



**MATEMATICKO-FYZIKÁLNÍ
FAKULTA**
Univerzita Karlova

BAKALÁŘSKÁ PRÁCE

Willy Svoboda

Bodové procesy na sféře

Katedra pravděpodobnosti a matematické statistiky

Vedoucí práce: RNDr. Jiří Dvořák, Ph.D.
Studijní program: Matematika
Studijní obor: Obecná matematika

Praha 2022

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

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

V dne

Podpis autora

Velké poděkování za odborné vedení mé bakalářské práce patří RNDr. Jiřímu Dvořákovi, Ph.D., za všechny profesionální rady, které byly pro mě velice užitečné, a v neposlední řadě děkuji za všechnu trpělivost a vstřícnost, se kterými jsem se setkával po celou dobu zpracovávání mé práce.

Na tomto místě bych chtěl také poděkovat z celého srdce mé mamce za neustálou podporu nejen při mém studiu.

Název práce: Bodové procesy na sféře

Autor: Willy Svoboda

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Jiří Dvořák, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Bodový proces lze jednoduše popsat jako náhodnou lokálně konečnou množinu. Můžeme tak modelovat polohu stromů v lese, místa hlášení událostí ve městě nebo ve světě, například polohy epicenter zemětřesení. V práci zavedeme základní typy bodových procesů v euklidovském prostoru a na sféře, podrobněji vysvětlíme, jaké situace lze modelovat pomocí bodových procesů, uvedeme definici základních vlastností a položíme teoretický základ pro zavedení K -funkce (a její modifikace pro kótované bodové procesy). Hlavním přínosem práce je zavedení kótovaných bodových procesů na sféře a položení teoretického základu, přičemž kóty bodů nám budou přinášet další netriviální informaci o bodech, kterou chceme zkoumat. Na konci práce se budeme věnovat testování hypotézy, zda kóty jsou nezávislé, přičemž testujeme pomocí permutačního Monte Carlo testu využívající kótami převáženou K -funkci pro kótované bodové procesy na sféře.

Klíčová slova: bodové procesy, K -funkce, sféra, kótovaný bodový proces na sféře, kóta, nezávislost kót, kótami převážená K -funkce

Title: Point processes on the sphere

Author: Willy Svoboda

Department: Department of probability and mathematical statistics

Supervisor: RNDr. Jiří Dvořák, Ph.D., Department of probability and mathematical statistics

Abstract: A point process can be easily described as a random locally finite set. For example, we can model locations of arbitrary events in a city or in the world such as earthquake epicenter locations. In this thesis, we introduce basic types of point processes in a Euclidean space and on a sphere, describe what situations can be modelled by them, define basic properties, and lay down theoretical groundwork for the K -function (and its modifications for marked point processes). The main goal of this thesis is to introduce marked point processes on a sphere and to give theoretical framework, whereas the marks will give us another nontrivial information about the points, which we want to study further. In the conclusion of the thesis, we concern ourselves with testing whether those marks are mutually independent. We apply Monte Carlo permutation test using mark-weighted K -function for marked point processes on a sphere.

Keywords: point processes, K -function, sphere, marked point process on a sphere, mark, independence of marks, mark-weighted K -function

Obsah

Úvod	2
1 Bodové procesy v euklidovském prostoru	3
1.1 Motivace	3
1.2 Definice a vlastnosti	4
1.3 K -funkce	10
1.3.1 Definice	10
1.3.2 Odhad	11
2 Kótované bodové procesy v euklidovském prostoru	12
2.1 Motivace	12
2.2 Druhy kót	13
2.3 Budování základní teorie	17
2.4 Kótami převážená K -funkce	18
3 Bodové procesy na sféře	20
3.1 Motivace	20
3.2 Definice a vlastnosti	20
3.3 K -funkce	26
4 Kótované bodové procesy na sféře	28
4.1 Motivace	28
4.2 Definice a vlastnosti	29
4.3 K -funkce	34
4.4 Testování nezávislosti kót	34
4.4.1 Hypotéza a testová statistika	34
4.4.2 Testování hypotézy	35
4.4.3 Realizace	36
Závěr	42
Seznam použité literatury	43
Seznam obrázků	44
A Přílohy	45

Úvod

Bodový proces lze jednoduše popsat jako náhodnou lokálně konečnou množinu. Můžeme pomocí nich modelovat například polohu stromů v lese nebo místa hlášení nějakých událostí ve městě.

Vlastní přínos práce najdeme hlavně ve čtvrté kapitole s názvem Kótované bodové procesy na sféře, avšak je potřeba vybudovat teorii a základní intuici s dostatečným pochopením pojmů v základních případech, kterými se budeme postupně zabývat v prvních třech kapitolách. Tedy dalším přínosem je shrnutí potřebných vlastností a poznatků jednotlivých druhů bodových procesů potřebných k položení základů pro čtvrtou kapitolu a celkové seznámení s tímto tématem.

V první kapitole zavedeme bodové procesy v euklidovském prostoru, kde se seznámíme se základními pojmy a vysvětlíme si základní myšlenky zkoumání bodových procesů.

Následně v druhé kapitole rozšíříme naše znalosti o kótované bodové procesy, kdy jednotlivé body nesou netriviální informaci o sobě, tedy bodové procesy nám v takovém případě umí poskytnout mnohem více informací a můžeme provádět pokročilejší zkoumání.

Třetí kapitola bude tvořit teoretický přechod mezi bodovými procesy v euklidovském prostoru a vlastní prací autora, kótovanými bodovými procesy na sféře. Tato kapitola bude konkrétně zahrnovat adaptaci (nekótovaných) bodových procesů na sféru.

Ve čtvrté kapitole vybudujeme základy pro kótované bodové procesy na sféře. Vyslovíme hypotézu o nezávislých kótách, kde budeme testovat hypotézu pomocí simulací. Testovou statistikou bude kótami převážená K -funkce, proto si v prvních třech kapitolách zavedeme její analogické verze (v první a třetí kapitole se nepracuje s kótami, proto se funkce nazývá pouze K -funkce).

Práce se zaměřuje na vybudování většího porozumění bodových procesů a za tímto účelem se zde nachází často velice podrobné odstavce a více příkladů i s ilustracemi. Je tedy určena i čtenáři bez větších předchozích znalostí bodových procesů.

1. Bodové procesy v euklidovském prostoru

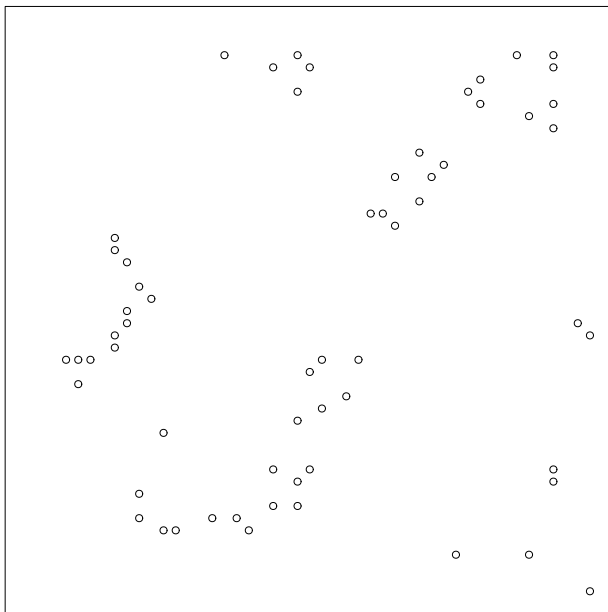
1.1 Motivace

Než si uvedeme definici bodových procesů, začneme s motivací a několika příklady. Představme si náhodnou lokálně konečnou množinu v euklidovském prostoru, pro jednoduchost v \mathbb{R}^2 . Ať si vybereme libovolnou kompaktní podmnožinu této roviny, vždy tam najdeme konečný počet bodů, kde každý bod může zastupovat určitou věc z reálného života. U takového seskupení nás může zajímat nějaká celková charakteristika, jako je nejmenší vzdálenost k dalšímu bodu, zda dochází někde k seskupování bodů a jinde naopak existuje významně velká oblast bez bodů a podobně. Než se budeme zabývat pouze případy v euklidovském prostoru, uvedeme si pro ilustraci několik příkladů z reálného života, které lze řešit nebo zkoumat pomocí bodových procesů. Ilustrativně zde uvedeme i situaci bodového procesu na sféře, i když se bodovým procesům na sféře budeme podrobněji věnovat ve 3. kapitole:

Příklad 1.1: V kalifornském parku rostou sekvoje obrovské a jejich poloha je zobrazena na obrázku 1.1 pomocí jednotlivých bodů. Jedná se o reálná data veřejně dostupná v balíku `spatstat` u programu `R`, u kterých vidíme, že došlo k vytvoření skupinek stromů. V takovém případě budeme říkat, že dochází ke shlukování. K němu dochází v přírodě velice často už jen z principu rozmnožování některých rostlin, kdy se dceřiné rostliny nachází v blízkosti mateřské rostliny.

Příklad 1.2: Na Měsíci si lze všimnout mnoha kráterů různých velikostí. Přírozenou otázkou může být, zda jsou krátery na povrchu Měsíce rozloženy rovnoměrně, či zda na některém místě dochází k nějaké anomálii. Ve čtvrté kapitole můžeme zkoumat i pokročilejší otázky týkající se velikostí kráterů, například zda velikost kráteru závisí na poloze kráteru.

Příklad 1.3: Představme si město, ve kterém každé ráno vyráží pouliční obchodníci do ulic vyhlížející vhodné místo pro postavení stánku. Vybírání místa může být mnohem složitější záležitost, než se na první pohled zdá. I když se obchodník postaví na stejné místo jako předešlý den, nemusí to pro něho být výhodné buď z důvodu malých zisků v dané lokalitě, či že se v blízkém okolí toho dne nachází už dostatek jiných konkurenčních stánků a pro dalšího stánkaře je tak nalákání zákazníků právě ke svému stánku ztlačně obtížnější. Polohy všech stánků se tak každý den mohou měnit. Nelze přitom očekávat nějaké rovnoměrné rozprostření po městě, neboť jsou ve městě ulice, kterými proudí velké množství lidí, a může se tam uživit větší počet stánkařů než v místech, kde intenzita pohybu lidí je citelně tak malá, že přítomnost jiného obchodníka v okolí by prakticky znamenala nedostatečný denní výdělek. Zahuštění jednotlivých částí města stánkaři bude ztlačně záviset na intenzitě pohybu lidí danými místy.



Obrázek 1.1: Poloha sekvojí obrovských v kalifornském parku nám poukazuje na shlukování, tedy vytváření míst s větší hustotou bodů než v jiných částech obrázku.

1.2 Definice a vlastnosti

Po motivační části si nyní zavedeme bodový proces, k čemuž ovšem budeme potřebovat několik pomocných definic. Dále si vysvětlíme důležitou vlastnost některých bodových procesů, kterou je stacionarita, protože ji budeme v dalších částech práce potřebovat. Pomocí ní dojde k zjednodušení funkce intenzity, další z vlastností bodových procesů, na nezáporné reálné číslo a bude naším důležitým nástrojem pro zavedení Palmova rozdělení a střední hodnoty vzhledem k Palmovu rozdělení, které potřebujeme v další podkapitole. Ovšem než ukončíme tuto podkapitolu, budeme muset ještě probrat, co je to okrajový efekt a uvedeme jednu z možných korekcí, konkrétně mínusovou korekci. Nyní pojďme zavést bodové procesy.

V této a v 2. kapitole většinu definic a tvrzení přejímáme ze zdroje (1) a (2). Budeme pracovat s euklidovskými prostory \mathbb{R}^d , proto v těchto kapitolách budeme uvažovat d jako přirozené číslo.

Mějme úplný separabilní metrický prostor (Y, ρ) . Pro borelovské množiny budeme používat značení $\mathcal{B}(Y)$ a pro omezené borelovské množiny budeme používat značení $\mathcal{B}_0(Y)$. Pokud bude zřejmé, o jaký prostor se jedná, budeme zkracovat značení na \mathcal{B} , respektive \mathcal{B}_0 .

Definice 1. *Míra μ na (Y, \mathcal{B}) je **lokálně konečná**, jestliže je konečná na \mathcal{B}_0 . Symbolem $\mathcal{M} \equiv \mathcal{M}(Y)$ budeme značit množinu všech lokálně konečných měr na (Y, \mathcal{B}) a symbolem $\mathcal{M}_b \equiv \mathcal{M}_b(Y)$ množinu všech konečných měr na (Y, \mathcal{B}) . Dále označme*

$$\mathcal{N} \equiv \mathcal{N}(Y) = \{\mu \in \mathcal{M} : \mu(B) \in \mathbb{N} \cup \{0, \infty\} \text{ pro každou } B \in \mathcal{B}\}$$

množinu všech lokálně konečných měr nabývajících pouze celočíselných hodnot.

Definice 2. Zavedme následující σ -algebry na \mathcal{M} a \mathcal{N} :

$$\begin{aligned}\mathfrak{M} &= \sigma\{\mu \mapsto \mu(B) \text{ měřitelné, } B \in \mathcal{B}\}, \\ \mathfrak{N} &= \{M \cap \mathcal{N} : M \in \mathfrak{M}\}.\end{aligned}$$

\mathfrak{M} je nejmenší sigma-algebra měr na borelovských množinách, pro kterou platí, že všechny projekce $\mu \mapsto \mu(B)$ jsou měřitelná zobrazení. Její stopa v množině všech lokálně konečných měr nabývajících pouze celočíselných hodnot je druhou zadanou σ -algebrou \mathfrak{N} .

Definice 3. Bud' (Ω, Σ, P) pravděpodobnostní prostor. **Bodový proces** na Y je měřitelné zobrazení

$$X : (\Omega, \Sigma, P) \rightarrow (\mathcal{N}, \mathfrak{N}).$$

Bodový proces je **jednoduchý**, jestliže $P(X \in \mathcal{N}^*) = 1$, kde

$$\mathcal{N}^* = \{\nu \in \mathcal{N} : \nu(\{x\}) \leq 1 \text{ pro každé } x \in Y\}.$$

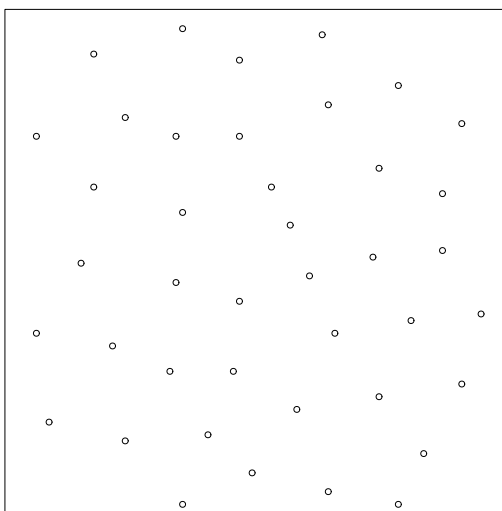
Poznámka: V celé práci budeme mlčky uvažovat jednoduché bodové procesy (nejen u těch dosud zavedených, ale analogicky lze tuto vlastnost definovat i u pozdějších typů bodových procesů). Tedy bez explicitního vyjádření budeme předpokládat nulovou pravděpodobnost, že dva body bodového procesu budou mít naprosto shodnou polohu v \mathbb{R}^d (respektive na sféře pro 3. a 4. kapitolu).

Nyní máme zavedené bodové procesy a můžeme si zavést důležitou vlastnost, stacionaritu. Vysvětlíme si rozdíl mezi stacionárním bodovým procesem a situací, kdy není stacionární.

Definice 4. Bodový proces X nazveme **stacionárním**, pokud jeho rozdělení je invariantní vzhledem k translaci v \mathbb{R}^d , to znamená rozdělení posunutého bodového procesu $X + y = \{x + y : x \in X\}$ je stejné jako rozdělení X pro všechna $y \in \mathbb{R}^d$.

Pokud bodový proces nebude stacionární, budeme ho v této práci označovat jako nestacionární. Pro porozumění rozdílu mezi stacionárním a nestacionárním bodovým procesem nám pomůže následující text a ilustrace.

Jeden z často používaných příkladů stacionárního bodového procesu jsou reálná data veřejně dostupná v balíčku `spatstat` obsahující polohu středů 42 buněk pozorovaných mikroskopem v histologickém řezu. Pozorovací okno dané pozorováním přes mikroskop bylo na našem obrázku 1.2 přeskálováno do jednotkového čtverce. Můžeme si všimnout pravidelné struktury, nikde nedochází k výraznému shlukování ani k jiným nepravidlostem. Představme si, že máme jistý bodový proces v euklidovském prostoru. Pokud bude tento bodový proces stacionární, tak i když bychom jakkoliv posunuli pozorovací okno, stále bychom měli vidět stejné vlastnosti, jaké jsme pozorovali u původního pozorovacího okna. Nemusíme tedy řešit jakou část bodového procesu pozorujeme, protože při posunutí pozorovacího okna by se naše informace o bodovém procesu neměla změnit, pokud je opravdu stacionární.

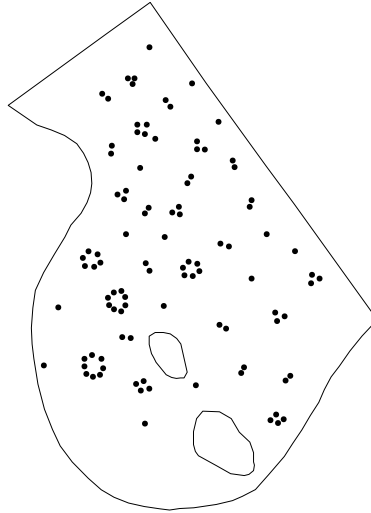


Obrázek 1.2: Poloha středů buněk pozorovaných pod mikroskopem.

Jiná situace nastane u nestacionárního bodového procesu. Vlastnosti bodového procesu se budou měnit v závislosti na poloze. Jednoduchým příkladem může být libovolný větší les. Pokud se postavíme ke stromu na úplném okraji lesa, tak v okolí 50 metrů od tohoto stromu napočítáme zřejmě mnohem méně sousedních stromů než u stromu, který se nachází někde uprostřed lesa (bez přítomnosti mýtiny či jiné okolnosti omezující počet stromů v jeho okolí), neboť strom na kraji lesa bude obklopen stromy typicky jen z některých světových stran.

Jiný příklad obsahující přímo data z reálného života (opět balíček `spatstat`) je poněkud úsměvný, ale lze na něm krásně ilustrovat nestacionární bodový proces. Jednoho slunečného dne ve 3 hodiny odpoledne byla zaznamenána poloha lidí usazených na trávniku ve veřejném parku Gordon Square v Londýně. Lze si jednoduše představit, že pokud se dostaví do parku skupinka lidí, která se spolu zná, budou mít tendenci si sednout blízko k sobě, typicky do kroužku. Pokud ale do parku přijde návštěvník, který nikoho jiného v parku nebude znát, nejspíše si radši sedne s dostatečným odstupem od ostatních v rámci možností volného prostoru. Poloha stávajících lidí tedy ovlivňuje polohu dalších nových lidí a dostáváme tak bodový proces, kde body jsou v nějaké závislosti vůči sobě. Situace s reálnými daty je vyobrazena na obrázku 1.3, kde lze vidět nepravidelný tvar trávniku, protože je vymezen cestami vedoucími parkem. Dva ostrůvky na obrázku jsou ve skutečnosti okrasné záhony.

Představme si rozšíření parku do celé roviny. V takovém „nekonečném parku“ budeme mít plno různých objektů, neboť parky nejsou na celém území monotónní a už na tom malém území zobrazeném na obrázku se nachází záhony, které ovlivňují polohu, kam si lidé sednou. Také byl slunný den a někteří lidé dávají přednost si sednout na místo, které je zastíněné okolními stromy. Dalšími faktory, které ovlivní v nekonečném parku místa usazení, budou ostatní objekty obvyklé v parku jako cesty, potůčky, jezírka (která ale naopak mohou přilákat za slunečného dne do svého okolí více lidí) a podobně. Když tedy budeme mít pozorovací okno v tomto nekonečném parku, rozložení lidí v něm bude záviset právě na tom, jaké objekty v pozorovacím okně zrovna budou, neboť jiné rozdělení bude v okolí jezera a jiné ve více zalesněné části parku. Takový bodový proces bude kvůli



Obrázek 1.3: Místa usazení lidí v Gordon Square.

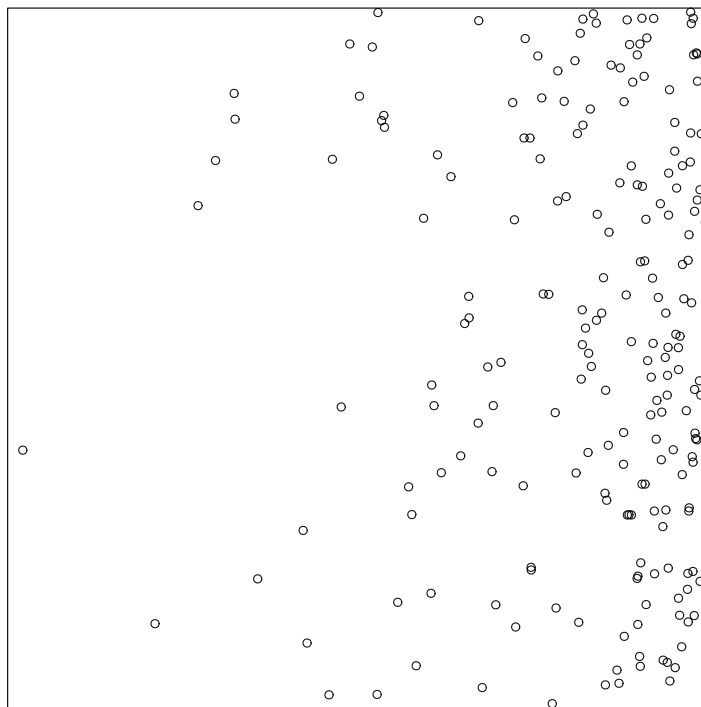
takovýmto objektům zřejmě nestacionární.

Jiná situace bodového procesu je na obrázku 1.4. Tato realizace vznikla cílenou volbou rozdělení bodového procesu za pomoci programu R. Zde máme jasnou představu očekávání, jak bude vypadat bodový proces mimo uvedené pozorovací okno. Čím více bychom pozorovací okno posunuli doprava, tím větší „hustotu“ bodů bychom viděli. Zmíněnou „hustotou“ ve skutečnosti myslíme pojem funkce intenzity, která se u bodových procesů hojně využívá a je základním nástrojem, jak pracovat s bodovými procesy. Intuitivně víme, co si představit za tímto pojmem, zavedeme si ji obecně pro bodové procesy v definici 5 (ale také si uvedeme speciální tvar pro stacionární bodové procesy v definici 6, neboť v této práci budeme v euklidovských prostorech pracovat zejména se stacionárními bodovými procesy). Tedy čím více posuneme pozorovací okno doprava, tím větší hodnoty funkce intenzity lze pozorovat a budeme dostávat zcela jiný pohled na bodový proces. Ovšem bez znalosti vlastností daného nestacionárního bodového procesu nemůžeme obecně říci nic o části bodového procesu mimo pozorovací okno, můžeme pouze pracovat s tím, co vidíme v pozorovacím okně.

Shrňme si předchozí odstavce. Pokud bude bodový proces nestacionární, jednotlivé vlastnosti bodového procesu se budou měnit napříč euklidovským prostorem, zatímco ve stacionárním případě můžeme pracovat s libovolnou částí prostoru díky invariantnosti vzhledem k translaci, aniž by docházelo k nějaké změně.

Poznámka: V dalších částech práce, ve kterých se budeme věnovat bodovým procesům v euklidovském prostoru, pokud neřekneme jinak, budeme předpokládat stacionaritu.

Zdefinovali jsme si stacionární bodové procesy. Můžeme tak nejen zdefinovat funkci intenzity, ale rovnou uvést i speciální případ pro stacionární bodové procesy s pojmem **intenzita**, kterou budeme později v práci potřebovat.



Obrázek 1.4: Nestacionární bodový proces - funkce intenzity roste směrem doprava.

Značení: Necht X je bodový proces na \mathbb{R}^d . $X(A)$ značí počet bodů procesu obsažených v A , kde A je podmnožinou \mathbb{R}^d .

Definice 5. Necht X je bodový proces na \mathbb{R}^d . Míra intenzity Λ je množinová funkce splňující rovnost $\mathbb{E}X(A) = \Lambda(A)$ pro všechny borelovské množiny A . Pokud míra intenzity má Radon-Nikodymovu derivaci vzhledem k Lebesgueově míře na \mathbb{R}^d , definujme funkci intenzity jako nezápornou měřitelnou funkci $\lambda(x)$ splňující rovnost $\mathbb{E}X(A) = \Lambda(A) = \int_A \lambda(x) dx$ pro všechny borelovské množiny A .

Při předpokladu stacionarity se nám definice funkce intenzity zjednoduší, neboť intenzita bude konstantní.

Definice 6. Necht X je stacionární bodový proces na \mathbb{R}^d . Intenzita je nezáporné reálné číslo λ splňující $\mathbb{E}X(A) = \lambda|A|$ pro všechny borelovské množiny A , kde $|A|$ je Lebesgueova míra množiny A .

Poznámka: Intenzita je tedy střední hodnota počtu bodů na množině o velikosti rovné 1 a pro stacionární bodový proces je konstantní.

Pro další část práce potřebujeme zavést Palmovo rozdělení. Než se dostaneme k definici převzaté z (2) ze strany 178, uvedeme motivaci, proč podmíněné rozdělení zavádět. Při práci s veličinami charakterizujícími stacionární bodový proces budeme pracovat se zkoumáním okolí nějakého „typického“ bodu. Nejšikovnější volba takového bodu je počátek, který nám zjednoduší zápisy a práci s veličinami popisujícími bodový proces. Nicméně pravděpodobnost jevu „Stacionární bodový

proces X má bod v počátku“ je nulová a vzniká tak problémová situace, jak vyjádřit příslušnou veličinu z důvodu podmiňování jevem s nulovou pravděpodobností. Z tohoto důvodu využijeme stacionaritu bodového procesu X a definujeme **Palmovo rozdělení** jako pravděpodobnostní míru P_o splňující rovnost

$$\lambda|A|P_o(X \in V) = \mathbb{E}\left(\sum_{x \in X \cap A} \mathbb{1}_V(X - x)\right), \quad (1.1)$$

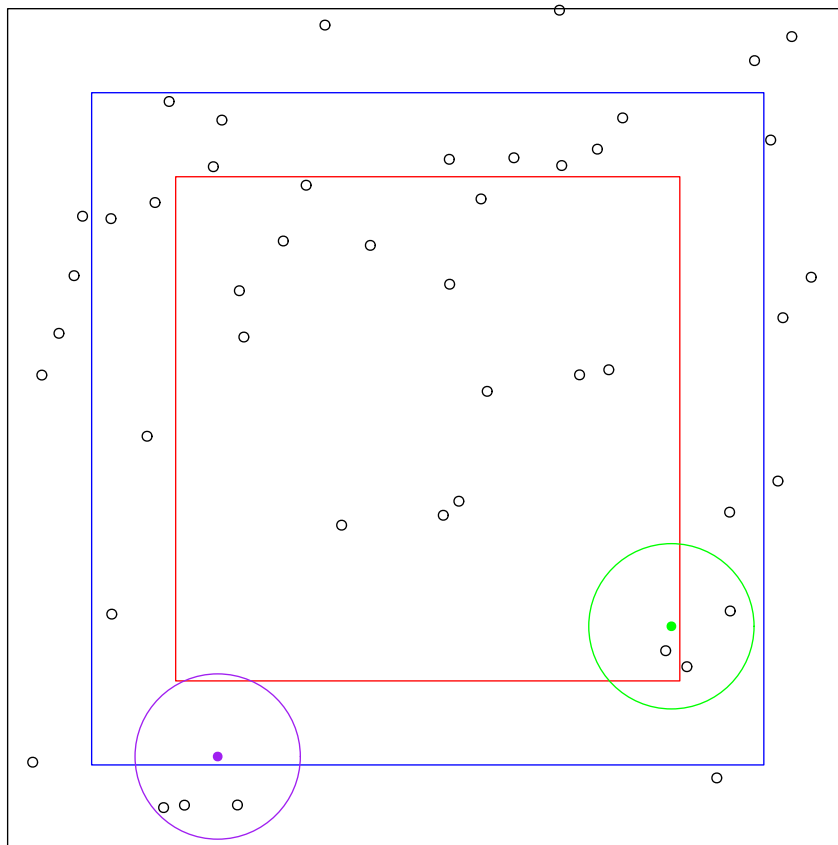
kde A je měřitelná a konečná podmnožina euklidovského prostoru, V je množina lokálně konečných měr, která může splňovat nějakou vlastnost, a $x \in X \cap A$ značí body bodového procesu X , které leží v podmnožině A a $X - x$ značí bodový proces, kde jsou všechny body posunuty o $-x$ (tedy $X - x$ je bodový proces obsahující bod v počátku).

Rovnost (1.1) nám říká, že střední hodnota výrazu, který značí kolik bodových procesů $X - x$, kde $x \in X \cap A$, (vzniklých z posunutí celého bodového procesu X tak, aby konkrétní bod x původního bodového procesu X ležící v podmnožině A byl nově v počátku) má vlastnost $X - x \in V$, vydělena intenzitou a velikostí podmnožiny A , je rovna podmíněné pravděpodobnosti jevu $X \in V$ při znalosti „počátek je bodem v X “.

Obdobně lze zavést střední hodnotu funkce f vzhledem k Palmovu rozdělení, kde funkce f přiřadí posunutému bodovému procesu $X - x$ pro $x \in X \cap A$ reálné číslo, například počet bodů rozdílných od počátku, které jsou od bodu v počátku vzdáleny nejvýše r (budeme později využívat u K -funkce, kde při označení posunutého bodového procesu jako Z můžeme tuto funkci zapsat jako $(Z(b(0,r) \setminus \{0\}))$ s využitím výše zavedeného značení $Z(A)$ značící počet bodů v Z ležících v $A \subset \mathbb{R}^d$). Tedy **střední hodnotu vzhledem k Palmovu rozdělení** s využitím zmíněného značení výše definujeme pomocí rovnice

$$\lambda|A|E_o(f(X)) = \mathbb{E}\left(\sum_{x \in X \cap A} f(X - x)\right). \quad (1.2)$$

V praxi ovšem nepracujeme s celým bodovým procesem, ale pracujeme pouze s množinou $W \subset \mathbb{R}^d$, ve které pozorujeme realizaci procesu a kterou nazýváme „pozorovací okno“. Pro $d = 2$ má pozorovací okno typicky obdélníkový či dokonce čtvercový tvar a na tomto tvaru budeme v následující části ilustrovat **okrajový efekt**. Tento efekt tvoří problém při odhadování funkcí bodového procesu. Představme si situaci, kdy chceme odhadnout funkci počítající počet bodů v okolí nějakého bodu procesu s argumentem $r > 0$, kde r určuje nejdelší vzdálenost, kterou body mohou mít, abychom je ještě započítali. Pokud ale výchozí bod, jehož okolí o poloměru r budeme zkoumat, je k okraji pozorovacího okna W blíže než je daná velikost r , může nám nastat situace, kdy do tohoto okolíčka nezapočítáme všechny body, neboť mohou ležet mimo pozorovací okno, a dojde k špatnému odhadnutí této funkce. Na ilustraci tohoto případu se můžeme podívat na obrázek 1.5, kde si jako pozorovací okno představme část ohraničenou modrým obrysem čtverce. Pokud si vezmeme fialový bod a budeme se chtít podívat na počet jeho sousedů v jeho okolí o nenulovém pevně daném poloměru r vyznačeném fialovou kružnicí, v pozorovacím okně neuvidíme žádný bod procesu, zatímco mimo pozorovací okno bychom našli 3 sousedy. Odhadování některých vlastností by se tedy razantně změnilo při znalosti, že v okolí daných bodů není prázdné okolí, natož kdyby byl známý počet sousedů.



Obrázek 1.5: Okrajový efekt a minusová korekce.

Abychom mohli funkce charakterizující bodový proces odhadnout lépe, existují různé typy korekcí okrajového efektu. V této práci budeme zásadně využívat **mínusovou korekci**, kdy pro dané r zúžíme oblast W na její podmnožinu $W_{\ominus r}$, která má všechny body od hranice W ve vzdálenosti alespoň r a zároveň leží stále v množině W . Podobné zúžení by se provádělo i pro další hodnoty r . Budeme ovšem pracovat pouze s takovými r , pro která $W_{\ominus r}$ vychází jako neprázdná množina. Při odhadování funkcí budeme pak zkoumat okolíčka bodů pouze ležících v $W_{\ominus r}$ a tak zabráníme tomu, aby nastal problém s okrajovým efektem. Při opětovném nahlédnutí do obrázku 1.5 si můžeme všimnout vyznačení zmenšeného pozorovacího okna $W_{\ominus r}$ pomocí červeného obrysu menšího čtverce, zatímco pozorovací okno je stále znázorněno modrým obrysem čtverce. I když zelený bod je při okraji zmenšeného pozorovacího okna $W_{\ominus r}$, jeho okolí o poloměru r je celé v pozorovacím okně W , a tak napočítáme správně místo 1 všechny 3 sousedy. Pomocí této korekce lze vytvářet některé nestranné odhady.

1.3 K -funkce

1.3.1 Definice

V praxi se nám u bodových procesů často může hodit charakteristika, která bude odhadovat průměrný počet bodů v okolí nějakého typického bodu procesu.

Taková myšlenka je jen krůček od zavedení K -funkce, ve které je tato charakteristika její součástí. K -funkci je pro účely tohoto textu důležité zavést, neboť ji budeme používat jako testovou statistiku u hypotézy, kterou formulujeme ve čtvrté kapitole, konkrétně ji využijeme pro vyhodnocení jednotlivých simulací potřebných k testování dané hypotézy.

Protože kótami převáženou K -funkci na sféře potřebujeme zavést, budeme v každé kapitole uvádět její modifikaci pro danou situaci. Kótami převáženou K -funkci pak dostaneme jako přímou analogii předchozích případů. Na tomto místě uvedeme její nejzákladnější podobu, tedy bez kót, které zavedeme v druhé kapitole, a v euklidovském prostoru.

Zavedeme nyní základní podobu a to K -funkci pro stacionární bodové procesy v euklidovském prostoru. Protože pracujeme se stacionárním bodovým procesem, tak za typický bod můžeme vzít libovolný bod bodového procesu a přesunout celý bodový proces tak, aby tento typický bod byl v počátku, čemuž jsme se věnovali v předchozí podkapitole při zavádění střední hodnoty vzhledem k Palmovu rozdělení pomocí rovnice (1.2). Bude nás zajímat podmíněná střední hodnota počtu sousedních bodů v jeho okolí o poloměru r . To můžeme zapsat jako $\mathbb{E}_0(X(b(0,r) \setminus \{0\}))$. Pokud toto vydělíme intenzitou, tak dostáváme K -funkci, kterou chápeme jako funkci o argumentu r určující velikost okolí daného typického bodu. Tedy K -funkce pro stacionární bodové procesy je definována následující rovností:

$$K(r) = \frac{\mathbb{E}_0(X(b(0,r) \setminus \{0\}))}{\lambda}, \quad r \geq 0.$$

Ve čtvrté kapitole nebudeme ovšem pracovat přímo s její teoretickou podobou, ale budeme využívat její odhad.

1.3.2 Odhad

K -funkci budeme odhadovat pomocí počtu pozorovaných dvojic bodů, které jsou k sobě blíže, než je kladná vzdálenost r . Protože nepracujeme s celým bodovým procesem, ale máme typicky jen nějakou část bodového procesu danou pozorovacím oknem, tak budeme využívat minusovou korekci. Pro každý bod ležící ve zmenšeném pozorovacím okně napočítáme počet bodů v jeho okolí o poloměru r bez započítání toho bodu samotného. Protože intenzitu typicky neznáme a pro naše účely ani nepotřebujeme přesně znát, bude pro nás dostatečné odhadnout $\lambda^2 K$ a to pomocí odhadu

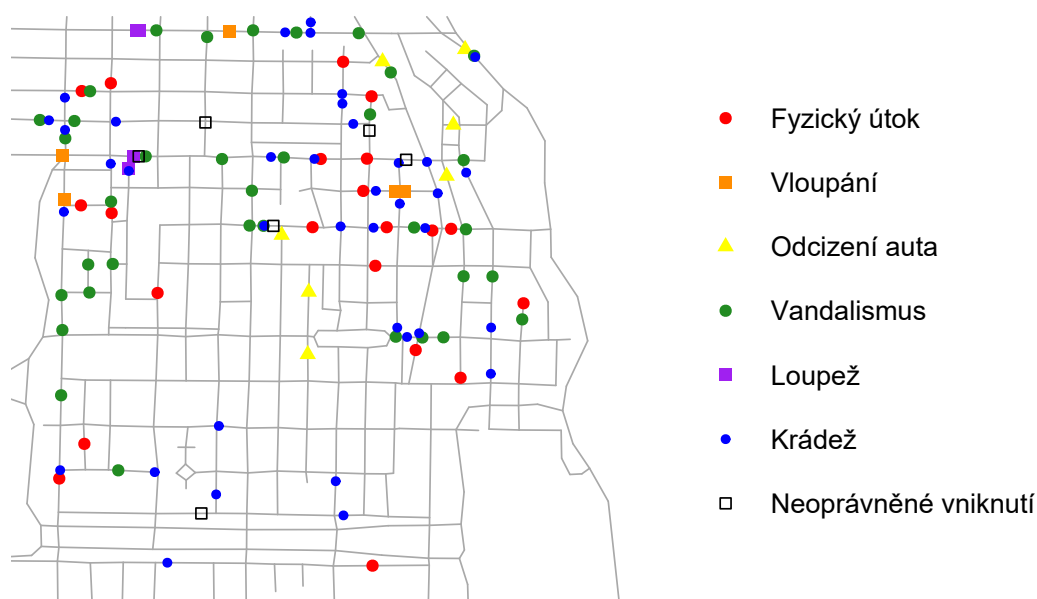
$$\widehat{\lambda^2 K(r)} = \frac{\sum_{x \neq y} \mathbb{1}_{[x \in W_{\Theta r}, \|x-y\| \leq r]}}{|W_{\Theta r}|}, \quad r \geq 0.$$

Dá se ukázat, že jde o nestranný odhad.

2. Kótované bodové procesy v euklidovském prostoru

2.1 Motivace

Následující příklad není příkladem bodového procesu v \mathbb{R}^2 , ale jedná se o bodový proces na lineární síti, kde se situace koncepčně liší od naší teorie, avšak dobře nám to poslouží pro ilustraci, co jsou to kóty a jak je lze používat i v reálných situacích.



Obrázek 2.1: Pouliční zločiny ve městě Chicago.

Na obrázku 2.1 vidíme pouliční zločiny v jedné z částí města Chicago, která je v okolí místní univerzity. Zločiny byly hlášeny v období od 25. dubna do 8. května v roce 2002. Původní mapa s těmito zločiny byla publikována v chicagském týdeníku. Data jsou v balíčku `spatstat`.

Mapa nezobrazuje pouze polohu uskutečněného zločinu, ale pomocí tvaru a barvy symbolu můžeme zjistit i druh zločinu. Můžeme se tak dozvědět, zda se na daném místě udála loupež, odcizení auta, fyzický útok, nebo například vloupání do domu, který stojí u vyznačeného místa na mapě, tedy nejedná se čistě jen o zločiny, které se po celou dobu odehrávají přímo na ulici, ale i o ty v její nejbližší blízkosti. Lze tak nejen zkoumat, kde se odehrávají nějaké zločiny, ale také, co se stává v jednotlivých oblastech za zločin, zda se zločin stejného charakteru odehrává často na podobných místech, jako vidíme shlukování u zelených puntíků zastupující vandalismus, dokonce můžeme zkoumat, zda daný druh zločinu může záviset na poloze, například odcizení auta v daném časovém období bylo hlášeno buď uprostřed pozorované části města, nebo v jeho severovýchodní části. V některých jiných příkladech bychom si zase mohli pokládat jiné otázky typu, zda objekty s danou vlastností v nějakém smyslu ovlivňují jiné objekty s jis-

tou vlastností. V případě se zločiny bychom si tak mohli položit otázku, jestli je výskyt vloupání ovlivňován zločiny, při kterých dochází ke vandalismu.

Jak jsme viděli u úvodního příkladu, bodový proces v takovém případě nese o sobě mnohem více informací, než jsou polohy bodů, neboť každý bod má pomocí kóty v sobě uchovanou další netriviální informaci a můžeme se toho dozvědět mnohem více. Proto má smysl zavádět bodové procesy s kótami, kterým budeme říkat **kótované bodové procesy**.

2.2 Druhy kót

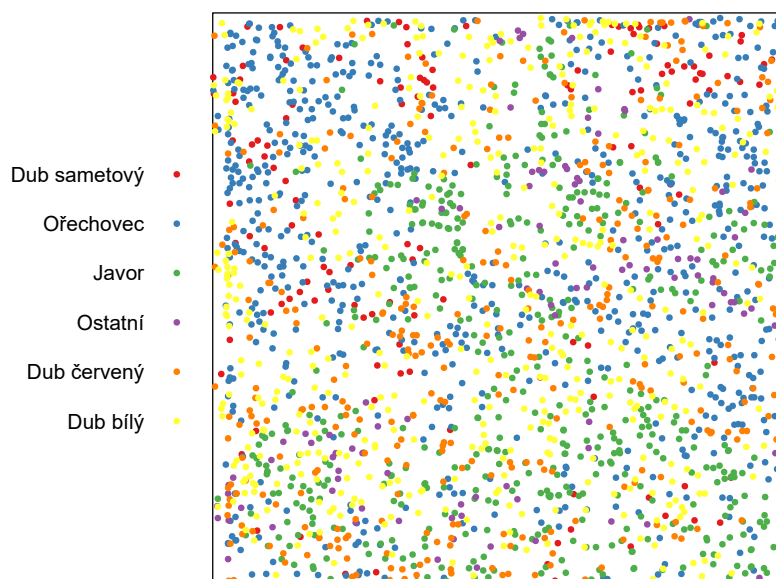
Podobně jako v minulé kapitole budeme chtít položit teoretické základy tak, abychom ke konci kapitoly mohli modifikovat K -funkci pro situaci, kdy chceme pracovat i s kótami. Nejdříve se podíváme na zavedení kótovaných bodových procesů a na typy kót.

Definici přejímáme z (3) [strana 20].

Definice 7. *Kótovaný bodový proces na prostoru S s kótami v prostoru Θ je bodový proces M na prostoru $S \times \Theta$ takový, že počet bodů M v množině $K \times \Theta$ je konečný skoro jistě pro všechna kompaktní $K \subset S$. To znamená, že odpovídající podkladový proces (bodový proces, který vznikne z M , když odstraníme kóty) je lokálně konečný.*

Zavedení kótovaného bodového procesu si lze představit následovně: mějme bodový proces X a seřadme jeho body do posloupnosti $\{x_n\}_{n \in \mathbb{N}}$. Přiřadíme-li každému bodu x_n reálné číslo $m(x_n)$, které budeme nazývat **kótu** bodu x_n , dostaneme **kótovaný bodový proces** $M = \{[x_n; m(x_n)]\}$.

I když jsme o kótě mluvili jako o reálném čísle, typicky kóta nabývá nezáporných hodnot (obvykle se jedná o velikost objektu: velikost kráteru, tloušťka stromu v lese, rozloha města, nadmořská výška vrcholku sopky a podobně) nebo ji dokonce zúžíme na přirozená čísla (počet obyvatel města) či dokonce na konečný výčet kategorií, kde každé číslo z množiny hodnot zastupuje konkrétní kategorii (druhy stromů v lese, typy zločinů ve městě). Přidělování kóty bodu může být závislé na poloze bodu, pro ilustraci si můžeme představit situaci, kdy kóta bude měřitelnou funkcí normy $\|x\|$, či si představme reálnější situaci, kdy mnoho rostlin na stejném místě znamená větší boj o živiny, a tak jejich vzrůst bude typicky menší, zatímco více osamocené rostliny se nemusí o živiny s ostatními dělit a jejich velikost může být výrazně větší. V obou případech dochází ke korelaci kót a bodů. Ve zcela jiném případě může být kóta přiřazena nezávisle na poloze bodu, kdy kóty jsou navzájem nezávislé a pochází ze stejného rozdělení (budeme používat zkratku i.i.d., pokud kóty budou nezávislé a stejně rozdělené), jinými slovy jedná se o kótovaný bodový proces s i.i.d. kótami. Velice jednoduchým příkladem bodového procesu s i.i.d. kótami je situace, kdy u každého bodu si hodíme spravedlivou kostkou a výsledek zapíšeme pomocí kóty. V takovém případě velikost kóty opravdu nezávisí na poloze bodu, na velikosti jiných kót a každá kóta pochází ze stejného rozdělení. Situace s i.i.d. kótami je nejjednodušším typem kótovaných bodových procesů.



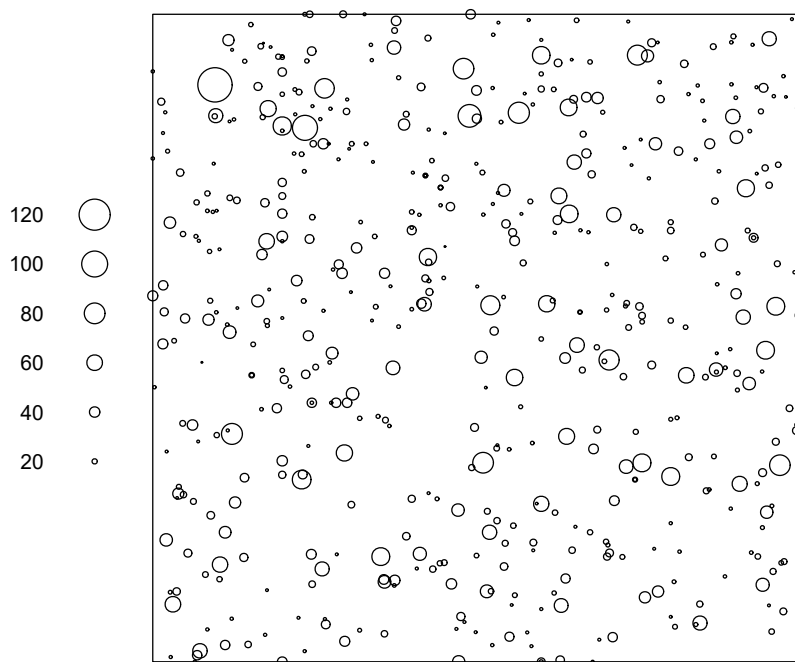
Obrázek 2.2: Rozmístění stromů v lese Lansing.

Bodové procesy mohou mít vícerozměrné kóty, kdy každá složka kóty bude popisovat určitou vlastnost daného bodu, respektive objektu v reálném světě. Můžeme mít například nejen polohu měst v České republice danou polohou bodu, ale kóta může obsahovat informace o počtu obyvatel, rozloze města, roku založení a podobně. V naší práci budeme pracovat pouze s jednorozměrnými kótami. Budeme tedy mít dva základní typy kót. Buď bude kóta nabývat konečného výčtu hodnot, tedy kóta bude pocházet z diskrétního rozdělení, nebo bude pocházet ze spojitého rozdělení.

Situaci, kdy kóta nabývá konečného výčtu hodnot, můžeme vidět na obrázku 2.2. Jedná se o data z balíčku `spatstat` obsahující polohu stromů v Lansing Woods ve státu Michigan (USA) ve výřezu 924 stop krát 924 stop. U každého stromu máme navíc informaci, o jaký z 5 nejčastějších druhů stromů se jedná, případně ho máme zařazeného do kategorie „Ostatní“. Můžeme tak zkoumat složení lesa z pohledu druhů stromů.

Druhou situací jsou kóty pocházející ze spojitého rozdělení. Opět si můžeme uvést příklad z lesa, avšak místo druhového složení lesa nás tentokrát budou zajímat velikosti jednotlivých stromů. Opět využijeme data z balíčku `spatstat`, tentokrát máme výřez 100 metrů krát 100 metrů z národního parku Waka v africké zemi Gabon. U každého stromu je změřena výčetní tloušťka ve výčetní výšce, to znamená ve výšce 1,3 metru je změřena tloušťka stromu a máme ji uvedenou v centimetrech. Kóty jsou zvětšené oproti poloze stromů tak výrazně, že na obrázku dochází k překryvu symbolů, i když stromy ve skutečnosti nemusí být v přímém dotyku. Na obrázku 2.3 vidíme, že v celém pozorovacím okně lze najít stromy různých velikostí, nikde není absence malých stromů a i velké stromy najdeme skoro všude.

Ukázali jsme si hlavní typy kót. Nyní můžeme zavést důležitou vlastnost, a tou bude stacionarita. Definice stacionárních kótovaných bodových procesů je



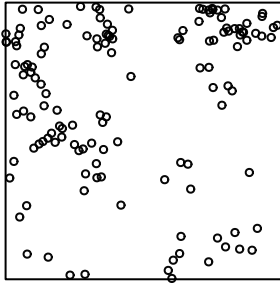
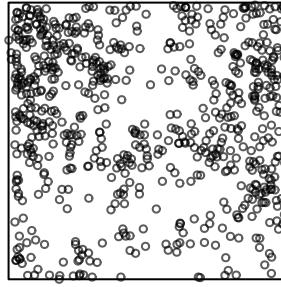
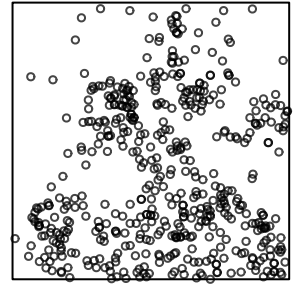
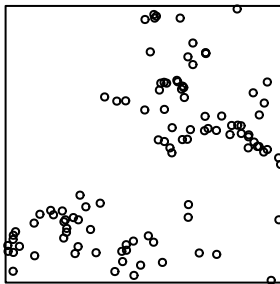
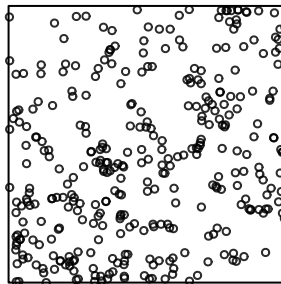
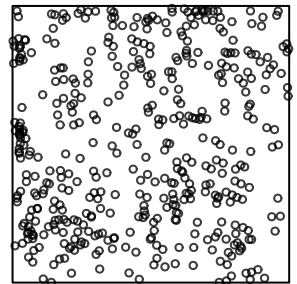
Obrázek 2.3: Rozmístění a velikosti stromů v lese Waka.

zcela analogická, odpovídající definici 4 v první kapitole. Opět požadujeme, aby rozdělení bylo invariantní vzhledem k translaci.

Definice 8. *Kótovaný bodový proces M nazveme **stacionárním**, pokud jeho rozdělení je invariantní vzhledem k translaci v \mathbb{R}^d , t.j. rozdělení posunutého kótovaného bodového procesu $M_x = \{[x_n + x; m(x_n)]\}$ je stejné jako rozdělení M pro všechna $x \in \mathbb{R}^d$.*

I když nám intuitivně může být jasné, co si představit pod pojmem stacionarita, v praxi může být velice obtížné odhadnout, zda lze u dané situace předpokládat stacionaritu. Hned u následujícího příkladu uvidíme, že podle obrázku bychom tipovali, že se jedná o nestacionární bodový proces, nicméně ve skutečnosti celá záležitost může být velice obtížná z důvodu mnoha různých faktorů, které z obrázku už nevyozorujeme. Avšak u druhého příkladu naopak uvidíme, že se zcela zřejmě jedná o nestacionární kótovaný bodový proces.

Podívejme se opět na obrázek dat z Lasing Woods (obrázek 2.2). Závisí poloha stromu na tom, jakého druhu je? Ekvivalentně, ovlivňuje polohu stromu jeho kóta? Situace na obrázku je poněkud nepřehledná z důvodu velkého množství bodů. Pokud si ale rozdělíme bodový proces na vícero různých procesů, kde v rámci jednoho procesu budou jen body se stejnou kótou, můžeme lépe vidět rozprostření jednotlivých stromů v lese, jak lze vidět na obrázku 2.4. Dub sametový se v pravé dolní části nachází velice řídko, javor se v pravé horní části a zejména v levé horní části moc nenachází. Méně zastoupené stromy v kategorii „Ostatní“ se nachází prakticky přibližně jen na polovině území. Naopak u ořechovce vidíme velice jednoznačnou dominanci hlavně v levém horním rohu pozorovacího okna. Jedná se o oblast 924 stop krát 924 stop, kde může rozmístění jednotlivých

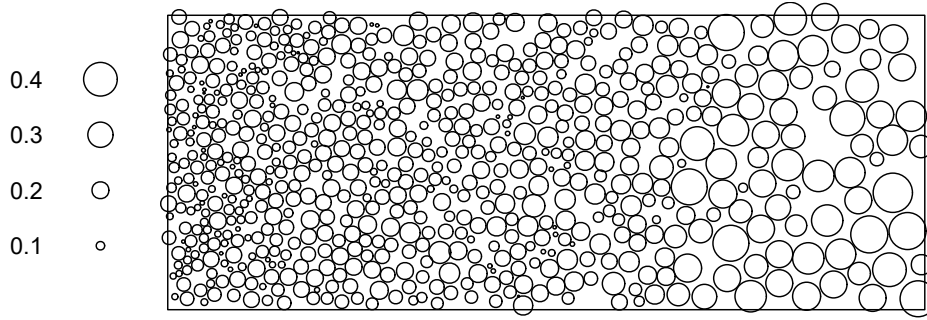
Dub sametový**Ořechovec****Javor****Ostatní****Dub červený****Dub bílý**

Obrázek 2.4: Rozmístění stromů v lese Lansing podle druhů.

druhů ovlivňovat mnoho různých faktorů, od chemického složení půdy až po mezidruhové interakce, navíc dochází k vlivům i v rámci stejného druhu (například shlukování zapříčinené způsobem rozmnožování, které jsme viděli již na obrázku 1.1). Obecně nelze vůbec jednoduše vyvodit závěry bez komplexnějších informací o okolí a jejich vlivu na stromy konzultovaných s odborníkem.

Další typ nepravidelnosti struktury kóvaného bodového procesu může být krásně viděn na obrázku 2.5, kde závislost kóty na poloze bodu je zcela oči-vidná. Směrem doprava se velikost kóty zvětšuje. Jedná se přitom o data, kdy se zaznamenala poloha a velikost částic po procesu postupného usazování částic bronzového prášku různých velikostí a sinterování. Pozorovací okno má na šířku 17 milimetrů, na výšku 7 milimetrů. Více informací o daném procesu lze najít v balíku `spatstat`, jedná se o data `bronzefilter`.

Stacionarita je i v této kapitole velice důležitá a budeme ji potřebovat při budování základní teorie, abychom mohli zformulovat modifikaci K -funkce pro kóvané bodové procesy, kterou budeme nazývat **kótami převážená K -funkce**. Ovšem než se k ní dostaneme, musíme položit nějaké teoretické základy.



Obrázek 2.5: Podélný rovinný řez částicemi při filtrování.

2.3 Budování základní teorie

Pro definici kótami převážené K -funkce budeme potřebovat obdobně jako v první kapitole Palmovo rozdělení a střední hodnotu vzhledem k Palmovu rozdělení. Než se k nim ale dostaneme, potřebujeme položit pár nejzákladnějších teoretických poznatků, abychom byli schopni pracovat se střední hodnotou pro kótované bodové procesy. Tyto poznatky budeme přejímat z páté kapitoly (2).

Pokud stacionární bodový proces M s intenzitou λ bude mít kóty z diskrétního rozdělení, bez újmy na obecnosti lze každý typ kóty očíslovat přirozeným číslem, tudíž hodnotu kóty můžeme označit jako T , a pravděpodobnost, že kóta nabude hodnoty $i \in \mathbb{N}$, označíme

$$p_i = P(T = i).$$

Protože bodový proces s kótami z diskrétního rozdělení můžeme rozdělit na jednotlivé bodové procesy obsahující pouze kóty stejného typu, dostáváme nekótované bodové procesy se zmenšenou intenzitou λp_i . Z první kapitoly víme, že očekávaný počet bodů s kótou i pro množinu $B \subset \mathbb{R}^d$ u bodového procesu M je

$$\mathbb{E}(M(B \times \{i\})) = \lambda p_i |B|, \quad i \in \mathbb{N},$$

kde $|B|$ je velikost množiny B . Střední hodnota bodů se stejnou kótou je přímou analogií situace z první kapitoly, nicméně k bodovým procesům s kótami ze spojitého rozdělení musíme přistupovat zcela jinak.

Mějme kótovaný bodový proces M s intenzitou λ , kde kóty pochází ze spojitého rozdělení T . Zavedme si distribuční funkci kóty F_T jako pravděpodobnost, že napozorujeme hodnotu kóty menší nebo rovnou m , to znamená

$$F_T(m) = P(T \leq m), \quad m \in \mathbb{R}.$$

S využitím zavedené distribuční funkce kóty lze vypočítat střední hodnotu počtu bodů s kótou menší nebo rovnou m v množině $B \subset \mathbb{R}^d$

$$\mathbb{E}[M(B \times (-\infty; m])] = \lambda F_T(m) |B|, \quad m \in \mathbb{R},$$

respektive pokud nás bude zajímat střední hodnota bodů v množině B s kótami většími než $m_1 \in \mathbb{R}$, ale ne většími než $m_2 \in \mathbb{R}$ pro $m_1 \leq m_2$, lze to s využitím předešlého vztahu vyjádřit následovně

$$\mathbb{E}[M(B \times (m_1; m_2))] = \lambda|B|(F_T(m_2) - F_T(m_1)).$$

U kót ze spojitého rozdělení nás samozřejmě často zajímá střední hodnota kóty μ . K jejímu vypočtení zaveďme hustotu kóty, která je derivací distribuční funkce, pokud její derivace existuje, tedy $f_T(m) = F_T'(m)$.

Palmovo rozdělení a střední hodnotu vzhledem k Palmovu rozdělení pro kótované bodové procesy lze zavést podobně jako jsme to udělali u (1.1) a (1.2) pro nekótované bodové procesy. S využitím strany 305 z (2) uvedeme rovnou výsledný tvar. Předpokládáme, že pro kótu typického bodu (opět posuneme bodový proces o $-x$ tak, aby se tento bod nacházel v počátku) platí, že pochází z podmnožiny hodnot L . U kót z diskrétního rozdělení můžeme vzít $L = \{i\}$ pro libovolnou hodnotu z množiny hodnot kót, zatímco pro kóty ze spojitého rozdělení můžeme vzít například $(-\infty; m]$, kde m je libovolná hodnota z nosiče rozdělení, nicméně lze L položit rovnu celé množině hodnot, kterých nabývají kóty. Za této situace definujeme **Palmovo rozdělení** jako pravděpodobnost P_L (pokud taková existuje) splňující pro všechny borelovské podmnožiny $B \subset \mathbb{R}^d$

$$\lambda|B|P_L(M \in V) = \mathbb{E} \left(\sum_{[x; m(x)] \in M} \mathbb{1}_B(x) \mathbb{1}_L(m(x)) \mathbb{1}_V(M - x) \right),$$

kde $|B|$ znamená velikost množiny B a V je množina lokálně konečných měr, která může splňovat nějakou vlastnost.

Obdobně **střední hodnotu vzhledem k Palmovu rozdělení** pro funkci kótovaného bodového procesu $f(M)$ (kde funkce f přiřadí posunutému bodovému procesu $M - x$ pro $x \in M \cap B$ reálné číslo, například počet bodů rozdílných od počátku, které jsou od bodu v počátku vzdáleny nejvýše r) definujeme pomocí rovnice

$$\lambda|B|\mathbb{E}_L(f(M)) = \mathbb{E} \left(\sum_{[x; m(x)] \in M} \mathbb{1}_B(x) \mathbb{1}_L(m(x)) f(M - x) \right).$$

Tímto máme položen potřebný teoretický základ pro zavedení kótami převážené K -funkce.

2.4 Kótami převážená K -funkce

Z první kapitoly víme, že K -funkce pro nekótované bodové procesy a pro $r \geq 0$ je definována jako

$$K(r) = \frac{\mathbb{E}_0(X(b(0,r) \setminus \{0\}))}{\lambda}.$$

Nyní bychom chtěli modifikovat K -funkci tak, aby pracovala i s velikostí kót. Začneme triviální úpravou čitatele, který chceme přepsat pomocí sumy, ale aby se stále pracovalo s počtem sousedních bodů počátku za předpokladu, že v počátku máme bod procesu, tedy

$$K(r) = \frac{\mathbb{E}_0 \sum_{x \in X}^{x \neq 0} \mathbb{1}_{[0 \leq \|x\| \leq r]}(x)}{\lambda}, \quad r \geq 0,$$

což je stále ekvivalentní původní formulaci. Nyní modifikujme tvar. Už nepracujeme se stacionárním bodovým procesem X , ale se stacionárním kótovaným bodovým procesem M , který má kóty stejně rozdělené (ale nepředpokládáme nezávislost). Označme si $\mu = \mathbb{E}_0 m(0)$. K -funkce pracovala s počty dvojic, ve kterých vystupuje typický bod a které jsou od sebe vzdálené méně než $r \in \mathbb{R}$. Ty chceme převážít jejich kótami, tedy každý indikátor pro dvojici přenásobíme velikostí jejich kót. Abychom funkci normalizovali, vydělíme ji kvadrátem střední hodnoty kót. Po takovéto úpravě získáváme **kótami převáženou K -funkci** pro $r \geq 0$ definovanou

$$K_{mm}(r) = \frac{\mathbb{E}_0 \sum_{[x;m(x)]}^{x \neq 0} m(0)m(x) \mathbb{1}_{[0 \leq \|x\| \leq r]}}{\lambda \mu^2}.$$

Kdybychom přidali předpoklad nezávislosti kót, kótami převážená K -funkce se zredukuje „zpět“ na K -funkci. Pro bodový proces M s i.i.d. kótami totiž platí

$$\begin{aligned} K_{mm}(r) &= \frac{\mathbb{E}_0 \sum_{[x;m(x)]}^{x \neq 0} m(0)m(x) \mathbb{1}_{[0 \leq \|x\| \leq r]}}{\lambda \mu^2} = \frac{\mathbb{E}_0 m(0) \mathbb{E}_0 \sum_{[x;m(x)]}^{x \neq 0} m(x) \mathbb{1}_{[0 \leq \|x\| \leq r]}}{\lambda \mu^2} \\ &= \frac{\mu \mathbb{E}_0 \sum_{[x;m(x)]}^{x \neq 0} m(x) \mathbb{1}_{[0 \leq \|x\| \leq r]}}{\lambda \mu^2} = \frac{\mathbb{E}_0 \sum_{[x;m(x)]}^{x \neq 0} m(x) \mathbb{1}_{[0 \leq \|x\| \leq r]}}{\lambda \mu} \\ &= \frac{\mathbb{E}_0 \sum_{[x;m(x)]}^{x \neq 0} m(0) \mathbb{1}_{[0 \leq \|x\| \leq r]}}{\lambda \mu} = \frac{\mathbb{E}_0 m(0) \mathbb{E}_0 \sum_{[x;m(x)]}^{x \neq 0} \mathbb{1}_{[0 \leq \|x\| \leq r]}}{\lambda \mu} \\ &= \frac{\mu \mathbb{E}_0 \sum_{[x;m(x)]}^{x \neq 0} \mathbb{1}_{[0 \leq \|x\| \leq r]}}{\lambda \mu} = \frac{\mathbb{E}_0 \sum_{[x;m(x)]}^{x \neq 0} \mathbb{1}_{[0 \leq \|x\| \leq r]}}{\lambda} = K(r), \quad r \geq 0, \end{aligned}$$

kde jsme postupně využili toho, že kóty jsou i.i.d., nezávislosti kót s identifikátorem, zda daná dvojice kót je k sobě blíže než pevné r , a u poslední rovnosti jsme využili toho, že u kótovaného bodového procesu M využíváme pouze informaci o poloze bodů, tedy pracujeme pouze s nekótovaným bodovým procesem X .

Kótami převážená K -funkce je modifikace K -funkce, která je významně bližší k podobě testové statistiky, kterou budeme využívat na konci čtvrté kapitoly. Opět si uvedeme odhad a podobně jako v první kapitole, nebudeme potřebovat zvlášť odhadnout intenzitu λ a střední hodnotu kóty μ , postačí nám společný odhad

$$\lambda^2 \widehat{\mu^2 K_{mm}}(r) = \frac{\sum_{x \in X \cap W_{\Theta r}} \sum_{y \in X \cap W}^{x \neq y} m(x)m(y) \mathbb{1}_{[0 \leq \|x-y\| \leq r]}}{|W_{\Theta r}|}, \quad r \geq 0.$$

3. Bodové procesy na sféře

3.1 Motivace

Jak bylo pomocí příkladu 1.2 zmíněno, bodový proces v reálném světě nemusíme hledat pouze v rovině nebo v prostoru, ale můžeme zkoumat i bodové procesy na sféře, jako byla poloha kráterů na Měsíci. Uvedme si několik dalších příkladů pro ilustraci, co lze na sféře zkoumat.

Příklad 3.1: Zkoumání okolního vesmíru je pro mnohé lidi velice zajímavé. Avšak při obvyklém pozorování noční oblohy nelze jednoduše říci, jak daleko je které těleso vzdáleno od pozorovatele bez netriviálních znalostí o konkrétním tělesu nebo nějakých větších výpočtů. Proto se někdy pracuje s takzvanou nebeskou sférou. Jde o myšlenou sféru o dostatečně velikém poloměru, kde ve středu stojí pozorovatel a tělesa, které pozoruje, jsou zobrazeny pomocí projekce na sféru tak, aby mezi nimi byla zachována úhlová vzdálenost taková, jakou pozorovatel osobně pozoruje. Když si vezmeme pozorovací okno například kolem Polárky, můžeme zkoumat, zda jsou hvězdy v této oblasti rovnoměrně rozmístěné.

Příklad 3.2: Zemi běžně aproximujeme koulí, a tím se nabízí spousta situací, které můžeme zkoumat pomocí bodových procesů na sféře. Můžeme zkoumat epicentra zemětřesení, místa dopadu meteoritů či například polohu sopek. U posledního případu nás může zajímat situace, kdy v bodovém procesu budou pouze body zastupující polohu sopek, které byly aktivní v pevně daném období, například v posledních měsících, a můžeme zkoumat, zda někde dochází k významnému shlukování aktivních sopek, a tedy zvážit, zda zintenzivnit mapování situace v této oblasti.

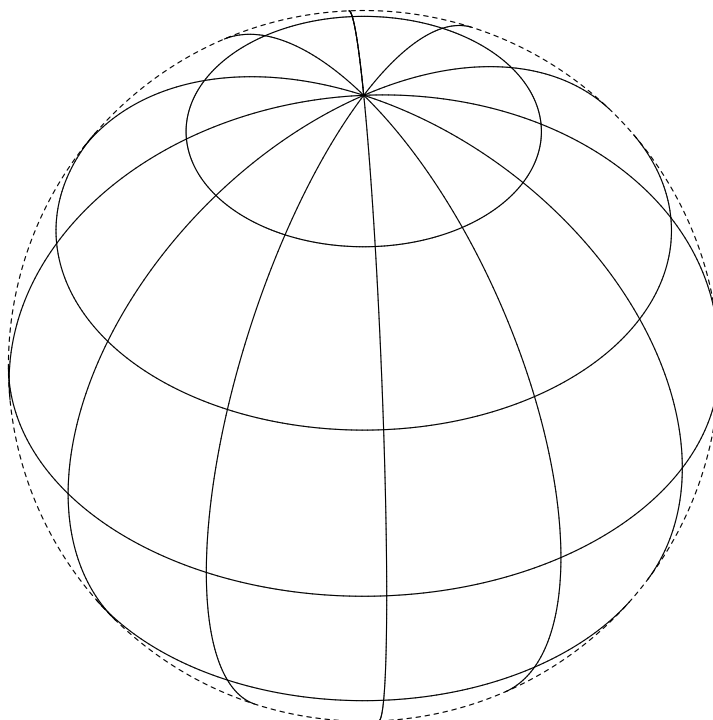
Příklad 3.3: Bodové procesy na sféře se nemusí týkat pouze vesmírných těles. Sféra může mít i relativně malé rozměry, například můžeme uvažovat detektor částic zaznamenávající odkud částice přiletěly. V takovém případě nás opět může zajímat, zda jsou polohy směrů, odkud přiletěly částice, rovnoměrně rozmístěny po sféře, případně můžeme zkoumat, zda dochází k významnému shlukování.

3.2 Definice a vlastnosti

Tato kapitola je adaptací první kapitoly na případ sféry. Budeme přebírat definice a tvrzení z (4). Začneme zavedením a popsáním sféry.

Sféra v \mathbb{R}^3 se středem v počátku $[0, 0, 0]$ a s kladným poloměrem ρ je množina $\mathbb{S}_\rho^2 = \{u \in \mathbb{R}^3 : \|u\| = \rho\}$, kde $\|u\|$ je norma vektoru u v \mathbb{R}^3 .

Označme si bod $[0, 0, \rho]$ **severním pólem** a $[0, 0, -\rho]$ **jižním pólem**. Pro udávání polohy bodu na sféře se používají obvykle sférické souřadnice s transformací

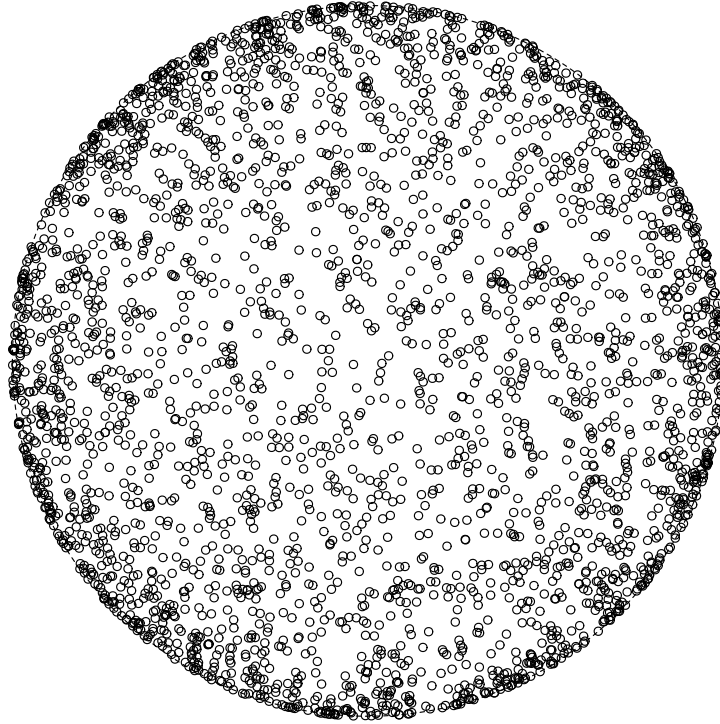


Obrázek 3.1: Zobrazení sféry a kulového vrchlíku - množiny bodů na sféře mající sférickou vzdálenost od pevně daného bodu sféry nejvýše $r \in [0, \pi\rho]$, kde ρ je poloměr sféry.

$$\begin{aligned}x &= \rho \sin \theta \cos \phi, \\y &= \rho \sin \theta \sin \phi, \\z &= \rho \cos \theta,\end{aligned}$$

kde $0 \leq \theta \leq \pi$ určuje úhel mezi daným bodem sféry a severním pólem, zatímco $0 \leq \phi < 2\pi$ odpovídá až na přeškálování tomu, čemu říkáme zeměpisná délka (místa se stejnou zeměpisnou délkou tvoří poledníky, tedy nejkratší spojnice severního a jižního pólu ležící na sféře). Při takovéto transformaci bude mít severní pól souřadnice $[\theta, \phi] = [0, 0]$, proto tento bod bývá označován **počátkem sféry**. Nazveme **sférickou vzdáleností** $d(u, v)$ bodů $u, v \in \mathbb{S}_\rho^2$ délku nejkratší spojnice bodů u a v ležící na sféře. Tato vzdálenost lze vypočítat pomocí vzorce $d(u, v) = \rho \arccos([u \cdot v]/\rho^2)$ pro $u, v \in \mathbb{S}_\rho^2$ s kartézskými souřadnicemi, kde $u \cdot v$ je skalární součin vektorů v \mathbb{R}^3 .

Na obrázku 3.1 vidíme zobrazení sféry s vyznačenými několika poledníky a rovnoběžkami (rovnoběžka je množina bodů na sféře mající stejnou hodnotu θ). Zároveň zde můžeme ilustrovat okolí bodu, které se na sféře nazývá kulový vrchlík. Vybereme-li libovolnou rovnoběžku, bude to hranice kulového vrchlíku se středem v počátku sféry, neboť všechny body jsou stejně vzdálené od počátku sféry (mají stejnou hodnotu θ) a všechny body sféry mající menší sférickou vzdálenost od počátku mají menší hodnotu θ . Jedná se vlastně obecně o průnik koule (mající střed na sféře a s poloměrem $r_{\text{euklidovské}} < 2\rho$) se sférou. Upozorníme, že uzávěr



Obrázek 3.2: Bodový proces na sféře. I když jsou body rozmístěny rovnoměrně, při kolmém průmětu do roviny vidíme „zhuštění“ na okrajích kruhu.

doplňku kulového vrchlíku je opět kulový vrchlík. Kulový vrchlík budeme označovat symbolem $C(u,s)$, kde u je střed kulového vrchlíku a $s \leq \pi\rho$ je sférická vzdálenost libovolného bodu hranice kulového vrchlíku a jeho středu.

Na obrázku 3.2 vidíme, že při kolmém průmětu do roviny vidíme „zhuštění“ bodů při okrajích obrázku. Musíme si tedy při prohlížení obrázků bodových procesů na sféře uvědomovat tuto „deformaci“ a nezískat neoprávněně pocit, že by se mohlo jednat o bodový proces s nerovnoměrně rozmístěnými body.

Další upozornění se týká pozorovacího okna. V prvních dvou kapitolách jsme pracovali s posouváním pozorovacího okna a zkoumali, zda se nám přitom mění vlastnosti pozorované části bodového procesu. Na základě toho jsme zadefinovali pojem stacionárního bodového procesu, u kterého jsme mohli pracovat s libovolnou částí prostoru, aniž by docházelo ke změnám vlastností bodového procesu. Něco podobného bychom chtěli zavést i u bodových procesů na sféře, avšak na sféře nelze posouvat bodové procesy tak, jak jsme dosud prováděli. Analogií k tomuto na sféře je **rotace**, kdy rotace R bodového procesu X je $RX = \{y : y = R(x), x \in X\}$ a nové polohy bodů lze vypočítat podle (4) na straně 31 pomocí vztahu (3.17), který je označován jako „Rodrigues’ rotation formula“. Podle tohoto autora, který se odkazuje na (5), tvoří jednotlivé rotace dohromady grupu rotací na sféře, která se označuje symbolem $SO(3)$. Toto značení budeme využívat například při budování analogické vlastnosti ke stacionaritě, kdy jsme bodové procesy posouvali v \mathbb{R}^d , $d \in \mathbb{N}$, kterým bude izotropie, kdy bude docházet k rotacím bodového procesu na sféře. Nejdříve ale musíme zadefinovat samotné bodové procesy.

Protože sféra je kompaktní, lze na ní to, co jsme si na začátku první kapitoly odvodili pro úplné metrické prostory, aplikovat. Jednotlivé definice jsme si formulovali pro úplný separabilní metrický prostor Y s borelovskými množinami $\mathbf{B}(Y)$ na Y značenými zkráceně \mathcal{B} . V naší situaci tyto definice využijeme pro sféru \mathbb{S}_ρ^2 a množinu borelovských množin na sféře, kterou budeme značit $\mathcal{B}_{\mathbb{S}_\rho^2}$. Dosazením do definice 1 dostaneme množinu všech lokálně konečných měr na sféře nabývajících pouze celočíselných hodnot $\mathcal{N}_{\mathbb{S}_\rho^2}$, obdobně z definice 2 získáme σ -algebru $\mathfrak{N}_{\mathbb{S}_\rho^2}$, což hned vzápětí využijeme pro definování bodového procesu na sféře.

Definice 9. *Bodový proces na sféře \mathbb{S}_ρ^2 je měřitelné zobrazení*

$$X : (\Omega, \Sigma, P) \rightarrow (\mathcal{N}_{\mathbb{S}_\rho^2}, \mathfrak{N}_{\mathbb{S}_\rho^2}).$$

Nyní můžeme uvést definici výše zmíněného pojmu izotropie.

Definice 10. *Bodový proces X na sféře \mathbb{S}_ρ^2 nazveme **izotropním**, pokud jeho rozdělení je invariantní vzhledem ke grupě rotací $SO(3)$. To znamená, že rozdělení bodového procesu po rotaci $RX = \{y : y = R(x), x \in X\}$ je stejné jako rozdělení X pro všechna $R \in SO(3)$.*

Definice nám říká, že pokud máme izotropní bodový proces na sféře, pak pozorovacím oknem W uvidíme stejné vlastnosti, jako bychom viděli v pozorovacím okně RW , kde bychom pozorovací okno přemístili libovolnou rotací $R \in SO(3)$. Tedy v takovém případě můžeme pracovat s libovolnou částí sféry, aniž by docházelo k nějakým změnám.

Ilustraci izotropního bodového procesu můžeme vidět na obrázku 3.3. Pokud bodový proces na sféře nebude izotropní, budeme ho nazývat neizotropním. Na obrázku 3.4 vidíme realizaci neizotropního bodového procesu, kde dochází k výraznému shlukování bodů v okolí severního pólu.

Analogii stacionarity pro bodové procesy na sféře máme nyní vybudovanou. Nicméně k definici K -funkce na sféře potřebujeme obdobně jako v první kapitole zavést funkci intenzity, Palmovo rozdělení a střední hodnotu vůči Palmovu rozdělení. Opět budeme přebírat definice z (4), kde některé definice jsou formulované obecně, avšak my je budeme uvádět pouze pro sféru. Než zdefinujeme funkci intenzity, připomeňme značení $X(B)$ pro počet bodů procesu X v množině B .

Definice 11. *Nechť máme bodový proces X na sféře \mathbb{S}_ρ^2 a $B \in \mathcal{B}_{\mathbb{S}_\rho^2}$. Definujme množinovou funkci Λ vztahem $\Lambda(B) = \mathbb{E}X(B)$. Pokud je Λ konečná pro všechny omezené borelovské množiny na \mathbb{S}_ρ^2 , pak Λ nazveme **mírou intenzity**.*

Pokud bude bodový proces X na sféře \mathbb{S}_ρ^2 izotropní, pak míra intenzity bude konstantní a bude λ -násobkem velikosti kulové plochy odpovídající množině B , pro kterou počítáme Λ . Nezápornou konstantu λ nazveme **intenzitou** bodového procesu na sféře.

Palmovo rozdělení je v (4) odvozeno jinak, než jsme k němu přistupovali v první a v druhé kapitole, vyžaduje Campbellovu míru, kterou musíme nejprve zdefinovat. Za připomínku stojí, že takto odvozené Palmovo rozdělení lze nalézt i v (1) a to v sedmé kapitole.



Obrázek 3.3: Izotropní bodový proces.

Definice 12. Necht máme bodový proces X na sféře \mathbb{S}_ρ^2 . Definujme množinovou funkci C vztahem $C(B \times N) = \mathbb{E}[X(B)\mathbb{1}_N(X)]$ pro $B \in \mathcal{B}_{\mathbb{S}_\rho^2}$ a $N \in \mathcal{B}_{\mathcal{N}_{\mathbb{S}_\rho^2}}$. Pokud je C konečná pro všechny omezené borelovské množiny $B \in \mathcal{B}_{\mathbb{S}_\rho^2}$ a všechny $N \in \mathcal{B}_{\mathcal{N}_{\mathbb{S}_\rho^2}}$, pak C nazveme **Campbellovou mírou**.

Dále zavedeme Palmovo rozdělení a redukované Palmovo rozdělení. Později budeme mít pomoci střední hodnoty vzhledem k redukovanému Palmovu rozdělení zdefinovanou K -funkci. Připomeňme, že \mathcal{N}_Y je množina všech lokálně konečných měr nabývajících pouze celočíselných hodnot pro daný prostor Y . Toto značení využijeme i pro sféru \mathbb{S}_ρ^2 v následující definici.

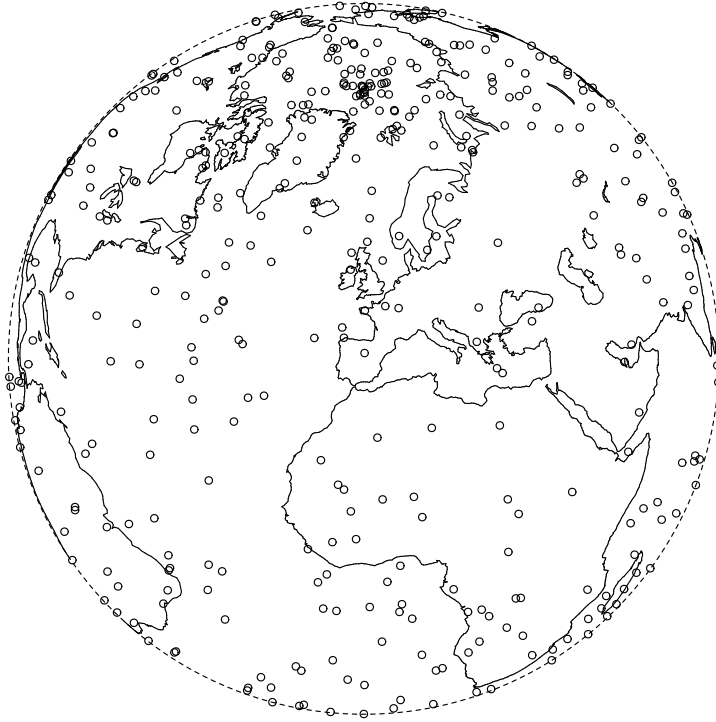
Definice 13. Necht máme bodový proces X na sféře \mathbb{S}_ρ^2 s funkcí intenzity Λ a Campbellovou mírou C .

a) Pravděpodobnostní míra P_X^x pro $x \in \mathbb{S}_\rho^2$ se nazývá **Palmovo rozdělení** X , pokud platí současně

(i) pro všechna $x \in \mathbb{S}_\rho^2$ (skoro všude vzhledem k Λ) je P_X^x pravděpodobnostní míra na $(\mathcal{N}_{\mathbb{S}_\rho^2}, \mathcal{B}_{\mathcal{N}_{\mathbb{S}_\rho^2}})$,

(ii) pro všechna $B \in \mathcal{B}_{\mathbb{S}_\rho^2}$ a všechny $N \in \mathcal{B}_{\mathcal{N}_{\mathbb{S}_\rho^2}}$ lze hodnoty Campbellovy míry X vyjádřit jako

$$C(B \times N) = \int_B P_X^u(N) \Lambda(du).$$



Obrázek 3.4: Neizotropní bodový proces.

b) Pro všechna $x \in \mathbb{S}_\rho^2$ je **redukováno Palmovo rozdělení** $P_X^{!x}$ definováno vztahem

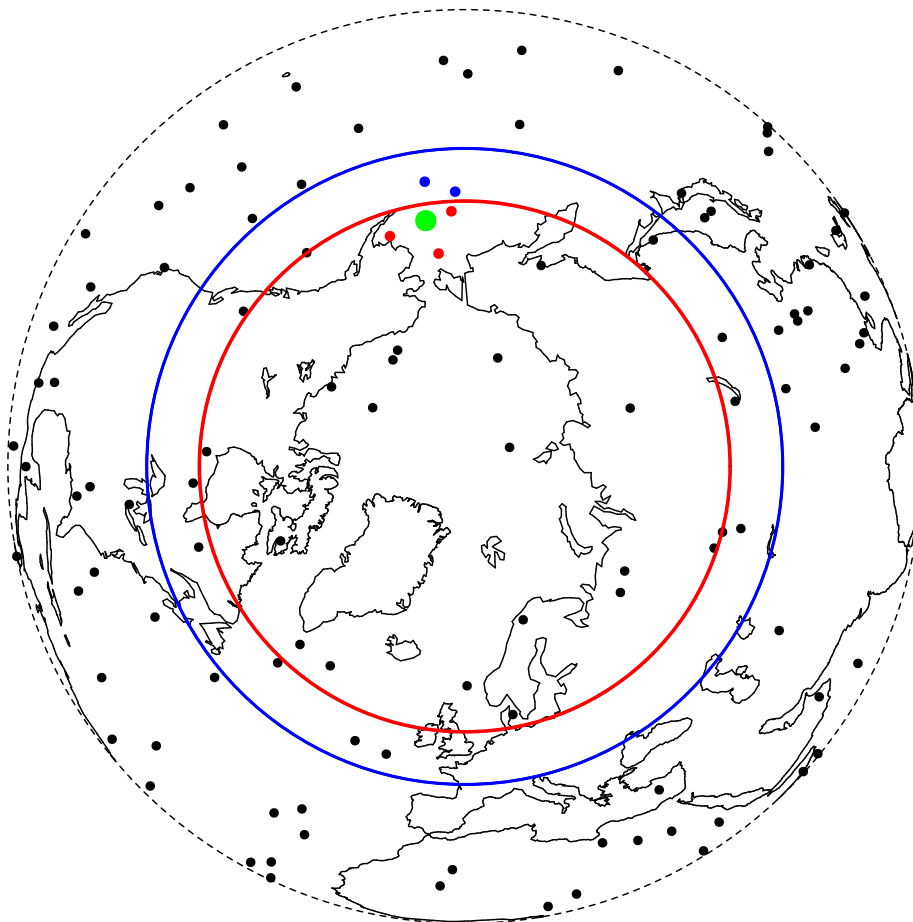
$$P_X^{!x}(N) = P_X^x(X \setminus \{x\} \in N),$$

pro všechny $N \in \mathcal{B}_{\mathcal{N}_{\mathbb{S}_\rho^2}}$.

Pokud bude bodový proces na sféře \mathbb{S}_ρ^2 izotropní, pak podle lemmatu 3.1 v (4) je Palmovo rozdělení invariantní vůči všem rotacím z $SO(3)$. To je v souladu s intuicí, co bychom očekávali. Nyní už jen uvedme minusovou korekci a budeme mít vše připraveno pro K -funkci a její odhad.

Problém okrajového efektu a jak ho vyřešit pomocí minusové korekce jsme diskutovali v první kapitole. Na sféře \mathbb{S}_ρ^2 je myšlenka stejná a minusová korekce principově funguje stejně. Je pouze potřeba dát si pozor, že už nepracujeme s euklidovskou vzdáleností, ale se sférickou vzdáleností, kterou budeme v této i v následující kapitole jednoduše označovat jako r .

Uvedme si jednoduchý příklad pomocí obrázku 3.5. Máme realizaci bodového procesu na sféře a zajímá nás průměrný počet sousedních bodů v okolí bodu procesu. Ovšem typicky nemusíme mít informace o realizaci bodového procesu na celé sféře, ale vidíme jen nějakou část a to pozorovací okno W , které jsme si na obrázku ohraničili modře. Pokud si vezmeme okolí nějakého bodu o dostatečně velkém r , tak část okolí bude ležet mimo pozorovací okno W . Proto provedeme minusovou korekci, kdy si pro daná r kladná, pro která to má smysl, vezmeme podmnožinu pozorovacího okna W takovou, že obsahuje pouze body, které jsou



Obrázek 3.5: Mínusová korekce na sféře.

od hranice W vzdáleny alespoň r . Tu pojmenujeme jako **zmenšené pozorovací okno** $W_{\ominus r}$ (pokud by byla tato množina prázdná, tak pak pro dané r nemá mínusová korekce smysl). Zmenšené pozorovací okno $W_{\ominus r}$ jsme na obrázku ohraničili červeně. Nyní pokud budeme brát pouze body z $W_{\ominus r}$, tak už nám nebude hrozit okrajový efekt a uvidíme všechny jejich sousedy do vzdálenosti r .

Na obrázku jsme si zeleně zvýraznili bod, pro který počítáme počet sousedů v jeho okolí. Takto jsme správně napočítali 5 sousedů, které jsme barevně označili. Pokud bychom ale neprovedli mínusovou korekci a pozorovacím oknem by byla právě jen červeně ohraničená oblast, tak bychom pak dva modře vyznačené body neviděli a nesprávně bychom počet sousedů pomocí pouze červených bodů odhadli 3.

3.3 K -funkce

Definice K -funkce na sféře je analogická té z první kapitoly, pouze přizpůsobená tomu, že bodový proces je nyní na sféře. Opět se jedná o střední hodnotu počtu sousedních bodů v okolí typického bodu (což jsou vlastně body nacházející se v kulovém vrchlíku se středem ve výše zmíněném typickém bodě) vydělený intenzitou. Než si uvedeme definici, připomeňme, že kulový vrchlík označujeme

symbolem $C(u,s)$, kde u je střed kulového vrchlíku a $s \leq \pi\rho$ je sférická vzdálenost libovolného bodu hranice kulového vrchlíku a jeho středu. Pak $X(C(u,s))$ bude značit počet bodů v kulovém vrchlíku $C(u,s)$.

Definice 14. *Nechť X je izotropní bodový proces na sféře \mathbb{S}_ρ^2 s intenzitou $\lambda > 0$. K -funkci pro bod $u \in \mathbb{S}_\rho^2$ definujeme vztahem*

$$K(r) = \frac{\mathbb{E}_X^{!u} X(C(u,r))}{\lambda}, \quad 0 \leq r \leq \pi\rho,$$

kde $\mathbb{E}_X^{!u}$ je střední hodnota vzhledem k redukovanému Palmovu rozdělení $P_X^{!u}$.

Dle lemmatu 4.2 z (4) lze K -funkci vyjádřit i jiným způsobem, který bude na pohled podobnější s později zmíněným odhadem K -funkce. Dle tohoto lemmatu platí pro každou neprázdnou $B \in \mathcal{B}_{\mathbb{S}_\rho^2}$ vztah

$$K(r) = \frac{\mathbb{E}[\sum_{x \in X \cap B} X(C(x,r) \setminus \{x\})]}{\lambda \mathbb{E}X(B)}, \quad 0 \leq r \leq \pi\rho.$$

Použijeme odhad K -funkce z definice 5.4 (a) v (4), který bude souviset s posledním vyjádřením K -funkce. Upozorníme, že tento odhad nebude odhadovat společně s K -funkcí i čtverec intenzity jako v první kapitole, tedy $\lambda^2 K(r)$, ale bude odhadovat $\lambda K(r)$.

Definice 15. *Nechť X je izotropní bodový proces na sféře \mathbb{S}_ρ^2 s nenulovou intenzitou λ a W je pozorovací okno. Pak pro všechna r , pro která platí $0 \leq r \leq \pi\rho$ a $W_{\ominus r}$ je neprázdná množina, je odhad $\lambda K(r)$ roven*

$$\widehat{\lambda K(r)} = \frac{\sum_{x \in X \cap W_{\ominus r}} \sum_{y \in X \cap W}^{x \neq y} \mathbb{1}_{[0,r]}(d(x,y))}{X(W_{\ominus r})}.$$

Jak vidíme, odhad souvisí více s upraveným tvarem K -funkce, neboť pokud za množinu B zvolíme zmenšené pozorovací okno $W_{\ominus r}$, pak střední hodnotu bodů ležících v množině $W_{\ominus r}$ odhadneme pozorovaným počtem bodů v této množině. Dále v čitateli odhadu pro každé x ležící v množině $W_{\ominus r}$ spočteme počet sousedů mající sférickou vzdálenost od x nejvýše r , a tyto počty sečteme, což je vlastně odhad střední hodnoty sumy, kdy v jednotlivých sčítancích máme veličinu udávající počet bodů v kulovém vrchlíku $C(x,r)$ (které jsou různé od středu kulového vrchlíku), což je jinými slovy počet sousedních bodů majících sférickou vzdálenost od bodu x nejvýše r .

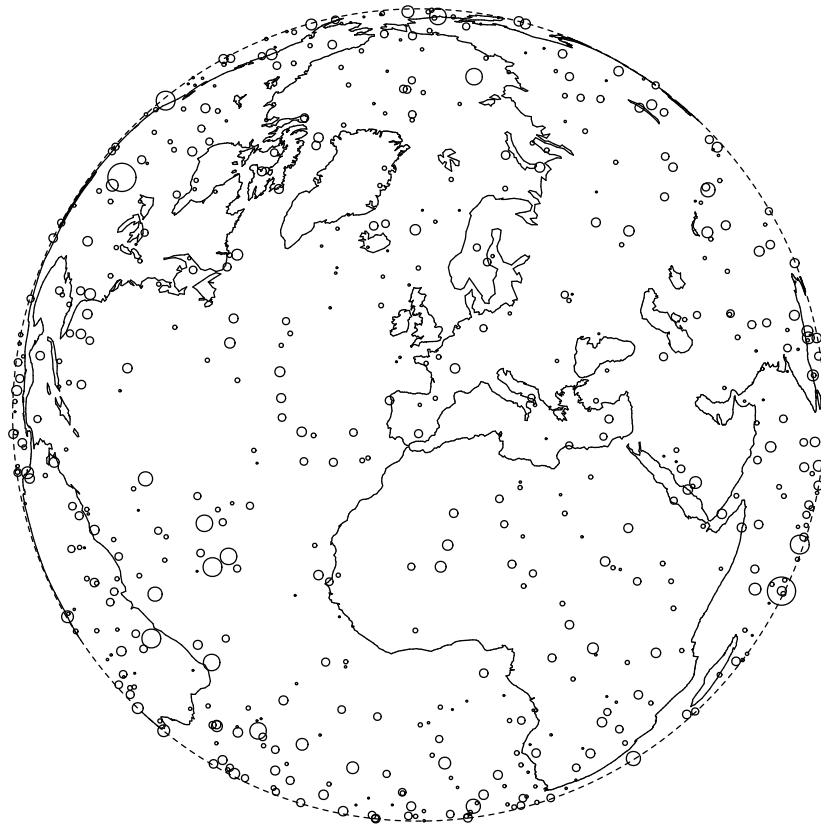
Nyní máme vše připraveno, abychom se mohli věnovat kótovaným bodovým procesům na sféře.

4. Kótované bodové procesy na sféře

4.1 Motivace

V této kapitole chceme pracovat u bodových procesů na sféře s více informacemi, než je poloha bodů. Chceme využívat další netriviální informaci o bodech, a proto má smysl zavádět kótované bodové procesy na sféře. Uvedme si pro ilustraci pár situací z reálného světa, které lze pomocí nich zkoumat.

Příklad 4.1: Jedna ze situací, kterou lze modelovat pomocí bodových procesů na sféře, jsou dopady meteoritů na Zemi a jejich velikost v momentě dopadu. Jedná se o situaci, kdy bez nějakých větších informací nemáme důvod předpokládat, že bude nějaká oblast s výrazně větším výskytem dopadů meteoritů a ani nemáme žádný důvod očekávat, že by menší meteority měly dopadat spíše na jedno území, zatímco větší na jiné. Tedy očekávali bychom rovnoměrné rozložení poloh dopadů meteoritů a nezávislost velikosti meteoritu na poloze dopadu a na velikosti jiných meteoritů. Zmíněnou situaci jsme simulovali a vykreslili na obrázku 4.1.



Obrázek 4.1: Polohy dopadů meteoritů jsou zaznamenány pomocí polohy středu kroužku a velikost meteoritu v okamžiku dopadu (tedy proporcionální velikost kóty) odpovídá poloměru kroužku.

Obdobně můžeme postupovat i u příkladu 1.1, tedy můžeme se nejen ptát a zkoumat, zda rozmístění kráterů na Měsíci je v nějakém smyslu rovnoměrné, ale i zda velikosti kráterů jsou nezávislé a stejně rozdělené. Hypotéze a testování, zda jsou kóty nezávislé a stejně rozdělené (zkráceně i.i.d.), se budeme věnovat na konci kapitoly.

Příklad 4.2: Můžeme navázat na příklad 3.1 s nebeskou sférou a zkoumáním bodů zastupující objekty ve vesmíru, případně jen hvězdy. Z povrchu Země vidíme vždy jen část noční oblohy a už párkrát jsme si mohli položit otázku, zda je některá část oblohy jasnější než jiná. Můžeme si tedy vzít například nějaké okolí Polárky na nebeské sféře a zkoumat, zda má takový bodový proces na sféře místa s většími velikostmi kót než je obecně obvyklé na jiných oblastech sféry.

Příklad 4.3: Zemětřesení je další situací, kterou lze zkoumat pomocí kótovaných bodových procesů na sféře a zároveň nás u ní nezajímá jen poloha, ale hlavně i velikost kóty, v tomto případě magnitudo zemětřesení, které se může měřit pomocí škály známé pod názvem Richterova stupnice. Na obrázku 4.2 vidíme zemětřesení zaznamenaná USA v roce 1985, využili jsme data ze zdroje (6). Poloha vykreslených bodů určuje polohu epicenter odpovídajících zemětřesení, zatímco velikost symbolů určuje magnitudy daných zemětřesení. Pro lepší orientaci, na kterých místech byla jak silná zemětřesení, jsme obarvili zeleně body zastupující epicentrum zemětřesení s magnitudou do 3, žlutě jsou zemětřesení s magnitudou mezi 3 a 4, červená je pro magnitudu od 4 do 5 a tmavým plným puntíkem jsou zemětřesení s magnitudou větší než 5. Nejsilnější tehdy zaznamenané zemětřesení podle dostupných dat mělo magnitudo 6,6, což je považováno za silné zemětřesení. Na obrázku vidíme nejen, že se většina zemětřesení odehrávala na západním pobřeží USA a okolí Aljašky, ale právě zde byla zaznamenána největší zemětřesení.

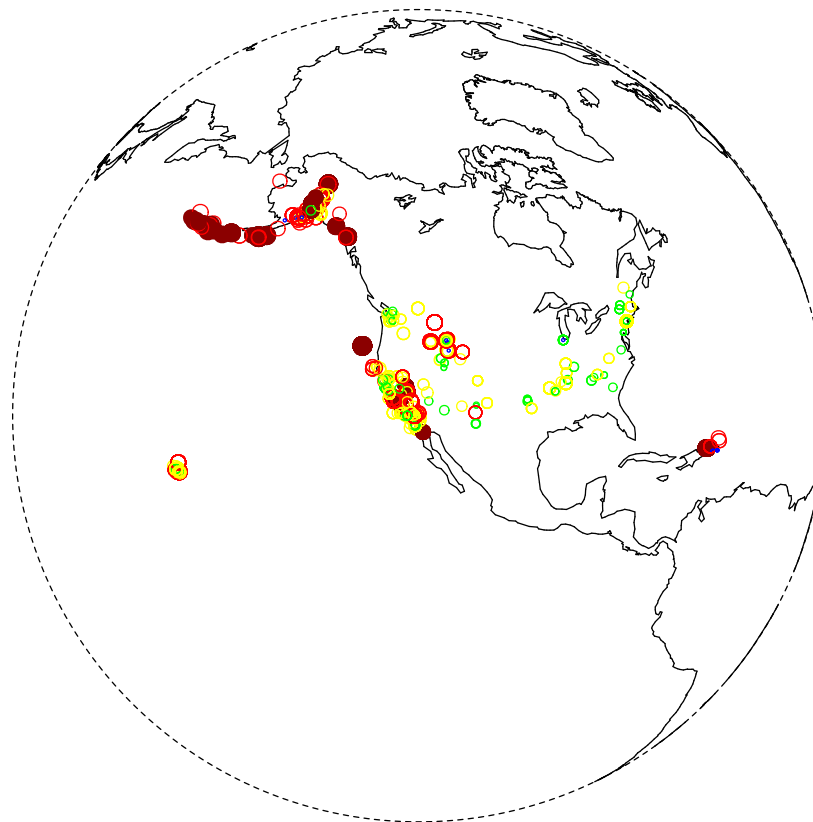
4.2 Definice a vlastnosti

Tato kapitola je přímou adaptací druhé kapitoly, která je věnována kótovaným bodovým procesům v euklidovském prostoru, a třetí kapitoly, která je věnována bodovým procesům na sféře. Budeme proto při odvozování a vytváření definic vycházet z teoretických poznatků těchto dvou kapitol.

Nejprve zdefinujeme kótované bodové procesy na sféře.

Definice 16. *Kótovaný bodový proces na sféře S_ρ^2 s kótami v prostoru Θ je bodový proces M na prostoru $S_\rho^2 \times \Theta$ takový, že počet bodů M v množině $K \times \Theta$ je konečný skoro jistě pro všechna kompaktní $K \subset S_\rho^2$. To znamená že odpovídající podkladový proces (to je bodový proces, který vznikne z M , když odstraníme kóty) je lokálně konečný.*

Kóty mohou být vícerozměrné, tedy v každé složce vektoru mohou nést nějakou jinou informaci. Například v příkladě 4.3 kdybychom se neomezili pouze na polohu epicentra (= poloha bodu) a na magnitudo daného zemětřesení, které nám dávalo jedinou další informaci o daném zemětřesení, ale využili i další informace z dat a neomezili se pouze na rok 1985, mohli bychom tak mít jako kótu vektor,



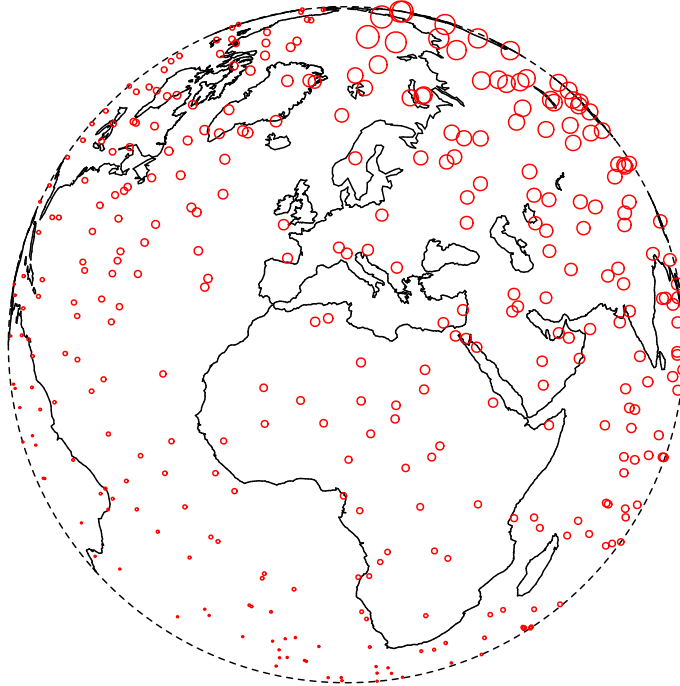
Obrázek 4.2: Zemětřesení zaznamenaná USA v roce 1985.

který by v dalších složkách mohl mít čas, kdy přesně začalo zemětřesení, hypocentrum (přesněji jak hluboko pod epicentrem se nachází hypocentrum, tedy místům, který je skutečným středem zemětřesení) nebo i kategoriální veličinu, která svou hodnotou určuje zdroj informací o daném zemětřesení, jak je uvedeno v (6).

V naší práci budeme převážně pracovat pouze s jednorozměrnými kótami, u kterých jsme v druhé kapitole uváděli nejzákladnější druhy, se kterými se můžeme setkat. Pouze v jednom ilustračním příkladě později zmíníme situaci, kdy kótovaný bodový proces na sféře bude mít vícerozměrnou kótu, jinak obecně mlčky budeme předpokládat jednorozměrnou kótu.

Kóty se mohou navzájem ovlivňovat nebo být dokonce závislé na poloze bodu. Jednoduchým příkladem, kdy poloha bodu a kóta tohoto bodu budou závislé, je například sledování mezistátní dopravy. Poloha každého dopravního prostředku bude v bodovém procesu na sféře zastoupena polohou bodu, typ dopravního prostředku bude zaklíčován prostřednictvím kóty. Pak se body s kótami, které odpovídají vlakové dopravě, budou nacházet nutně na pevnině, zatímco body s kótami odpovídající námořní dopravě se typicky budou nacházet mimo pevninu nebo na okraji pevniny, pokud loď bude v přístavu. Poznamenejme, že i když se body s kótami odpovídající vzdušné dopravě mohou nacházet během letu jak nad pevninou, tak nad mořem, tak část dopravních prostředků nebude v danou chvíli v provozu, ale bude na letištích, tedy bude docházet ke shlukování bodů se stejnou kótou.

Ještě jednodušším příkladem pro závislost kóty a polohy bodu je bodový pro-



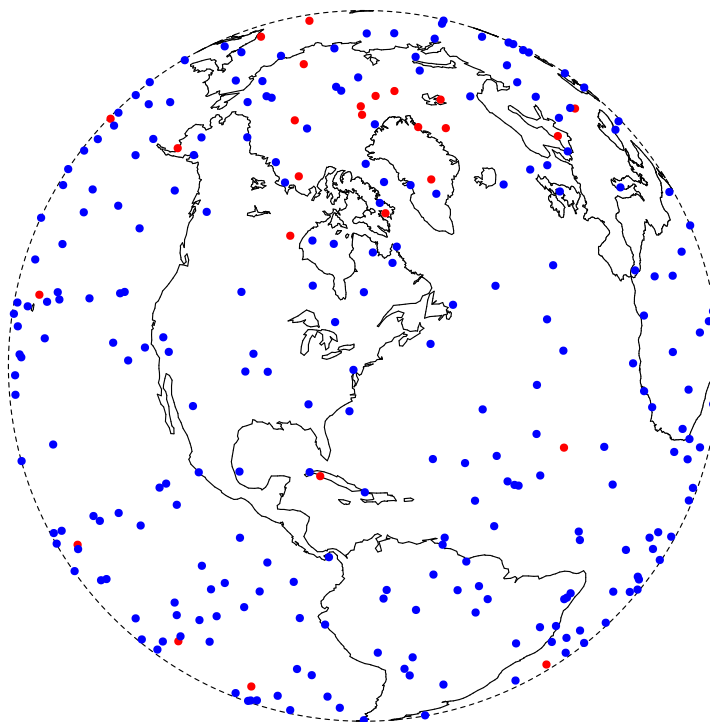
Obrázek 4.3: Body nemají stejně rozdělené kóty - velikost závisí na poloze.

ces s kótami pocházejícími ze spojitého rozdělení, kdy velikost kóty je měřitelnou funkcí polohy bodu, jak to můžeme vidět i na realizaci na obrázku 4.3, kdy body v jižním Atlantickém oceánu jsou vizuálně opravdu malé, na severu Afriky se nachází body se středně velkou kótou a v okolí severní části Asie je velikost symbolů na pohled opravdu velká. Jedná se o simulovaná data, která nemají v reálném životě žádnou skutečnou interpretaci, pouze nám slouží jako jednoduchý ilustrační příklad.

Důležitou vlastností kótovaných bodových procesů na sféře je izotropie. Připomeňme, že kótovaný bodový proces lze zapsat jako $M = \{[x_n; m(x_n)]\}$, kde $m(x_n)$ je kóta bodu x_n .

Definice 17. *Kótovaný bodový proces M na sféře S_ρ^2 nazveme **izotropním**, pokud jeho rozdělení je invariantní vzhledem ke grupě rotací $SO(3)$, to jest rozdělení bodového procesu po rotaci $RM = \{[R(x_n); m(x_n)]\}$ je stejné jako rozdělení M pro všechna $R \in SO(3)$.*

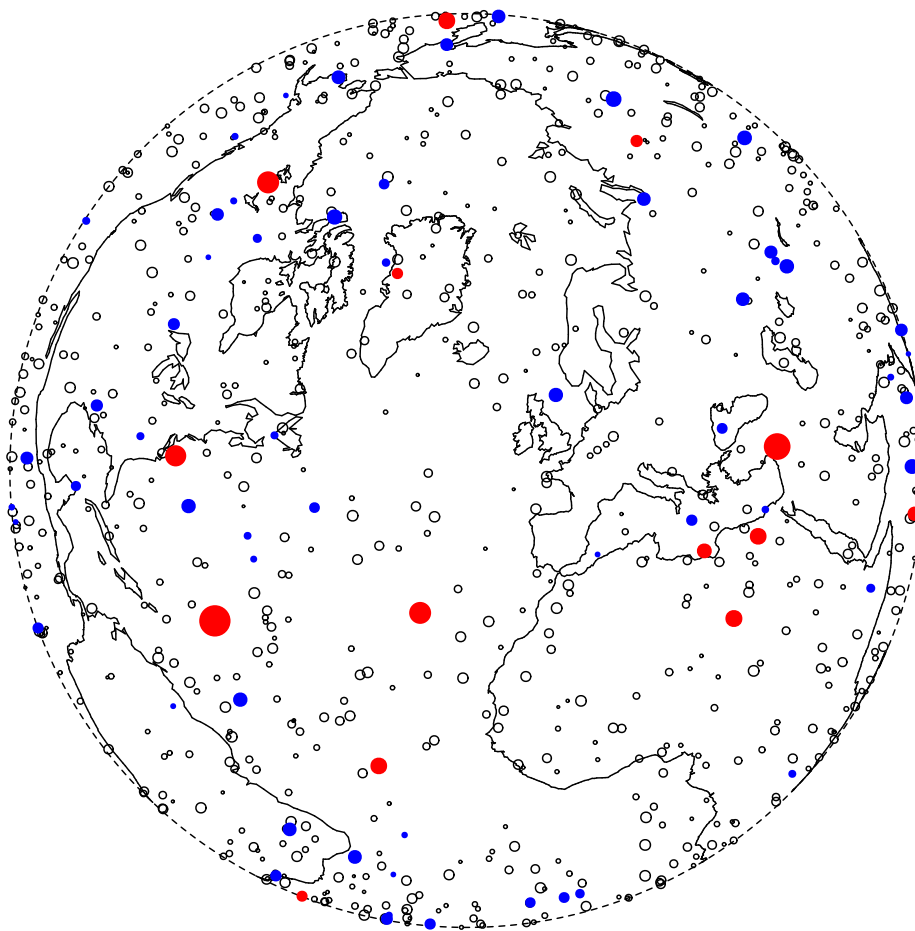
Definice je analogická k definici ve třetí kapitole, avšak pro kótované bodové procesy nám implikuje stejné rozdělení kót. Intuitivně pokud máme libovolné pozorovací okno W a po jeho rotaci RW se nemá změnit rozdělení bodového procesu, které pomocí RW pozorujeme, tak by se měly vlastnosti rozdělení zachovat včetně rozdělení kót, které bychom v novém pozorovacím okně RW očekávali. Naopak, jestli rotací pozorovacího okna očekáváme významnou změnu velikostí kót, pak zde zjevně nebude stejné rozdělení jako v původním okně a není tak splněna definice izotropního kótovaného bodového procesu. Dále, pokud by kóty nebyly stejně rozdělené, pak by kótami převážená K -funkce závisela na tom, jaký bod bychom zvolili jako „typický“, neboť jak už jsme viděli u kótami převážené K -funkce v druhé kapitole, tak číselně závisí na hodnotě kóty typického



Obrázek 4.4: Červené body jsou nerovnoměrně rozložené po sféře, modré body jsou rovnoměrně rozdělené na sféře.

bodů. Což už by nebyla tak vhodná funkce, která by měla v nějakém smyslu charakterizovat izotropní (pro druhou kapitolu stacionární) bodový proces, když by závisela na místě, kde jí počítáme, zatímco rotací (respektive v druhé kapitole posunutím) se vlastnosti bodového procesu v pozorovacím okně nemají měnit, tedy má zůstat stejná rozdělenost. Implikace stejné rozdělenosti kót při předpokladu izotropie je tedy plně v souladu s naší intuicí.

Na obrázku 4.4 jsme nasimulovali situaci, na kterou se můžeme dívat ze dvou pohledů. První pohled popisuje přímo naši „konstrukci“ obrázku. Provedli jsme realizaci izotropního bodového procesu, kde jsme body na sféře zobrazili modře. Nezávisle na procesu s modrými body jsme simulovali neizotropní bodový proces s červenými body, kde daná realizace má výrazné shlukování v oblasti severního pólu. Při tomto pohledu můžeme porovnávat izotropní a neizotropní bodové procesy. Pokud bychom ovšem neznali postup simulování, pak bychom mohli na celou situaci pohlížet jiným pohledem, kdy bychom obrázek vnímali jako realizaci kótovaného bodového procesu, kde kóty mohou nabývat jedné ze dvou hodnot. Tento pohled podporuje v rozvoji naší intuice nečekat, že pokud kótovaný bodový proces nebude izotropní, tak bychom stále mohli očekávat nepříliš odlišné vlastnosti u bodů se stejnou kótou. V tomto příkladě je rozdíl v povaze natolik extrémní, že kóty jednoho druhu tvoří izotropní bodový proces, zatímco kóty jiného typu velice razantně porušují pravidelnost, dochází k výraznému shlukování do jedné oblasti. Další kótovaný bodový proces, při kterém body se stejnou kótou mohou mít jiné vlastnosti než body s jiným typem kóty, jsme uváděli výše v příkladu s dopravními prostředky.



Obrázek 4.5: Kótovaný bodový proces s vícerozměrnou kótou.

Ne vždy lze u kótovaného bodového procesu jednoduše poznat, zda lze předpokládat izotropie nebo ne. Je obecně potřeba znát okolnosti, které ovlivňují body i kóty. Bez hlubších znalostí o kótovaném bodovém procesu nelze obecně dělat žádné závěry, například u realizace na obrázku 4.5 bychom těžko odhadovali, zda izotropii lze předpokládat. Jedná se o realizaci složitějšího bodového procesu s vícerozměrnou kótou, kde v jedné složce máme informaci do jaké kategorie bod patří (na obrázku rozlišeno barevně, pro přehlednost modré a červené body jsou vyobrazeny plným puntíkem) a v druhé složce je spojitá veličina (na obrázku rozlišeno velikostí symbolu). Složky kóty jsou přitom vzájemně závislé. Červené body dosahují větších velikostí než modré a černé dosahují obecně menších velikostí než modré a červené. Také vidíme různou četnost bodů. Přitom je to realizace bodového procesu, který splňuje definici izotropního bodového procesu. Tento proces lze simulovat tak, že při vygenerování polohy se nejdříve určí kategorie bodu bez závislosti na poloze (u každé kategorie budeme mít pravděpodobnost s jakou se vyskytne, jedná se tedy o multinomické rozdělení), pak se určí velikost symbolu, která bude pocházet ze spojitého rozdělení příslušného k dané kategorii bodu. Žádná složka kóty tedy nijak nezávisí na poloze.

4.3 K -funkce

Připomeňme, že upravený tvar K -funkce pro izotropní bodové procesy na sféře má pro každou neprázdnou $B \in \mathcal{B}_{\mathbb{S}_\rho^2}$ tvar

$$K(r) = \frac{\mathbb{E}[\sum_{x \in X \cap B} X(C(x,r) \setminus \{x\})]}{\lambda \mathbb{E}X(B)}, \quad 0 \leq r \leq \pi\rho.$$

Obdobně jako ve druhé kapitole budeme opět chtít tuto funkci upravit tak, aby pracovala i s velikostmi kót. Přepíšme číselník, který pracuje s počtem sousedních bodů na kulovém vrchlíku. Připomeňme značení pro sférickou vzdálenost bodů x a y na sféře pomocí $d(x,y)$.

$$K(r) = \frac{\mathbb{E}[\sum_{x \in X \cap B} \sum_{y \in X \cap B}^{x \neq y} \mathbb{1}_{[0,r]}(d(x,y))]}{\lambda \mathbb{E}X(B)}, \quad 0 \leq r \leq \pi\rho.$$

Nyní už tento tvar modifikujeme pro kótovaný bodový proces na sféře, kde každý indikátor převážíme součinem hodnot kót v bodech, pro které vyhodnocujeme daný indikátor. Jmenovatel přenásobíme kvadrátem střední hodnoty kóty μ . Definujme pro izotropní kótovaný bodový proces s kvantitativními kótami (kóty mají konkrétní numerický význam, hodnoty lze smysluplně uspořádat a rozdíly hodnot mají reálnou interpretaci) a pro každou neprázdnou $B \in \mathcal{B}_{\mathbb{S}_\rho^2}$ **kótami převáženou K -funkci bodového procesu na sféře \mathbb{S}_ρ^2 předpisem**

$$K_{mm}(r) = \frac{\mathbb{E}[\sum_{[x;m(x)]}^{x \in M \cap B} \sum_{[y;m(y)]}^{x \neq y \in M} m(x)m(y) \mathbb{1}_{[0,r]}(d(x,y))]}{\lambda \mu^2 \mathbb{E}X(B)}, \quad 0 \leq r \leq \pi\rho.$$

Pokud bychom přidali předpoklad nezávislosti kót, analogicky jako v druhé kapitole lze ukázat, že se kótami převážená K -funkce bodového procesu na sféře zjednoduší na K -funkci bodového procesu na sféře uvedené v třetí kapitole.

Obdobně jako v třetí kapitole odhadneme $\lambda^2 \mu^2 K_{mm}$ přímočaře s využitím mínusové korekce pro pozorovací okno W a r , pro které platí $0 \leq r \leq \pi\rho$ a $W_{\ominus r}$ je neprázdná množina, pomocí odhadu

$$(\lambda \mu^2 \widehat{K_{mm}}(r)) = \frac{\sum_{[x;m(x)]}^{x \in M \cap W_{\ominus r}} \sum_{[y;m(y)]}^{x \neq y \in M} m(x)m(y) \mathbb{1}_{[0,r]}(d(x,y))}{X(W_{\ominus r})}. \quad (4.1)$$

Od začátku práce jsme budovali teorii abychom se mohli dostat právě k tomuto odhadu, neboť s tímto budeme pracovat při testování hypotézy, zda kótovaný bodový proces na sféře má nezávislé kóty.

4.4 Testování nezávislosti kót

4.4.1 Hypotéza a testová statistika

Uvažujme model, ve kterém máme izotropní kótované bodové procesy na sféře \mathbb{S}_ρ^2 . Naši nulovou hypotézou je nezávislost kót proti alternativě, že kóty nejsou nezávislé. V modelu, ve kterém uvažujeme pouze izotropní kótované bodové procesy na sféře, bude platit, že kóty jsou stejně rozdělené. Naše nulová hypotéza je tak ekvivalentní s hypotézou i.i.d. kót.

Podotkněme, že velikost kóty může záviset na velikosti jiné kóty a přitom stále být ze stejného rozdělení. Představme si náhodnou veličinu U , která pochází z normálního rozdělení se střední hodnotou rovnou nule. Pak náhodná veličina $-U$ pochází ze stejného rozdělení, avšak s náhodnou veličinou U je závislá dokonce lineárně. Za alternativy bychom tedy mohli čekat ovlivňování velikostí kót například velikostí kóty sousedního bodu. Může docházet ke kladné korelaci kót blízkých bodů (uvidíme hned u následujícího příkladu se zemětřeseními) a nebo naopak může docházet k záporné korelaci kót blízkých bodů (v okolí bodu s velkou kótou budou spíše body s malými kótami a obdobně v okolí bodu s malou kótou budou spíše body s velkými kótami).

Jak bylo zmíněno dříve, testovou statistikou bude odhad kótami převážené K -funkce pro kótované bodové procesy na sféře definovaný rovnicí (4.1). Pokud kóty nebudou i.i.d., může docházet k nějaké nepravidelnosti, například častěji budou u sebe kóty podobné velikosti jak jsme viděli na obrázku 4.2. Na Aljašce zcela jasně docházelo k situaci, kdy vedle sebe byly velice silná zemětřesení, zatímco ve východní části USA měly jednotlivé body za „sousedy“ body zastupující epicentra méně silných zemětřesení. V našem odhadu se tak budou často potkávat velké hodnoty kót s dalšími velkými hodnotami kót a dojde k výraznému převážení dané dvojice bodů (daný sčítanec bude typicky mnohem větší pro stejnou hodnotu parametru r než bychom dostali pro i.i.d. body, v extrémním případě, v závislosti na možných velikostech kót, by dokonce většina sčítanců mohla být zanedbatelná oproti těmto velkým sčítancům). Odhad by v takovém případě mohl nabývat větších hodnot, než bychom běžně očekávali. Na druhou stranu, odhad může dosahovat i menších hodnot. Nastane to u kótovaných bodových procesů na sféře, které mají záporně korelované kóty (v okolí bodů s velkou kótou budou spíše body s malými kótami a v okolí bodů s malými kótami uvidíme spíše body s velkými kótami). Za extrémní hodnoty testové statistiky bereme tedy velmi malé i velmi velké hodnoty, které zkráceně nazveme extrémními hodnotami.

Rozdělení testové statistiky za nulové hypotézy neznáme. Je velice složité, neboť testová statistika je funkcionální. Proto k testování musíme přistupovat jinak. Nevíme z jakého rozdělení data pochází - neumíme vygenerovat nové polohy bodů ani nové velikosti kót, protože ani u těch nevíme, z jakého rozdělení pochází. Můžeme ale permutovat kóty a budeme tak testovat pomocí permutačního Monte Carlo testu.

4.4.2 Testování hypotézy

Mějme realizaci bodového procesu, která má celkový počet bodů právě $n \in \mathbb{N}$, a pro každé $i = 1, \dots, n$ známe polohu bodu x_i a jeho kótu $m(x_i)$. Máme tedy data $\{[x_i; m(x_i)]\}$ a označme si $T_0(r) = \lambda^2 \mu^2 \widehat{K_{mm}}(r)$, $r \in (0, R)$, kde $R \leq \pi \rho$ je vhodně zvolené kladné reálné číslo tak, aby zmenšené pozorovací okno $W_{\ominus r}$ bylo stále neprázdnou množinou.

Mějme s náhodných permutací na množině $\{1, \dots, n\}$ a označme je π_1, \dots, π_s .

Pro permutaci π_j , $j = 1, \dots, s$ získáváme permutovaná data $\{[x_i; m(x_{\pi_j(i)})]\}$. Pro každá permutovaná data spočteme $T_j(r) = \lambda^2 \mu^2 \widehat{K_{mm_j}}(r)$, $r \in (0, R)$.

Nyní když máme vyhodnocenou testovou statistiku pro všech s náhodných permutací i pro původní data, budeme chtít popsat, jak typická či extrémní je $T_0(r)$ v populaci $\{T_0(r), T_1(r), \dots, T_s(r)\}$.

K tomu potřebujeme ranking (neboli pořadí), například v prostorové statistice běžné „erl“ pořadí, které se používá v obálkových testech uvedené v (7). Pokud funkce bude mít pořadí nízké, pak se bude jednat o extrémní funkci, zatímco vysoké pořadí bude odpovídat typickým funkcím.

Pokud $T_0(r)$ odpovídající naším původním datům bude patřit mezi 5% nejextrémnějších případů, pak jsme pozorovali data natolik extrémně odlišná od toho, co bychom očekávali za nulové hypotézy, že zamítneme ve prospěch alternativy. Prokázali bychom tak, že kóty nejsou nezávislé.

4.4.3 Realizace

Testová statistika

Pro testování hypotézy bylo potřeba odhad kótami převážené K -funkce implementovat v programu R. Kód je součástí přílohy. Kótami převáženou K -funkci vyhodnocujeme celkově v 512 ekvidistantních bodech. Nejmenší hodnota argumentu je 0 a nejvyšší hodnotou je $\frac{\pi}{2}$. Rozhodli jsme se pro tuto volbu z toho důvodu, že je to tak i u nekótované verze `Ksphere` z balíčku `spherstat`, kterou budeme porovnávat nejen s naší kótami převáženou K -funkcí, ale při vyhodnocení testování hypotézy tyto dvě funkce od sebe odečteme. Nyní je porovnáme pro Poissonův bodový proces.

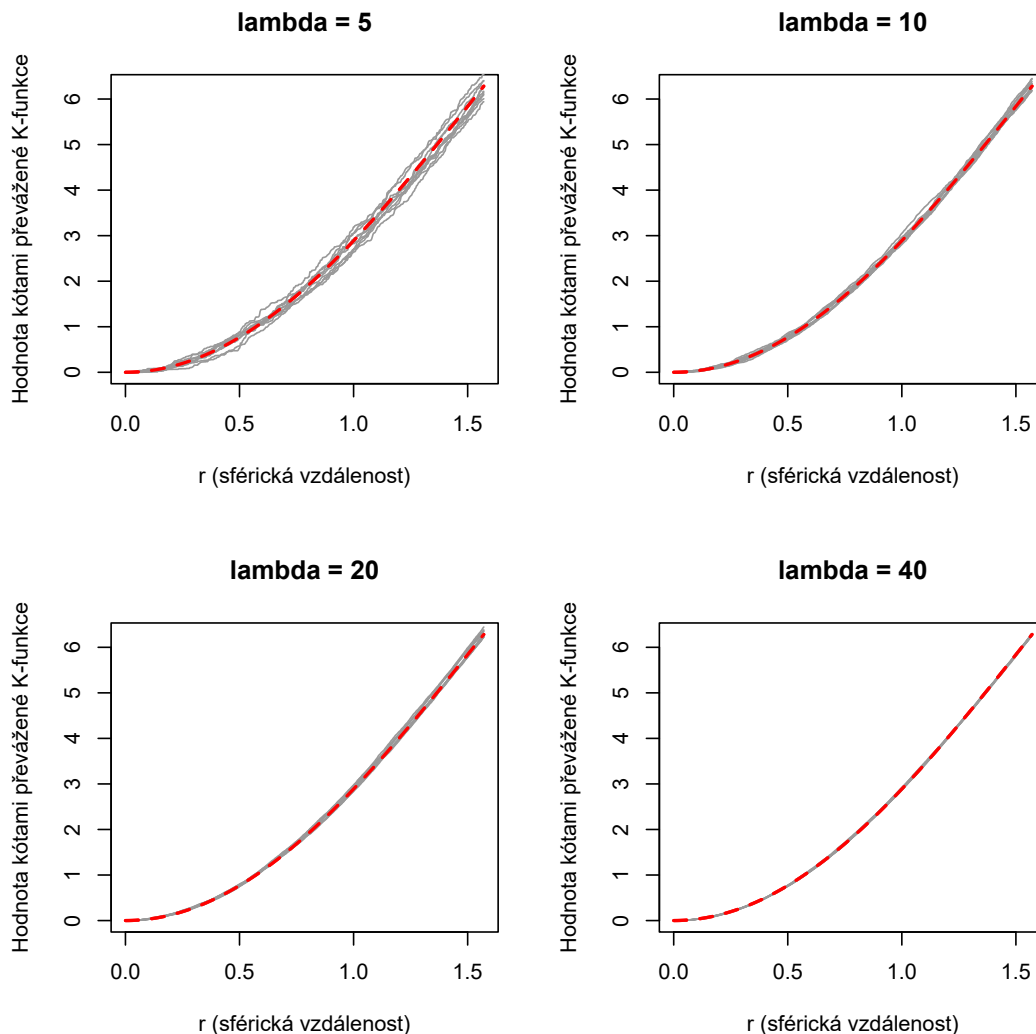
Víme, že v případě bodového procesu s nezávislými kótami se kótami převážená K -funkce redukuje na K -funkci definovanou ve třetí kapitole. Pro ověření správnosti naší implementace využijeme teoretickou hodnotu K -funkce Poissonova bodového procesu pomocí funkce `Ksphere` a budeme sledovat, zda náš implementovaný odhad kótami převážené K -funkce bude pro Poissonův bodový proces s nezávislými kótami konvergovat k této teoretické hodnotě s rostoucím počtem pozorovaných bodů.

Pro Poissonův bodový proces s argumentem λ platí, že střední počet bodů na množině o velikosti 1 bude roven λ . Na sféře o poloměru 1 vygenerujeme 10 Poissonových bodových procesů postupně pro každou hodnotu λ , které položíme rovno 5, 10, 20 a 40. Bodům budeme nezávisle na sobě přiřazovat kóty rovnoměrně nabývajících hodnot z intervalu $[0; 100]$. Dáme-li do společného grafu všechny křivky odpovídající odhadům se stejným parametrem Poissonova bodového procesu λ , můžeme posoudit, jak moc graficky se blíží naše odhady k teoretické hodnotě. Na obrázku 4.6 vidíme, že pro $\lambda = 20$ dochází k věrnému kopírování teoretické hodnoty, pro $\lambda = 40$ odhady už vizuálně nelze rozlišit od teoretické hodnoty. Náš implementovaný odhad se chová tak, jak očekáváme.

Data

K testování hypotézy využijeme data zemětřesení ze zdroje (8)¹, kde máme informaci o polohách a síle zemětřesení, která byla zaznamenaná v období od

¹Je potřeba u Event Type zaškrtnout Earthquake, jinak ve výběru budou zahrnuty i jiné události

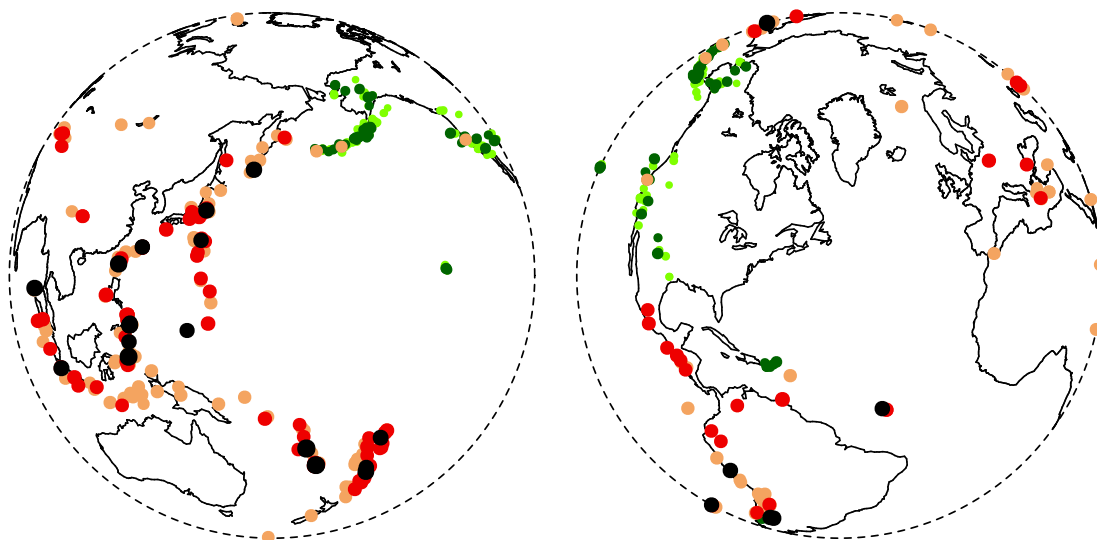


Obrázek 4.6: Odhadování kótami převážené K -funkce pro Poissonovy bodové procesy s parametrem λ a nezávislými kótami. Červeně je vyznačená teoretická hodnota nekótované verze K -funkce.

1. dubna do 7. dubna roku 2022 a zároveň magnitudo zemětřesení bylo alespoň 2,5. Celkem máme 515 bodů. Největší dosažené magnitudo v našem vzorku je rovno přesně 6 (hodnoty magnituda bývají zaokrouhlovány na desetiny, v některých případech na setiny).

Na obrázku 4.7 si můžeme prohlédnout polohy epicenter zemětřesení z našeho vzorku. Většina bodů leží na okrajích litosférických desek, zejména na okrajích Pacifické desky a podél jihovýchodní části Euroasijské desky. Naopak v oblasti Atlantického oceánu a jeho okolí v těchto sedmi dnech téměř nebyla zaznamenána zemětřesení s magnitudem alespoň 2,5.

Velikost bodu odpovídá velikosti magnituda, pro lepší grafickou přehlednost jsme body rozdělili do několika kategorií a dle toho body obarvili. Světle zelené body mají hodnoty v intervalu $[2,5; 3)$, tmavě zelené body odpovídají intervalu $[3; 4)$. Vidíme, že prakticky všechny body leží podél pobřeží Severní Ameriky, obzvláště u Aljašky se nachází významně velký shluk. Oproti zeleným bodům jsou body ostatních kategorií více roztroušené po povrchu Země. Třetí skupinu



Obrázek 4.7: Polohy a magnituda zemětřesení zaznamenaná za období 1. až 7. dubna 2022.

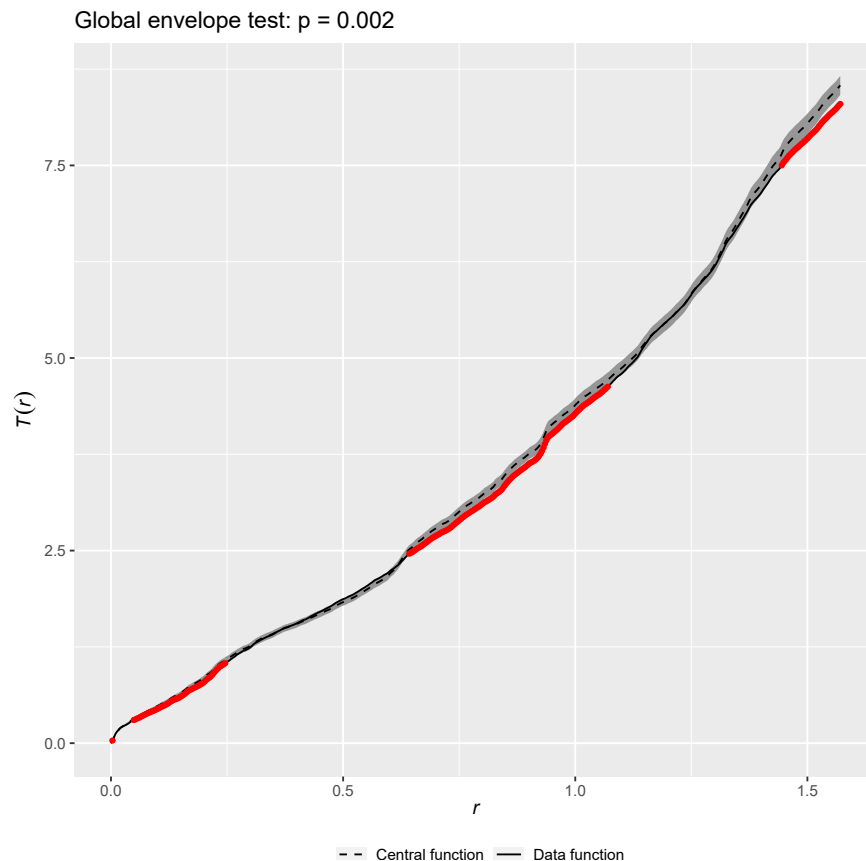
tvoří béžové body s intervalem $[4; 4,5)$, červené body odpovídají intervalu $[4,5; 5)$ a černé body jsou epicentry nejsilnějších zemětřesení, jejich magnituda jsou alespoň 5.

Pokud bychom se chtěli více podívat na hlavní charakteristiky kót, mimo zmíněného minima a maxima je pozoruhodné zmínit první kvartil, který je 2,9. Tedy více než čtvrtina dat náleží první kategorii, i když díky zakrytí světle zelených bodů tmavě zelenými body to může být při pohledu na obrázek zprvu mírně překvapující. Všechny zelených bodů je dohromady 189, což je zhruba 36,7 %. Téměř 300 bodů spadá do intervalu $[4; 5)$, proto taky byl výše rozdělen do dvou kategorií a lze zde najít medián, který je 4,2. Pouze 33 bodů je obarveno černě. Průměr 3,9 napovídá o významném vlivu zelených bodů, které nám mohou výrazně ovlivnit i testovou statistiku.

Vyhodnocení

Vzhledem k časové náročnosti byla testová statistika vyhodnocena pro původní data a právě 999 permutací, neboť vyhodnocování jedné permutace dat o 515 bodech trvalo okolo 90 sekund.

Jak jsme zmínili dříve, využijeme globální obálkový test implementovaný v R v balíčku GET, zaměřený na „erl“ pořadí. Testujeme na hladině $\alpha = 0,05$ s využitím 1000 vyhodnocených testových statistik v 512 bodech, přičemž jedna z těchto

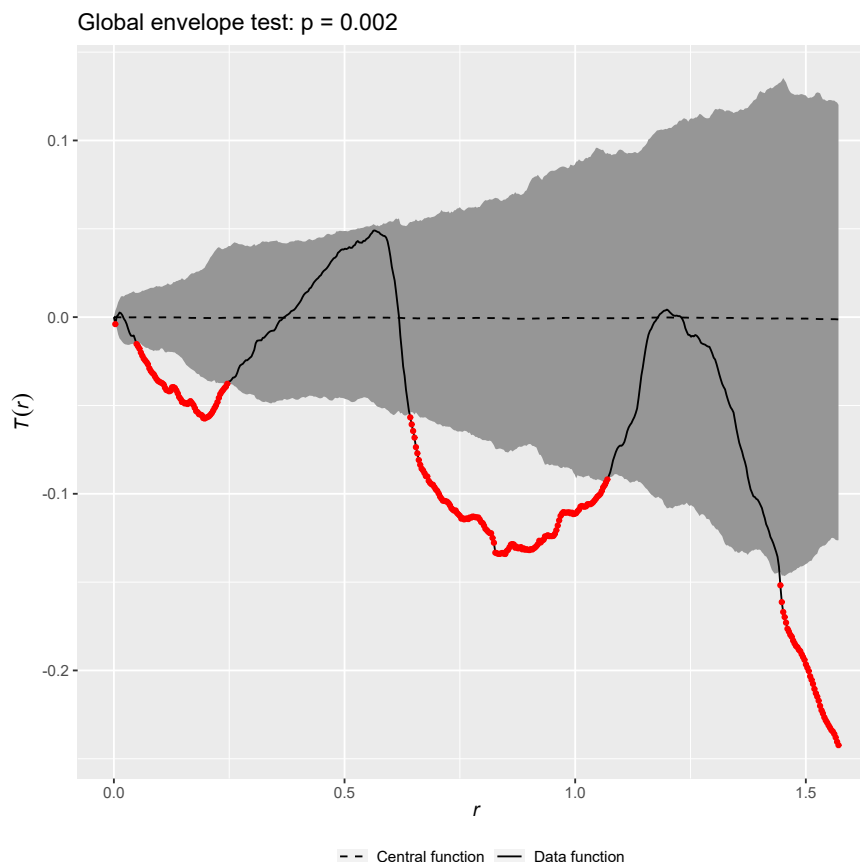


Obrázek 4.8: Obálkový test - nulovou hypotézu zamítáme.

statistik odpovídá původním datům a zadává se jako samostatný argument. Zajímá nás celkově extrémnost, tedy nebudeme se vymezovat pouze na porušení dané velkými hodnotami nebo pouze malými, ale vybereme si oboustrannou verzi testu.

Grafickým výstupem testu je obrázek 4.8. V tmavě šedivé oblasti obsahující čárkovanou křivku se nachází 95% nejtypičtějších křivek a křivka odpovídající původním datům je spojitě vykreslena černě, avšak pokud v některém z bodů bude hodnota testové statistiky ležet mimo šedivou oblast, tak v daném místě bude vykreslen červený bod (na dalším obrázku bude možno vidět, že červené obarvení křivky není spojité). Výhodou takového výstupu je nejen to, že víme, že naše data patří k 5% nejextrémnějších případů, ale také vidíme pro které hodnoty argumentu je testová statistika původních dat nejen extrémní, ale také zda je netypicky příliš velká, či malá. Ihned lze spatřit tři významné intervaly argumentu, ve kterých je naše testová statistika příliš malá.

Lze si všimnout, že pro náš případ bodového procesu máme už tolik bodů, že i při permutování kót výrazné změny hodnot nenastávají a obálka tvořena typickými křivkami je vizuálně velice úzká. Pro zřetelnější grafický výstup jsme tedy od všech statistik odečetli odhad nekótované verze K -funkce vypočtený pomocí příkazu `Ksphere`. Dostaneme tak „vycentrovaný“ grafický výstup, který vidíme na obrázku 4.9. Zde lze mnohem lépe usoudit, jak moc netypická je křivka odpovídající původním datům oproti 95% nejtypičtějším křivkám. Celkově v 248 z 512 bodů leží body testové statistiky mimo šedivou obálku, navíc p -hodnota testu je 0,002, proto můžeme na hladině 0,05 zamítnat nulovou hypotézu ve prospěch



Obrázek 4.9: Obálkový test při odečtení nekótované K -funkce.

alternativy, že kóty nejsou nezávislé.

Zamítnutí hypotézy v situaci se zeměřeseními není překvapivé, šlo to očekávat a také kvůli tomu byla taková data vybrána jako ilustrační příklad, u kterého by mohlo nastat zamítnutí nulové hypotézy ve prospěch alternativy. Překvapivější může být zamítnutí právě kvůli nízkým hodnotám. Uvedeme tedy některé faktory, díky kterým byla testová statistika původních dat netypicky malá pro většinu hodnot argumentu.

Připomeňme, že více než polovina kót byla v poměrně úzkém intervalu $[4; 5)$, zatímco menších hodnot mimo tento interval bylo výrazně více než hodnot velkých alespoň 5, konkrétně černých bodů bylo pouze 33 oproti 189 zeleným, které celkově tvořily více než třetinu celého vzorku. Podle prvního kvartilu s hodnotou 2,9 víme, že více než čtvrtina bodů je světle zelených a také celkový průměr byl pouze 3,9. To všechno v kombinaci s polohami bodů už vysvětluje, proč jsme po permutování kót dosahovali typicky větších hodnot testové statistiky, neboť zatímco nezelené skupiny původních bodů byly relativně hodně roztroušené po sféře, zelené body se nacházely prakticky jen u Severní Ameriky a to hlavně s obrovským shlukem u Aljašky, kde bylo celkově kolem 100 bodů, tedy přibližně pětina celého vzorku. Stačí, aby permutace kót dala část větších kót k Aljašce, pak každý bod, který bude mít ve svém okolí o poloměru r část tohoto koncentrovaného seskupení u Aljašky, s sebou přinese nějaké zvětšení kótami převážené K -funkce oproti situaci s původními daty. Je sice pár míst okolo Austrálie, kde jsou menší shluky bodů s většími kótami, avšak velikosti kót nejsou tak drasticky velké (kóty

jsou shora omezené 6), aby sčítance testové statistiky, ve kterých nastává součin pouze malých hodnot kót, byly zanedbatelné (kóty jsou zdola omezeny 2,5). Je to tedy velice zajímavý případ, kdy dosahujeme v původních datech menších hodnot testové statistiky než u permutovaných dat, aniž by muselo původně docházet k záporné korelaci kót pro sousední body. Upozorníme, že celá situace i následné testování při výběru jiných dat by se mohly změnit. A to nejen volbou sledovaných bodů podle velikosti kót (například jiným nastavením spodní hranice), ale zejména nastavením časového období, ze kterého pochází sledované body. Stačilo by mít období, ve kterém by se odehrál větší počet silnějších zemětřesení v oblasti Aljašky, a celá situace by se mohla změnit.

Závěr

V práci jsme zavedli bodové procesy v euklidovském prostoru, seznámili jsme se s reálnými situacemi, které lze modelovat pomocí bodových procesů a položili jsme teoretický základ pro odvození K -funkce, která nám pomáhá analyzovat bodové procesy.

V druhé kapitole jsme rozšířili užitečnost bodových procesů přidáním kót, které nesou o bodech netriviální informaci, která nás zajímá. Zdefinovali jsme tedy kótované bodové procesy a zkoumali několik základních vlastností. Celou kapitolu jsme završili modifikováním K -funkce, aby pracovala s kótami. Následně jsme ukázali že kótami převážená K -funkce se za předpokladu i.i.d. kót zredukuje na původní K -funkci.

Ve třetí kapitole jsme nejprve uvedli základní vlastnosti sféry od popisování polohy na sféře, přes sférickou vzdálenost a okolí bodů v podobě kulových vrcholků, až po rotace na sféře. Vzápětí jsme zavedli bodové procesy na sféře, objasnili jsme pojem izotropní bodový proces a věnovali jsme se podobě K -funkce pro bodové procesy na sféře.

Ve čtvrté kapitole jsme zavedli kótované bodové procesy na sféře, probrali jsme několik základních situací a vlastností, se kterými se můžeme u kótovaných bodových procesů setkat a zdefinovali jsme kótami převáženou K -funkci pro bodové procesy na sféře a uvedli jsme její odhad, který jsme využívali jako testovou statistiku pro testování hypotézy nezávislosti kót.

V celé práci jsme dbali na vybudování intuice a seznámení daného tématu i čtenářům, kteří s bodovými procesy nemají zkušenosti. Za tímto účelem se zde nachází často velice podrobné odstavce a více příkladů i s ilustracemi.

Vlastní přínos práce najdeme hlavně ve čtvrté kapitole s názvem Kótované bodové procesy na sféře, avšak bylo potřeba vybudovat teorii a základní intuici s dostatečným pochopením pojmů v základních případech, kterými jsme se postupně zabývali v prvních třech kapitolách. Tedy dalším přínosem je shrnutí nejpotřebnějších vlastností a poznatků jednotlivých druhů bodových procesů potřebných k položení základů pro čtvrtou kapitolu a celkové seznámení s tímto tématem.

Seznam použité literatury

- [1] RATAJ, J. (2006). *Bodové procesy*. Druhé opravené vydání. Karolinum, Praha.
- [2] ILLIAN, J., PENTTINEN, A., STOYAN, H. a STOYAN, D. (2008). *Statistical analysis and modelling of spatial point patterns*. Chichester, West Sussex, England.
- [3] BADDELEY, A. (2007). *Spatial Point Processes and their Applications*. Springer-Verlag Berlin Heidelberg.
- [4] LAWRENCE, T. J. (2017). *Point Pattern Analysis on a Sphere*. PhD thesis. The University of Western Australia.
- [5] KOKS, D. (2006). *A Roundabout Route to Geometric Algebra*, strany 147–184. Springer Science+Business Media.
- [6] NATIONAL CENTERS FOR ENVIRONMENTAL INFORMATION (2006). *Earthquake intensity database 1638 - 1985*. URL https://www.ngdc.noaa.gov/hazard/intintro.shtml?fbclid=IwAR0xcq9jvnaAcOINYLhqwRL6wcP62GdGYwYUYKXr_mMKIH8JuWpS4wno0ig. Navštíveno 21.4.2022.
- [7] MYLLYMÄKI, M., MRKVIČKA, T., GRABARNIK, P., SEIJO, H. a HAHN, U. (2017). *Global envelope tests for spatial processes*. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 79, 381-404.
- [8] U.S. GEOLOGICAL SURVEY (2017). *Search earthquake catalog*. URL <https://earthquake.usgs.gov/earthquakes/search/>. Navštíveno 8.6.2022.

Seznam obrázků

1.1	Poloha sekvojí obrovských v kalifornském parku nám poukazuje na shlukování, tedy vytváření míst s větší hustotou bodů než v jiných částech obrázku.	4
1.2	Poloha středů buněk pozorovaných pod mikroskopem.	6
1.3	Místa usazení lidí v Gordon Square.	7
1.4	Nestacionární bodový proces - funkce intenzity roste směrem doprava.	8
1.5	Okrajový efekt a mínusová korekce.	10
2.1	Pouliční zločiny ve městě Chicago.	12
2.2	Rozmístění stromů v lese Lansing.	14
2.3	Rozmístění a velikosti stromů v lese Waka.	15
2.4	Rozmístění stromů v lese Lansing podle druhů.	16
2.5	Podélný rovinný řez částicemi při filtrování.	17
3.1	Zobrazení sféry a kulového vrchlíku - množiny bodů na sféře mající sférickou vzdálenost od pevně daného bodu sféry nejvýše $r \in [0, \pi\rho]$, kde ρ je poloměr sféry.	21
3.2	Bodový proces na sféře. I když jsou body rozmístěny rovnoměrně, při kolmém průmětu do roviny vidíme „zhuštění“ na okrajích kruhu.	22
3.3	Izotropní bodový proces.	24
3.4	Neizotropní bodový proces.	25
3.5	Mínusová korekce na sféře.	26
4.1	Polohy dopadů meteoritů jsou zaznamenány pomocí polohy středu kroužku a velikost meteoritu v okamžiku dopadu (tedy proporcionální velikost kóty) odpovídá poloměru kroužku.	28
4.2	Zemětřesení zaznamenaná USA v roce 1985.	30
4.3	Body nemají stejně rozdělené kóty - velikost závisí na poloze.	31
4.4	Červené body jsou nerovnoměrně rozloženy po sféře, modré body jsou rovnoměrně rozdělené na sféře.	32
4.5	Kótovaný bodový proces s vícerozměrnou kótou.	33
4.6	Odhadování kótami převážené K -funkce pro Poissonovy bodové procesy s parametrem λ a nezávislými kótami. Červeně je vyznačená teoretická hodnota nekótované verze K -funkce.	37
4.7	Polohy a magnituda zemětřesení zaznamenaná za období 1. až 7. dubna 2022.	38
4.8	Obálkový test - nulovou hypotézu zamítáme.	39
4.9	Obálkový test při odečtení nekótované K -funkce.	40

A. Přílohy

Zde uvádíme kód výpočtu kótami převážené K -funkce pro bodové procesy na sféře, pomocí kterého jsme testovali hypotézu o nezávislosti kót.

```
MWksphere = function (X, win = sphwin(), r = NULL, correction = c("un"),
ratio = FALSE, lambdavalues = NULL, update = TRUE)
{
stopifnot(inherits(X, c("sp2", "sp3", "matrix")))
if (!inherits(X, "matrix")) {
win <- X$win
mX <- X$X$marks
X <- X$X[1:2]
}
stopifnot(inherits(win, "sphwin"))
if (is.null(r)) {
rmax <- rmax.rule.sphwin(win)
r <- seq(0, rmax, length = 512)
}
correction.given <- !missing(correction)
correction <- match.arg(correction, c("un"), several.ok = TRUE)
if (!is.null(lambdavalues) && inherits(lambdavalues, "sphppm")) {
model <- lambdavalues
if (update) {
XX <- sp2(X, win)
model <- update(model, XX)
}
lambdavalues <- fitted(model)
}
rad <- win$rad
nrX <- nrow(X)
lr <- length(r)
areaX <- area.sphwin(win)
if (is.null(lambdavalues)) {
lambda <- nrX/areaX
lambda2 <- nrX * (nrX - 1)/areaX^2
denom <- lambda2 * areaX
}
else {
denom <- areaX
}
W <- unname(X)
Wm = data.matrix(W)
Dmat <- pairdistsph(Wm)
Kdf <- data.frame(r = r, theo = 2 * pi * rad * (1 - cos(r/rad)))
desc <- c("sférická vzdálenost r", "theoretical Poisson %s")
K <- ratfv(Kdf, NULL, denom, "r", quote(K(r)), "theo",
. ~ r, range(r), c("r", "%s[pois](r)"), desc,
fname = "K", ratio = ratio)

if (any(correction == "un")) {
mdij <- matrix(rep(0, nrX*lr), nrow = nrX)
for(i in 1:nrX){
for(j in 1:lr){
for(l in 1:nrX){
if(Dmat[i,l] <= r[j]){
if(i != l){
```

```

                                mdij[i,j] <- mdij[i,j] + mX[i] * mX[l]
                                }
                                }
                                }
}
Em <- mean(mX)^2

Kraw <- data.frame(r = r, un = colSums(mdij) * areaX / (nrX * (nrX - 1)
* Em))

Kraw <- fv(Kraw, "r", quote(hat(K(r))),
"un", NULL, range(r), c("r", "%s[odhad](r)"),
desc = c("sférická vzdálenost r", "odhad kótami převážené K-funkce"),
fname = "K")
K <- bind.fv(K, Kraw)
fvnames(K, ".y") <- "un"
}
return(K)
}

```