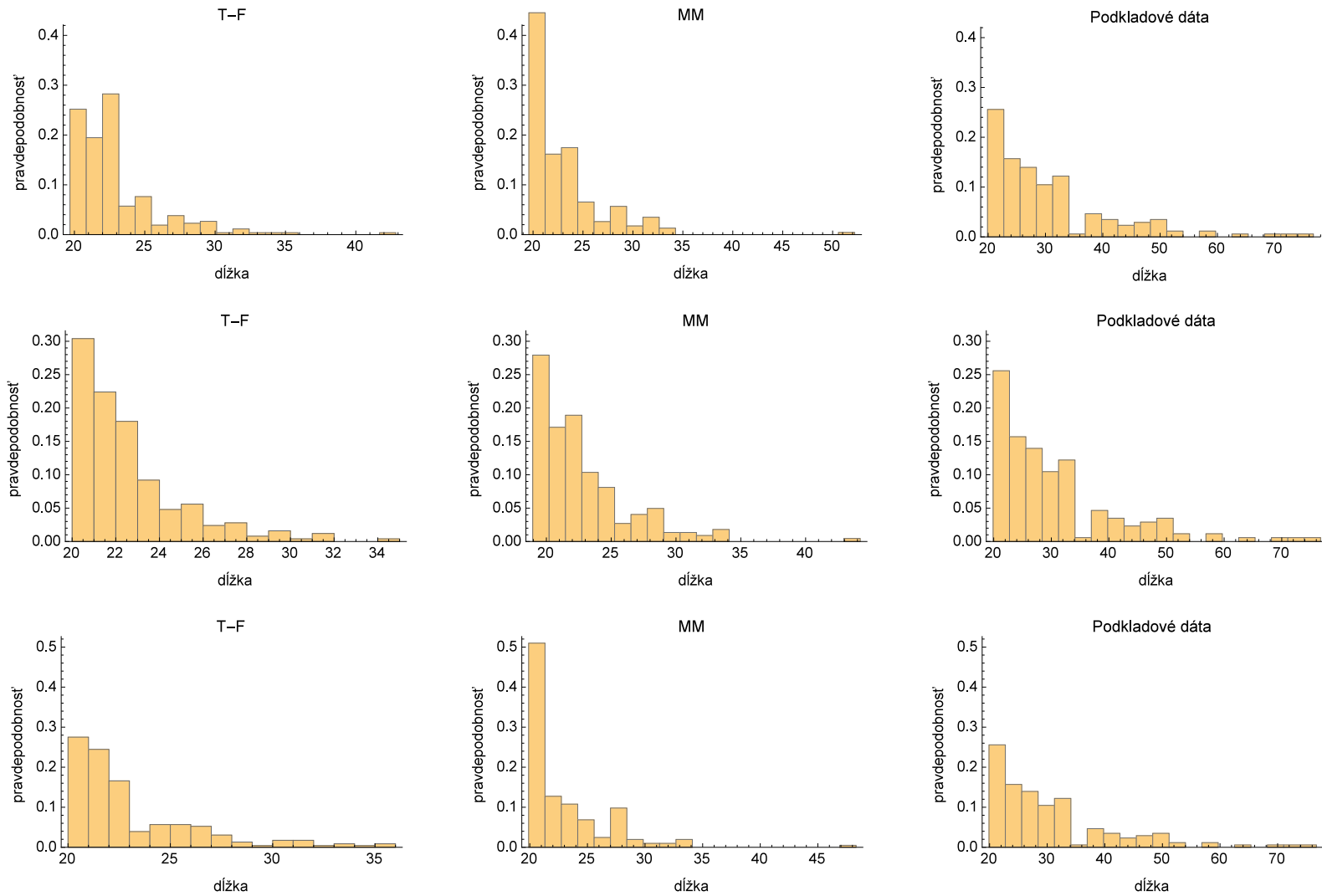
V tejto časti práce uvidíme niekoľko obrazových príloh, ktoré obsahovo prislúchajú ku kapitole 5. V rámci textu sa na obrázky odkazujeme a zvyčajne sú tu umiestnené tri grafy, prípadne histogramy vedľa seba. Ich výpovedná hodnota je podstatne vyššia, keď sú za sebou v rade a tým pádom je kvôli veľkosti nutné ich mať otočené o 90° . Je vhodné sledovať obrázky popri čítaní textu, najskôr sú zaradené obrázky prislúchajúce k bunke1, potom obrázky k bunkeSI a na záver dva obrázky k bunke2.



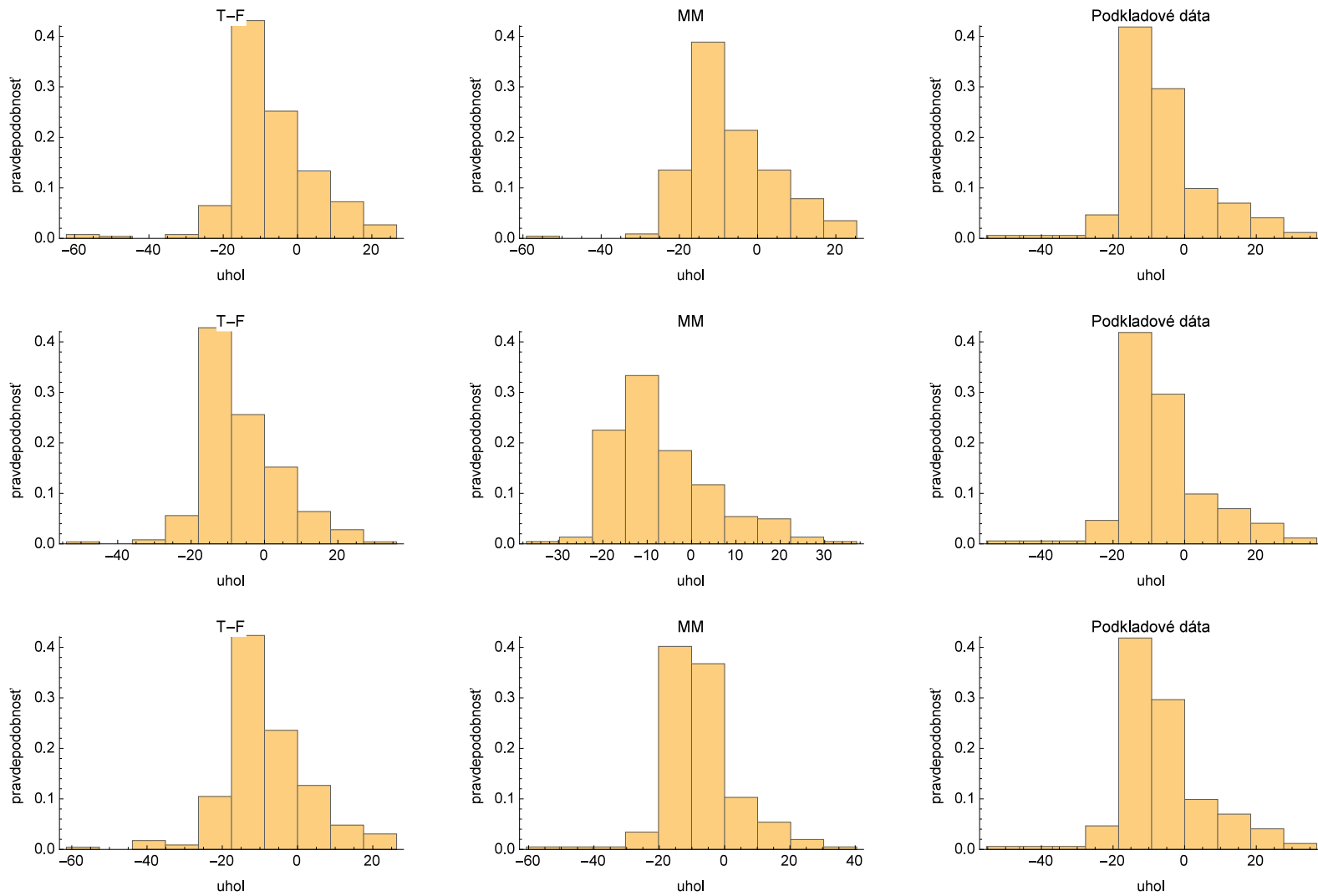
Obr. A.1: Vykreslenie systémov úsečiek po poslednej iterácii simulácie v porovnaní s pôvodnou realizáciou bunky1



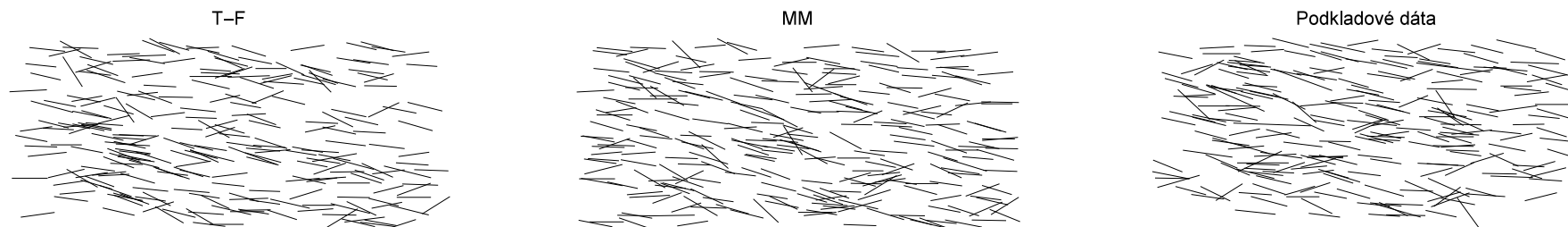
Obr. A.2: Histogramy počtu priesečníkov jednotlivých úsečiek so zvyškom systému po skončení simulácií v porovnaní s histogramom pre podkladové dáta bunky1.



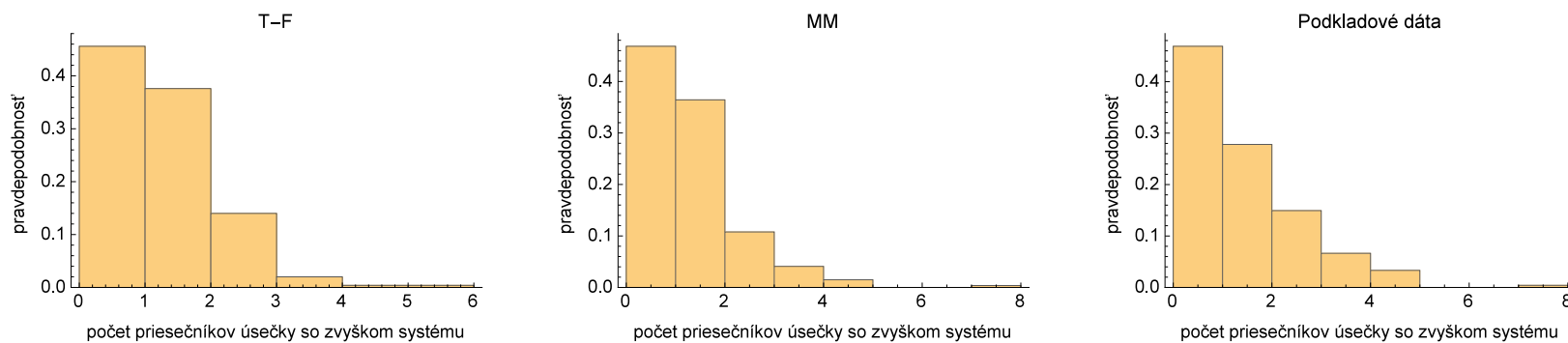
Obr. A.3: Histogramy rozdelení dĺžok zhora postupne pre realizáciu po 60-tisíc iteráciách, 80-tisíc a pre poslednú realizáciu v porovnaní s rozdelením dĺžok úsečiek bunky1



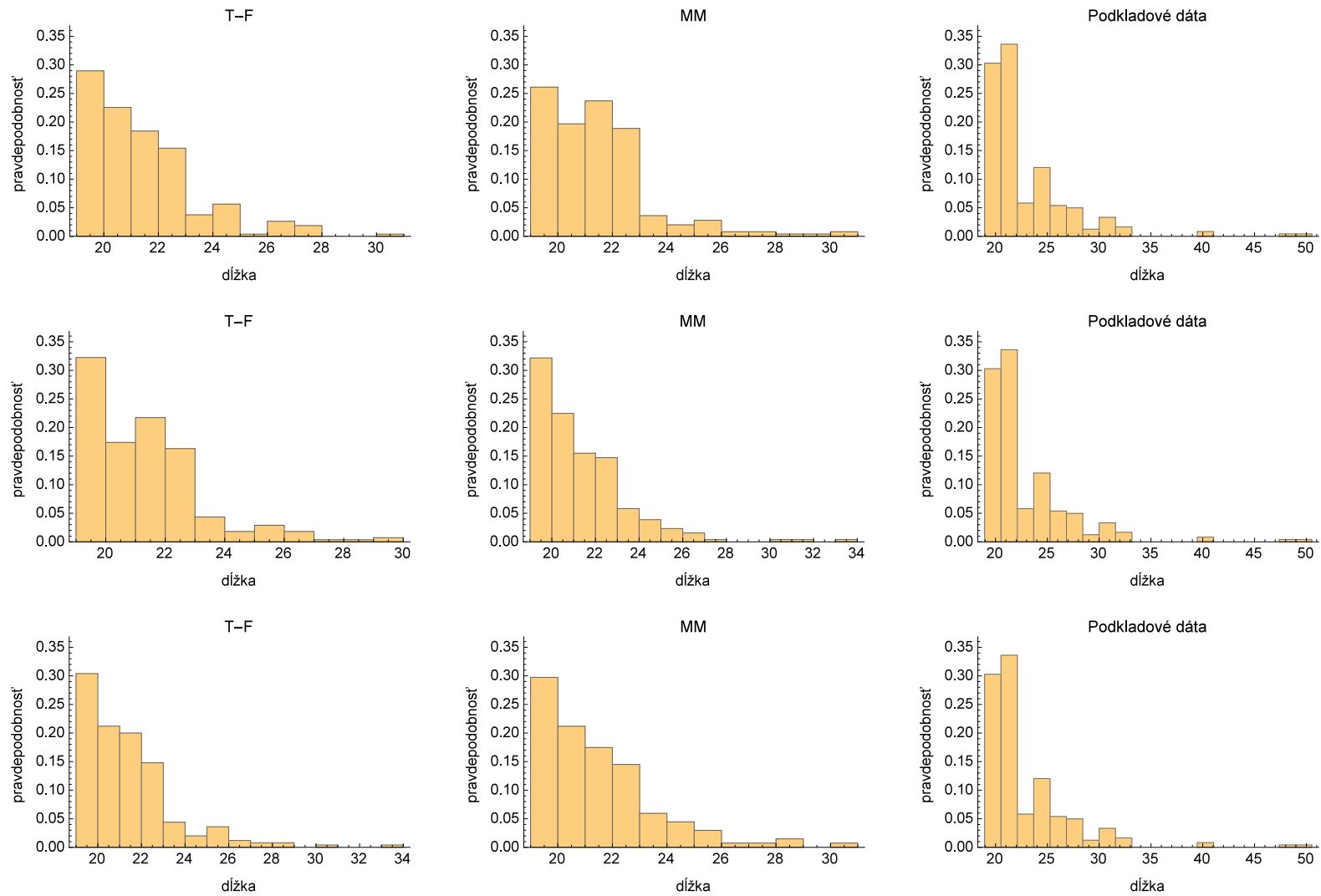
Obr. A.4: Histogramy rozdelení orientácií zhora postupne pre realizáciu po 60-tisíc iteráciách, 80-tisíc a pre poslednú realizáciu v porovnaní s rozdelením orientácií úsečiek bunky1



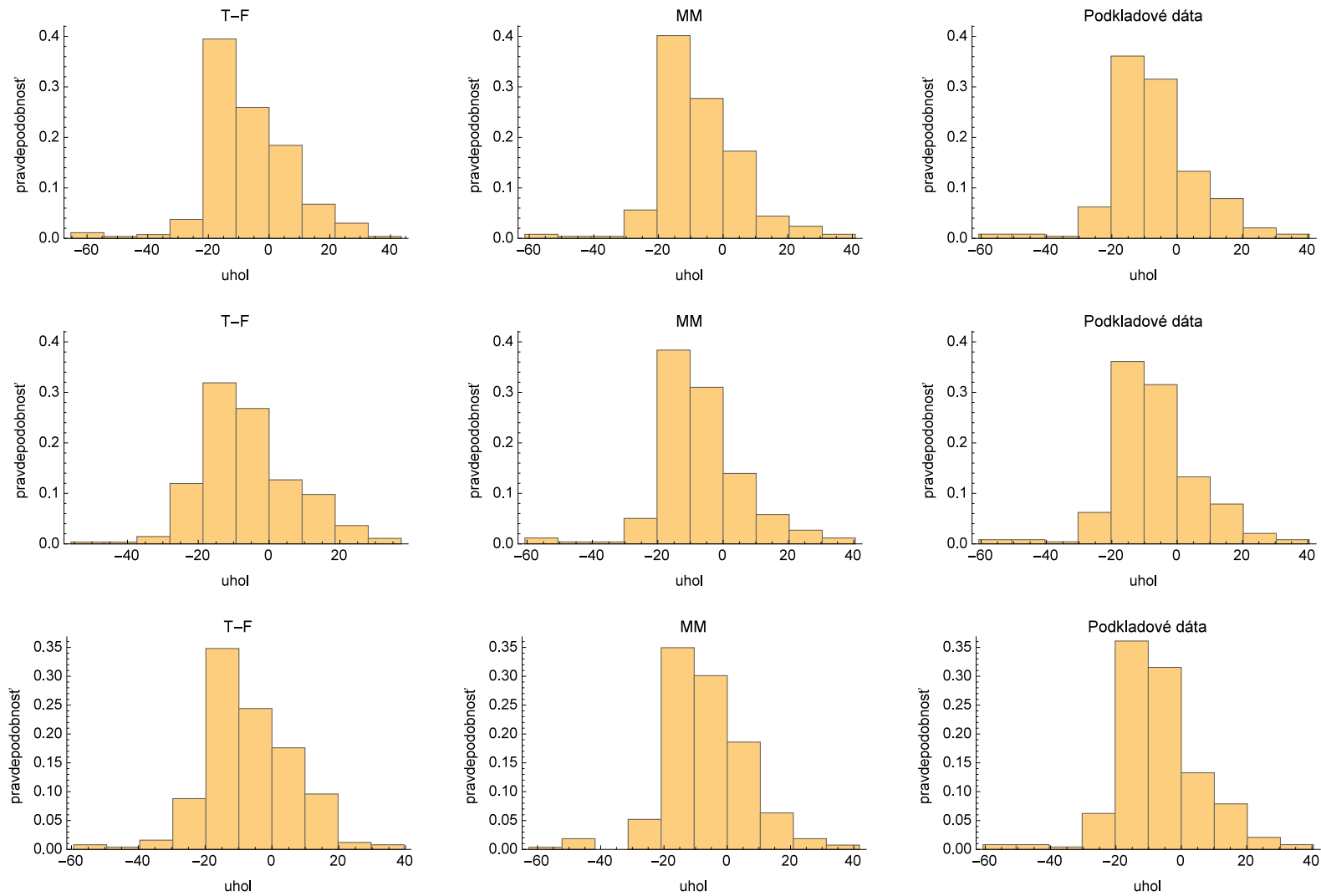
Obr. A.5: Vykreslenie systémov úsečiek po poslednej iterácii simulácie v porovnaní s pôvodnou realizáciou bunkySI



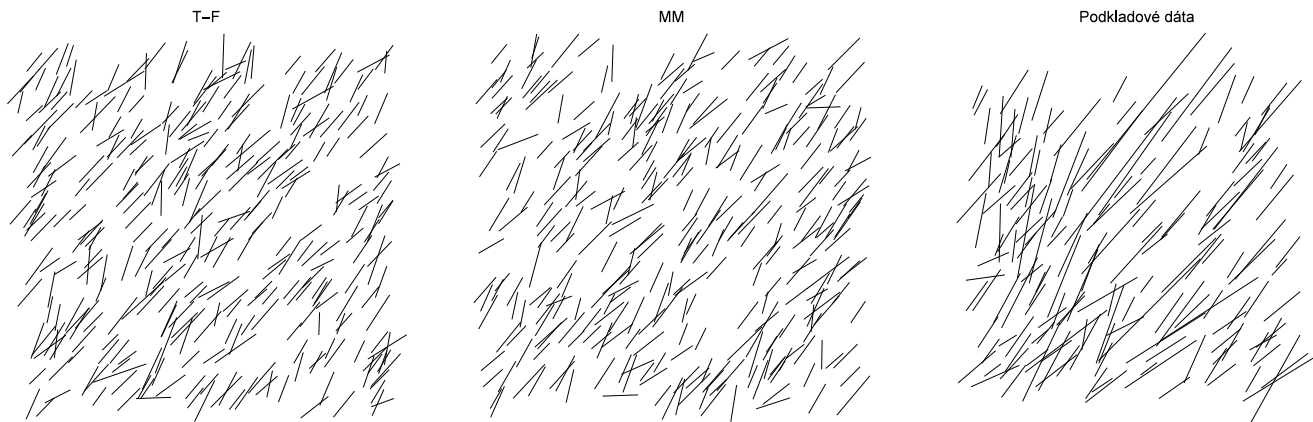
Obr. A.6: Histogramy počtu priesečníkov jednotlivých úsečiek so zvyškom systému po skončení simulácií v porovnaní s histogramom pre podkladové dáta bunkySI.



Obr. A.7: Histogramy rozdelení dĺžok zhora postupne pre realizáciu po 60-tisíc iteráciách, 80-tisíc a pre poslednú realizáciu v porovnaní s rozdelením dĺžok úsečiek bunkySI

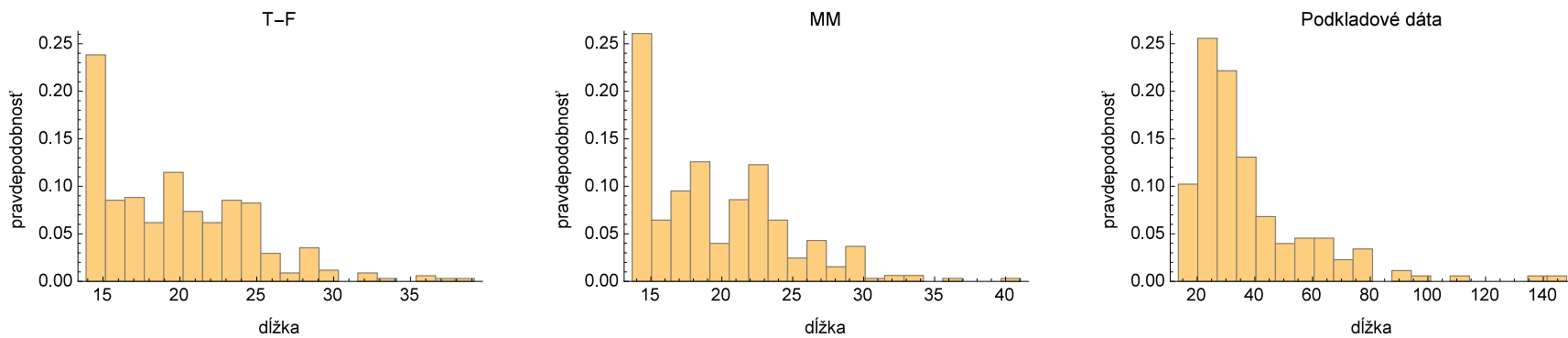


Obr. A.8: Histogramy rozdelení orientácií zhora postupne pre realizáciu po 60-tisíc iteráciách, 80-tisíc a pre poslednú realizáciu v porovnaní s rozdelením orientácií úsečiek bunkySI



Obr. A.9: Vykreslenie systémov úsečiek po poslednej iterácii simulácie v porovnaní s pôvodnou realizáciou bunky2

09



Obr. A.10: Histogramy rozdelení dĺžok pre poslednú realizáciu v porovnaní s rozdelením dĺžok úsečiek bunky2

Dodatok B

Obsah priloženého CD

Na priloženom CD sú uložené Mathematica notebooky prislúchajúce k jednotlivým analyzovaným bunkám. Máme ich pod názvami *bunka1.ntb*, *bunka2.ntb* a *bunkaSI.ntb*. Na CD sú taktiež zdrojové MS Excel súbory s názvami *bunka1_orezana_na_172bodov.xlsx* a *bunka2_orezana_na_176bodov.xlsx*.

Jednotlivé notebooky obsahujú celý algoritmus pre odhadovanie parametrov, následnú simuláciu procesom rodenia a zániku s odhadnutými parametrami a aj všetky popisné štatistiky uvádzané v texte práce.