

UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA

DIPLOMOVÁ PRÁCE

Odhady kuželoseček a kvadrik



Katedra matematické analýzy a aplikací matematiky
Vedoucí diplomové práce: **doc. RNDr. Eva Fišerová Ph.D.**
Vypracoval(a): **Martin Schejbal**
Studijní program: Aplikovaná matematika
Studijní obor Matematika a její aplikace
Forma studia: Prezenční
Rok odevzdání: 2022

BIBLIOGRAFICKÁ IDENTIFIKACE

Autor: Martin Schejbal

Název práce: Odhady kuželoseček a kvadrik

Typ práce: Diplomová práce

Pracoviště: Katedra matematické analýzy a aplikací matematiky

Vedoucí práce: doc. RNDr. Eva Fišerová Ph.D.

Rok obhajoby práce: 2022

Abstrakt: Diplomová práce hovoří o metodách vhodných pro statistické modelování kuželoseček a kvadrik. Teoretická část práce je zaměřena na matematickou podstatu tohoto typu modelování. Praktická část práce obsahuje řešení ilustrativních úloh a porovnání studovaných metod zejména z hlediska jejich přesnosti a výpočetní rychlosti. Úlohy jsou zpracovány v softwaru R a doplněny o obrázky z provedených simulací.

Klíčová slova: Regrese, Kružnice, Geometrická aproximace, Algebraická aproximace

Počet stran: 62

Počet příloh: 1

Jazyk: český

BIBLIOGRAPHICAL IDENTIFICATION

Author: Martin Schejbal

Title: Fitting conics and quadrics

Type of thesis: Master's

Department: Department of Mathematical Analysis and Application of Mathematics

Supervisor: doc. RNDr. Eva Fišerová Ph.D.

The year of presentation: 2022

Abstract: This master thesis is discussing several methods for fitting conics and quadrics. Theoretical part is about the mathematical background and derivation of said methods. Practical part is filled with simulations and its results. In the end our selected methods are compared in many ways.

Key words: Regression, Circle, Geometrical fit, Algebraic fit

Number of pages: 62

Number of appendices: 1

Language: Czech

Prohlášení

Prohlašuji, že jsem diplomovou práci zpracoval samostatně pod vedením paní doc. RNDr. Evy Fišerové Ph.D. a všechny použité zdroje jsem uvedl v seznamu literatury.

V Olomouci dne

.....

podpis

Poděkování

Na tomto místě bych chtěl poděkovat především své vedoucí diplomové práce paní doc. RNDr. Evě Fišerové Ph.D., za nekonečně velkou trpělivost, aby mi pomohla dovést tuto práci ke zdárnému konci. Také bych chtěl poděkovat kolegyni Mgr. Viktorii Nesrstové za konzultace a pomoc, které si velmi vážím.

Obsah

1	Úvod	7
1.1	Motivace	7
1.2	Historie	8
1.3	Praktické využití	9
2	Regresní analýza	11
2.1	Metoda nejmenších čtverců	12
2.2	Modely s chybami v proměnných	15
2.3	Parametrizace	17
2.4	Geometrické aproximace	19
2.5	Obecné řešení pro modely s chybami v proměnných	21
2.6	Vlastnosti regresních odhadů	22
2.7	Nelineární modely	23
3	Odhady kuželoseček a kvadrik	25
3.1	Parametrizace	25
3.2	Teorie	28
3.3	Typy aproximace kružnicí	29
3.3.1	Landauova metoda	30
3.3.2	Kasova metoda	31
3.3.3	Prattova metoda	33
3.3.4	Taubinova metoda	34
3.4	Kombinace metod	34
3.5	Další rozšíření	35
4	Praktické úlohy	37
4.1	Odhadování kružnice	39
4.2	Odhadování půlkružnice	46
4.3	Odhadování čtvrtkružnice	50
4.4	Speciální úlohy	53
4.5	Souhrn	58
5	Závěr	60

1 Úvod

Tato diplomová práce se zabývá speciálním případem regrese. Tento případ se týká prokládání pozorovaných dat kuželosečkami a kvadrikami. Jedná se o populární téma v moderní statistice, které přitahuje stále více pracovníků ze všech různých oborů. Práce je rozdělena do čtyř hlavních kapitol.

Nejprve si řekneme, jestli a proč je vlastně dobré věnovat se tomuto tématu. Pak se ohlédneme do historie a dozvíme se, kdo s touto teorií přišel jako první a na jaké problémy narazil. V průběhu celé práce pak ukážeme, jak tyto problémy řešit a které otázky před námi stále leží nevyřešeny. Na závěr první části vyjmenujeme několik konkrétních příkladů, kde se dá v praxi tato teorie využít a jaké výsledky můžeme získat. Nejprve uvedeme pár spíše historických příkladů a později i problémy dnešní doby, které se řeší právě pomocí teorie odhadu kuželoseček a kvadrik.

Ve druhé části se budeme věnovat zejména teorii obecné regresní analýzy. Připomeneme, co regrese obecně je a jak ji používáme. Rovněž v krátkosti pohovoříme o chybách a vlastnostech odhadů. Také představíme pár teoretických problémů a podíváme se, jak s nimi pracovat.

Poté se dostaneme do hlavní části práce, a sice k samotným odhadům kuželoseček. Představíme teoretický přístup k této problematice a různé metody, které budeme používat. Jedná se o jednu metodu geometrické aproximace a tři metody aproximace algebraické. Také se dozvíme jakou roli hraje parametrizace jednotlivých geometrických objektů. Použité metody rozebereme z pohledu teorie, ať jsme dobře připraveni na poslední kapitolu.

V poslední kapitole budeme v softwaru R řešit konkrétní problémy. Budeme používat metody zmíněné v předchozích částech práce. Na závěr budeme srovnávat jejich výsledky podle výpočetní rychlosti, přesnosti, stability a složitosti postupu. Cíl naší práce bude vypracován zejména v této kapitole. Chceme ukázat, kde má jaká metoda své silné stránky, a naopak v jakých případech je lepší sáhnout po metodě jiné. Výsledky budou doplněny obrázky ze softwaru R . Data, která použijeme, budeme sami generovat. Kód použitý ve výpočtech bude k dispozici jako příloha celé diplomové práce.

1.1 Motivace

Naše motivace ke studiu tohoto tématu je velmi podobná motivaci pro zkoumání klasické regrese. Téměř v každém oboru, kde analyzujeme nějaké úkazy nebo experimenty, hojně využíváme (nebo bychom určitě měli) dat. Výzkum bez dat lze jen těžko označit za opravdový výzkum. Jedná se pak pouze o teorii či spekulaci, kterou nelze ani potvrdit, ani vyvrátit. Proto vznikla potřeba naučit se s daty zacházet a různě vhodně je ohýbat, abychom mohli úspěšně dospět k nějakému validnímu závěru, zejména pomocí statistických testů. Data se ale ne vždy chovají tak, jak bychom si představovali, vlastně jen velmi zřídka kdy. Proto jsme

se v základním kurzu matematické analýzy učili například Taylorův polynom, a pak v kurzu pravděpodobnosti aproximovat data pomocí polynomů a dalších elementárních funkcí, které umíme dobře vyjádřit a použít bez významné ztráty informace. Ve statistice děláme v podstatě to samé. Snažíme se naše, často velmi divoká, data popsat pomocí nějakého „pěkného“ modelu. Pěkného ve smyslu, že se nám hodí do statistických testů, ale zároveň nenarušíme trend našich dat a neztratíme žádnou důležitou informaci uchovanou v datech. O tom hovoří celé jedno statistické téma s názvem regrese. Nejzákladnější ze všech aproximací je prokládání dat přímkou. Jedná se o typ lineární regrese, ta se ale bohužel nedá použít v mnoha případech, respektive lze ji použít, ale její přesnost nemusí být vždy dostačující. Výsledky tudíž budou jen málo použitelné.

Jak bylo řečeno v úvodu, my se budeme věnovat zejména jednomu konkrétnímu typu regrese. V případech, kdy je lineární regrese nedostačující, musíme sáhnout po složitějších modelech. Jeden z nich budeme diskutovat právě v této práci. Budeme data prokládat kuželosečkami, zejména pak kružnicí.

V tomto tématu narážíme ještě na jednu závažnou komplikaci. V klasické regresi se bavíme o jedné náhodné proměnné a druhé vysvětlující, jedná se tedy o vztah $y = f(x)$. Zde budou obě naše proměnné, vysvětlovaná i vysvětlující, ovlivněny náhodnou chybou. Takže se jedná o výrazně komplexnější problém, než je klasická regrese. V [15] je podrobněji uveden příklad problému, kde měříme objem a hmotnost kusů železa, pomocí těchto dat se snažíme odhadnout parametr hustoty ρ z notoricky známého vztahu $y = \rho x$, (resp. $m = \rho V$). Jaký odhad budeme chtít použít pro parametr ρ , není příliš důležité, a hlavně to není předmětem této práce, proto budeme nadále hovořit jen obecně o odhadu parametru. Pro žádné i nemůžeme použít přesný vzorec $y_i = \rho x_i$, právě proto, že měření obou proměnných mohou být, a typicky budou, zatížena chybou. Jak řešit tento problém, se dozvíme později v podkapitole hovořící o modelech s chybami v proměnných.

1.2 Historie

Regrese jako taková se poprvé začala používat již na začátku 19. století. Jednalo se tehdy o použití metody nejmenších čtverců, která je stále velmi populární v mnoha odvětvích matematiky díky své jednoduchosti, výpočetní rychlosti a přesnosti. Poprvé o této metodě hovořil Adrien-Marie Legendre v roce 1805 a Carl Friedrich Gauss v roce 1809. Oba matematici ji použili pro popis orbit objektů kolem slunce. Legendre v publikaci *Nouvelles Méthodes pour la Détermination des Orbites des Comètes*¹ a Gauss v publikaci *Theoria motus corporum coelestium in sectionibus conicis solem ambientum*². V podstatě obecný regresní problém byl kompletně vyřešen na začátku 19. století. Teprve později se začaly objevovat složitější problémy vyžadující hlubší analýzu.

¹Nové metody určování drah komet.

²Teorie pohybu nebeských těles obíhajících po eliptických dráhách kolem slunce.

Francis Galton v knize *Natural Inheritance* z roku 1889 poprvé zmínil metodu *regrese k průměru*, když popisoval jev, při kterém se výška dětí vysokých rodičů postupně zmenšovala směrem k průměru populace. První zmínka je z roku 1889, ale ve skutečnosti tento typ regrese objevil již v roce 1886. Zjednodušeně tato metoda říká, že máme-li měření jedné proměnné, které je extrémní, pak s rostoucím počtem opakování tohoto jednoho měření se budeme blížit střední hodnotě této proměnné. Tedy hodnota měření bude méně extrémní při dalších opakováních.

Problém regrese, ve které jsou obě proměnné do jisté míry náhodné, se poprvé objevil v sedmdesátých letech devatenáctého století. Mezi rokem 1975 a 1995 bylo toto téma velmi populární a v devadesátých letech dvacátého století byly všechny hlavní otázky týkající se tohoto typu lineární regrese zodpovězeny.

První zmínka o moderním (matematicko-statistickém) přístupu k odhadům kružnic je z padesátých let dvacátého století a je přinejmenším zajímavá. Britští inženýři společně s archeology zkoumali megalitické stavby z různých míst, aby zjistili, zda lidé, kteří je postavili, používali společnou jednotku vzdálenosti.

Celé toto téma je stále ve vývoji a hledají se další možnosti, jak vylepšit známé metody, ať už z hlediska výpočetní rychlosti, nebo přesnosti. Vzhledem k tomu, že se veškeré větší úlohy v dnešní době řeší na počítačích, je určitě dost prostoru pro zlepšení. Obecně se téma algoritmizace bude jen velmi těžko považovat za uzavřené, ať už se jedná o jakékoliv výpočty.

1.3 Praktické využití

Jeden z hlavních důvodů, proč je naše téma tak populární, je určitě široké spektrum uplatnění této teorie. Kromě zkoumání společné míry vzdálenosti našich předků, mohou archeologové odhadovat velikosti keramických předmětů ze starověku pomocí analýzy střepů nalezených ve výkopech, více např. v publikacích [10, 12, 21].

Kromě zmíněných, spíše historických, aplikací se najde i velké množství modernějších např. [11, 13]. V medicíně můžeme odhadovat průměr lidské duhovky kvůli čočkám z fotografie např. [16] nebo odhadujeme rozměry zubního oblouku pomocí analýzy snímku z rentgenu např. [7]. Také můžeme odhadovat velikost plodu z fotografie z ultrazvuku. Vše uvedené spadá pod společné téma rekonstrukce fotografií právě pomocí prokládání dat kuželosečkami.

Hlavní výskyt této teorie je v analýze fotografií a různých jiných obrázků. V případě rentgenu chrupu se podíváme na naši fotografii, a snažíme se aproximovat zubní oblouk pomocí paraboly. Poté odhadneme parametry této paraboly a na základě těchto odhadů můžeme určit například rozměry pro rovnátka. Obdobně postupujeme u ostatních případech rekonstrukce fotografie či obrázků.

Velký prostor pro aplikaci regrese kuželosečkami nabízí také fyzika. Když se podíváme na nově vzniklé částice v urychlovači částic, zjistíme, že se pohybují po kružnicových obloucích v magnetickém poli. Výzkumníci tak mohou měřit

jejich energii pomocí poloměru jejich trajektorií, neboť částice s velikou energií se pohybují po oblouku s malým průměrem.

2 Regresní analýza

Jak jsme již zmínili v úvodu, budeme se snažit odhadovat nějaké náhodné veličiny pomocí nenáhodných, tzv. vysvětlujících veličin. Proto se naše náhodné veličiny také nazývají vysvětlované proměnné. Obecně téma regrese slouží ke studiu vzájemných vztahů mezi veličinami, a to zejména závislosti. Existuje mnoho různých technik pro vypracování modelů a rovněž i mnoho různých metod, jak tyto modely analyzovat, interpretovat a ověřovat jejich statistickou významnost.

Nyní připomeneme teoretický pohled na klasickou regresi, abychom později mohli porovnat naše hlavní téma s poznatky v této kapitole.

Celou myšlenku, popsanou v předchozím odstavci, formálně vyjádříme jako hledání podmíněné střední hodnoty vysvětlované veličiny. Podmíněná se nazývá proto, že ty vysvětlující veličiny známe. Řekněme tedy, že budeme chtít odhadnout proměnnou Y pomocí nenáhodné veličiny X . Hledáme pak podmíněnou střední hodnotu veličiny Y jako funkci vysvětlujících proměnných

$$E(Y|X) = f(X).$$

Obvykle vyjádříme tuto střední hodnotu v obecném tvaru, v němž vystupují regresní parametry, které budeme později odhadovat na základě pozorovaných hodnot. Tedy pracujeme s výrazem

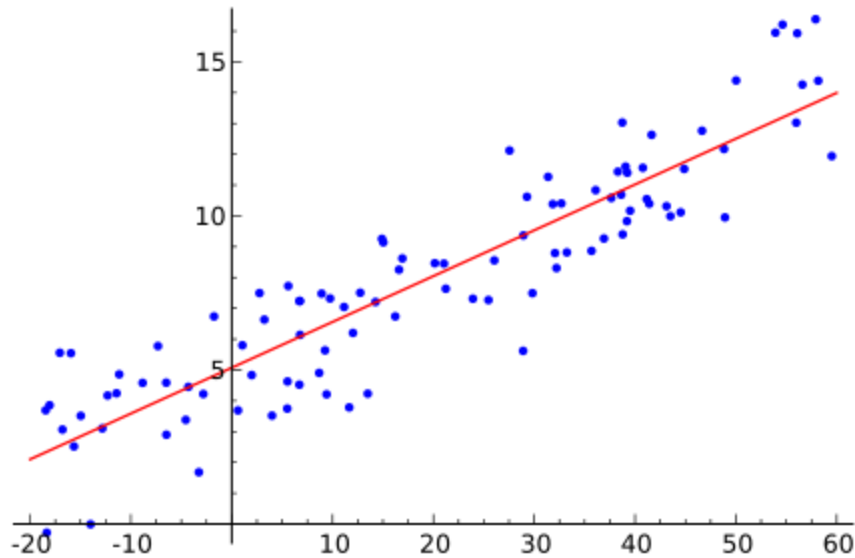
$$E(Y|X) = f(\beta_0, \beta_1, \dots, \beta_k).$$

Lineární regrese v obecném případě znamená odhad hodnot $[x_i, y_i]$ pomocí funkce $f(x, \beta_1, \beta_2, \dots, \beta_k)$, kterou lze vyjádřit jako lineární kombinací (odtud název lineární regrese) funkcí f_1, \dots, f_k tak, že platí $y = \beta_1 f_1 + \beta_2 f_2 + \dots + \beta_k f_k$. My se podíváme na případ aproximace dat přímkou. Je to nejjednodušší možná regrese a často velmi účinná. Model regresní přímky vypadá takto

$$Y = \beta_0 + \beta_1 X + \epsilon, \tag{1}$$

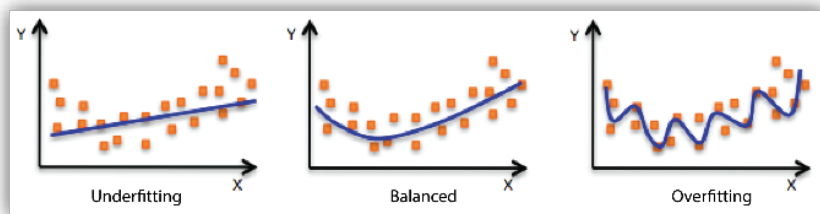
kde β_0, β_1 jsou neznámé koeficienty zvané regresní parametry a ϵ náhodné chyby. Typicky budeme předpokládat, že platí $\epsilon \sim N(0, \sigma^2)$ a $\text{cor}(\epsilon_i, \epsilon_j) = 0, \forall i \neq j$. Máme-li k dispozici odhady regresních parametrů, získáme vyrovnané (odhadnuté) hodnoty Y

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X. \tag{2}$$



Obrázek 1: Lineární regrese

Na obrázku 1 vidíme, že v případě použití těchto dat nám stačí lineární regrese, a není potřeba vymýšlet a hledat komplikovanější polynomy nebo jiné složitější funkce, abychom měli přesnější aproximaci. A navíc se určitě chceme vyhnout jevu zvanému overfitting. Zjednodušeně jde o použití příliš složitých metod na jednoduchý příklad. Obdobně se chceme vyhnout i opačnému případu, tedy underfittingu, kdy například proložíme veškeré datasety přímkou (viz obrázek 2). Obrázek je převzatý z [25].



Obrázek 2: Underfitting a Overfitting

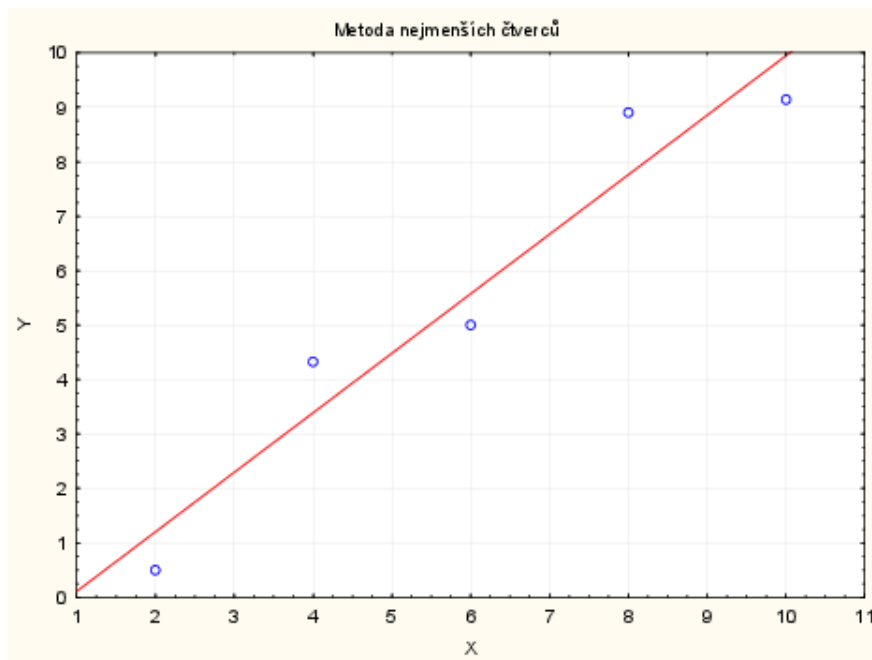
Jak ale zjistit, kterou přímkou nebo obecně regresní funkci zvolit? V následující kapitole připomeneme v krátkosti metodu nejmenších čtverců, která - jak již bylo řečeno - byla první představenou metodou regrese již na začátku 19. století.

2.1 Metoda nejmenších čtverců

Tato matematicko-statistická metoda je velmi často užívaná i mimo naše téma regrese, původně se touto metodou řešily soustavy rovnic, respektive se hledala

aproximace jejich řešení. Obecně tato metoda slouží hlavně k nalezení nejlepší aproximace z těch, které dostaneme pomocí eliminace chyby nebo přesněji její minimalizace. Takže využití této teorie je velmi bohaté, v podstatě se dá použít, kdekoli se setkáváme s nepřesnými daty a potřebujeme používat nějaké naměřené výsledky. Můžeme například pomocí metody nejmenších čtverců odhadnout pravděpodobný průsečík dvou přímek v rovině, jejichž rovnice není přesná, ale obsahuje opět nějakou chybu.

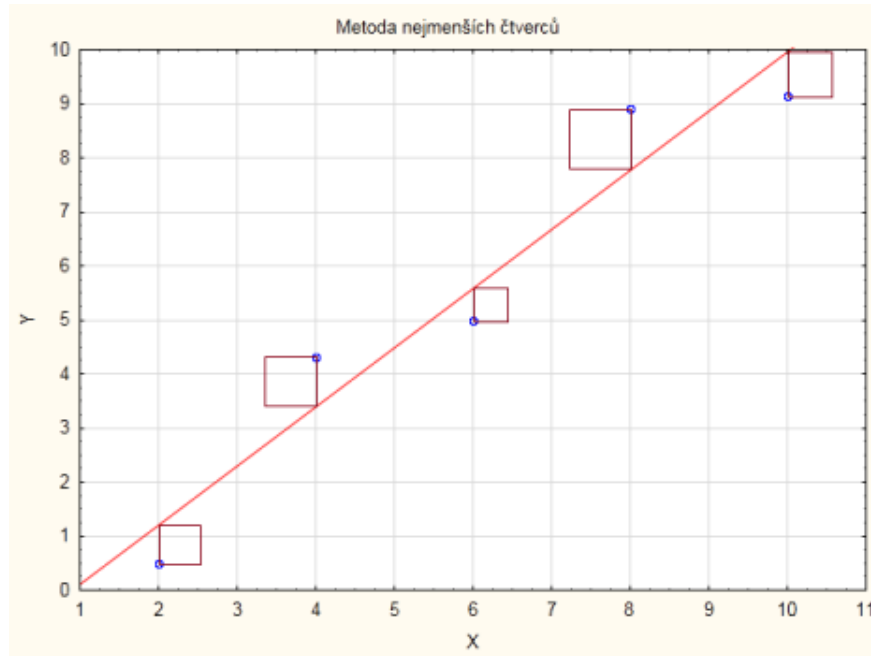
Pro jednoduchost začneme přímkou. Našimi daty můžeme proložit hned několik přímek, které si budou dost podobné a zdánlivě budou všechny odhadovat naměřená data dobře. Nám samozřejmě dobře odhadnutá data nestačí, a tak hledáme tu přímku, která naše data odhadne nejlépe. Podobně se v základním kurzu pravděpodobnosti bavíme o nejlepší nestranném odhadu náhodné veličiny, kde hledáme mezi všemi nestrannými odhady ten nejlepší. Zde máme všechny přímky – všechny lineární regrese – a budeme hledat tu nejlepší. Nejlepší ve smyslu nejmenší odchylky naší aproximace od skutečných naměřených dat. Jedna z prvních metod, která byla použita právě na tento problém, se nazývá metoda nejmenších čtverců.



Obrázek 3: Metoda nejmenších čtverců I

Na obrázku 3 máme regresní přímku červeně a konkrétní naměřená data jako modrá kolečka. Na obrázku 4 už je vidět, jak budeme zavádět metodu nejmenších čtverců.

Některé literatury označují tuto metodu jako minimalizaci součtu čtverců vertikálních vzdáleností, tedy vertikálních vzdáleností naměřené hodnoty od svého



Obrázek 4: Metoda nejmenších čtverců II

odhadu. Kromě vertikálních vzdáleností se dají počítat i jiné vzdálenosti, a vznikají tak modifikované metody nejmenších čtverců, ale k tomu se vrátíme později. V podstatě se budeme snažit najít takovou přímku, aby obsah vzniklých čtverců byl co nejmenší, proto i tento název. Čili budeme minimalizovat výraz

$$S_E = \sum_{i=1}^n e_i^2,$$

kde e_i jsou rezidua, tedy rozdíl mezi skutečnou naměřenou hodnotou y_i a naším regresním odhadem \hat{y}_i . Tedy e_i^2 je obsah i -tého čtverce z obrázku 4. Jestliže budeme odhadovat teoretickou hodnotu Y_i dostaneme \hat{Y}_i , dohromady se vztahem (1) tedy máme

$$S_E = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2.$$

Přičemž platí

$$Y_i = \beta_0 + \beta_1 X_i, \quad \hat{Y}_i = b_0 + b_1 X_i,$$

kde $b_0 = \hat{\beta}_0$ a $b_1 = \hat{\beta}_1$ jsou konkrétní odhady příslušných parametrů. Celkem pak dostáváme následující rovnost

$$S_E = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2.$$

Tento výraz budeme minimalizovat vzhledem k hodnotám regresních parametrů β_0, β_1 . Po několika málo krocích dostáváme následující výsledky pro oba naše parametry

$$\begin{aligned}\hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X}, \\ \hat{\beta}_1 &= \frac{\sum_{i=1}^n x_i y_i - \bar{y} \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i},\end{aligned}$$

kde $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ a $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$.

Pro metodu nejmenších čtverců platí následující zásadní věta.

Věta 2.1 *Odhady získané metodou nejmenších čtverců jsou nejlepší nestranné lineární odhady.*

Poznámka 2.2 *Navíc platí, že tyto odhady jsou silně konzistentní, tedy platí $\hat{\beta}_0 \rightarrow \beta_0$, (resp. $\hat{\beta}_1 \rightarrow \beta_1$) pro $n \rightarrow \infty$ s pravděpodobností 1.*

Vyjádření tohoto odhadu v maticovém tvaru by pak vypadalo následovně

$$\mathbb{Y} = \mathbb{X}\hat{\beta} + \epsilon,$$

kde navíc pro vektor $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$ platí $E(\hat{\beta}) = \beta$ a $\text{var}(\hat{\beta}) = \sigma^2(\mathbb{X}'\mathbb{X})^{-1}$, kde \mathbb{X} je matice prediktorů.

Jedna z velmi důležitých vlastností jakéhokoli odhadu je jeho nejistota. Ta se skládá z nejistoty měření (chyba měřícího zařízení) a náhodných chyb měření. Tato nejistota vymezuje interval, o kterém předpokládáme, že do něj měření padne s určitou pravděpodobností. Nejistota měření výsledných odhadů je typicky charakterizována varianční maticí.

2.2 Modely s chybami v proměnných

Jak již bylo zmíněno, v klasické regresní analýze máme typicky k dispozici data, která jsou zatížena chybou a chceme je odhadnout nenáhodnou proměnnou čili proměnnou nenesoucí chybu. V hlavním tématu této práce se budeme zabývat modely, kde vystupují chyby v obou proměnných, tedy ve vysvětlující i vysvětlované. Představíme si hned několik možností, jak k tomuto problému přistupovat.

V kapitole 1.1 byl nastíněn příklad s výpočtem hustoty železa pomocí naměřených objemů a hmotností. Obě tyto hodnoty mohou být, a pravděpodobně budou, zatíženy chybou. Může se stát, že kus železa, který měříme, bude obsahovat nějaké jiné prvky, a proto jeho hmotnost bude větší (resp. menší), než by měla být, kdyby se jednalo o homogenní látku. Nebo jde zkrátka o chybovost

našeho měřicího zařízení. Příčin takového problému je mnoho, zmiňujeme zde jen dvě, aby bylo vidět, jak velmi reálný a častý tento problém může být.

Další příklady modelů s chybami v proměnných můžeme najít v první kapitole v [8] (jako příklad můžeme uvést i výzkum týkající se přeživších atomové bomby), anebo v [20], kde najdete některé ekonomické modely.

První možnost, jak s takovým problémem naložit, je velmi intuitivní. Řekneme jednoduše, že jedno z našich měření (v našem případě hmotnost nebo objem) je správně, budeme tedy ignorovat míru chyby. Kterou z našich proměnných ale vybrat? Záleží na tom? Dostaneme jiný výsledek pro jinou volbu?

Zde se dostáváme k hlavnímu problému tohoto přístupu. Budeme-li považovat hodnoty x_i za přesné, dostaneme vztah $y = bx$ a položíme $\rho_1 = b$. V opačném případě, tedy když označíme hodnoty y_i za přesné, dostaneme vztah $x = b'y$ a položíme $\rho_2 = 1/b'$.

Přirozeně dostaneme pro ρ jiné hodnoty. Dostáváme tak dvě různé regresní funkce, a to určitě není správně. Kterou funkci tedy označit za řešení? Lze pomocí těchto funkcí dostat správné řešení? Například nějakou formou průměru?

Již v roce 1901 řekl K. Pearson, že tyto funkce jsou jiné při jiné volbě proměnné, kterou označíme za „přesnou“. Později bylo dokázáno, že oba tyto odhady ρ_1, ρ_2 jsou nekonzistentní a nemusí být vůbec nestranné. Dokonce již dnes víme, že odhad ρ_1 systematicky odhaduje tu správnou regresní funkci zdola a odhad ρ_2 shora. Vidíme, že tento přístup není ideální.

Další možností je přijmout fakt, že jsou naše naměřené hodnoty zatíženy chybou, a počítat s nimi tak. Potom určitě víme, že existují nějaké hodnoty \tilde{x}_i a \tilde{y}_i , které nabývají teoretické správné hodnoty. Tyto hodnoty na sobě závisí pomocí neznámé funkce g , tedy $\tilde{y}_i = g(\tilde{x}_i)$, $\forall i$.

Potom můžeme psát

$$x_i = \tilde{x}_i + \delta_i \quad y_i = \tilde{y}_i + \epsilon_i, \quad i = 1, \dots, n. \quad (3)$$

To znamená, že každé naše měření x_i , respektive y_i , se liší od té pravé hodnoty o δ_i , respektive ϵ_i . Předpokládáme, že $\delta_1, \dots, \delta_n, \epsilon_1, \dots, \epsilon_n$ jsou nezávislé náhodné veličiny se střední hodnotou rovnou nule.

V nejjednodušším případě můžeme předpokládat, že δ_i mají pro každé i stejný rozptyl σ_x a ϵ_i mají pro každé i stejný rozptyl σ_y . Dokonce je běžné předpokládat normalitu pro obě proměnné, tedy dostaneme předpoklady ve tvaru

$$\delta_i \sim N(0, \sigma_x^2), \quad \epsilon_i \sim N(0, \sigma_y^2). \quad (4)$$

Máme v podstatě dvě možnosti, jak naložit s těmi pravými hodnotami. První z variant se nazývá funkcionální model a jeho podstatou je považovat ty pravé hodnoty za pevně dané, a tedy nenáhodné. V tomto případě s nimi budeme zacházet jako s dalšími parametry.

Druhá možnost je považovat tyto pravé hodnoty za realizace nějakých skrytých proměnných s vlastním rozdělením pravděpodobnosti. Obvykle se uvádí

předpoklad, že $\tilde{x}_i \sim N(\mu, \sigma^2)$, a pak dostaneme \tilde{y}_i ze vztahu $\tilde{y}_i = g(\tilde{x}_i)$. Parametry normálního rozdělení u proměnné \tilde{x}_i odhadneme společně s parametry funkce $g(x)$. Tento model se nazývá strukturální.

2.3 Parametrizace

Když mluvíme o lineární regresi, je klíčové správným způsobem vyjádřit obecnou přímkou. Správně ve smyslu, že by naše vyjádření mělo popisovat každou přímkou v rovině a v ideálním případě jednoznačně. Až do osmdesátých let dvacátého století se používala velmi známá rovnice

$$y = a + bx.$$

Z jasných důvodů: je to jednoduchá lineární rovnice, umíme dobře interpretovat oba parametry (směrnice a průsečík s osou Y) a popisuje v podstatě každou přímkou v rovině. Tuto rovnici označujeme za tzv. směrnicový tvar přímky. Bohužel ale narazíme na problém, uvažujeme-li „skoro vertikální“ přímky. Dostáváme pro parametr b libovolně velké hodnoty, což pochopitelně nadělá spoustu problémů v numerických řešeních. Dříve by to nebyl takový problém, ale nyní, kdy se velká část úloh řeší numericky, už to problém je. Navíc odhady parametrů a, b mají nekonečné střední hodnoty a nekonečné rozptyly, což se nedá moc dobře interpretovat. Tento přístup tedy není optimální hned z několika závažných důvodů.

Druhá varianta, jak popsat přímkou v rovině, je pomocí tří parametrů najednou. Tedy obecnou rovnicí přímky

$$Ax + By + C = 0, \tag{5}$$

kde $A, B, C \in \mathbb{R}$. Jedná se o algebraický tvar přímky v rovině. Nabízí se podmínky $A \neq 0$ a $B \neq 0$, které zapíšeme dohromady vztahem

$$A^2 + B^2 > 0. \tag{6}$$

V případě, že by oba parametry byly rovny nule, dostáváme rovnici $C = 0$, a tedy buď celou rovinu (případ $C = 0$) nebo prázdnou množinu (případ $C \neq 0$). Ani jedna množina řešení není žádoucí, a proto tedy zavádíme takovou podmínku, která nám navíc zaručí, že momenty odhadovaných parametrů budou konečné. Dále pomocí této parametrizace zaručíme i numerickou stabilitu, ale k tomu více později.

V této parametrizaci nastává jiný problém. Použijeme-li vektory parametrů (a, b, c) a $(\alpha a, \alpha b, \alpha c)$ pro $\alpha \in \mathbb{R}$, dostáváme tutéž přímkou. To opět zakážeme pomocí podmínky

$$A^2 + B^2 = 1, \tag{7}$$

která nám zaručí jednoznačnost určení přímky. Z podmínky (7) snadno dostáváme podmínku (6), takže nám stačí pouze ta druhá v pořadí.

Nyní můžeme zavést parametrický prostor, na kterém budeme později řešit existenci a jednoznačnost řešení. Nahradíme současně parametry A, B jedním parametrem φ následovně

$$A = \cos \varphi, \quad B = \sin \varphi, \quad \text{kde } \varphi \in [0, 2\pi).$$

Po této substituci můžeme libovolnou přímku vyjádřit pomocí dvou parametrů, jako

$$x \cos \varphi + y \sin \varphi + C = 0.$$

Vzhledem k tomu, odkud bereme jednotlivé parametry, se celý náš prostor (zvaný parametrický prostor) stává nekonečným válcem. Parametr C udává „výšku“, a parametr φ udává směr normálového vektoru přímky. Takže se zřejmě opět nejedná o jednoznačné určení přímky. Každou přímku můžeme určit dvěma způsoby. Máme-li dvojici vektorů parametrů (φ, C) a $(\varphi + \pi, -C)$, jedná se o tutéž přímku. Jednoznačnost zaručíme přidáním jednoduché podmínky $C \geq 0$. Nyní již nemáme nekonečný válec, ale jen jeho „polovinu“. Dostáváme jednoznačné určení přímky až na jednu výjimku. Tou jsou přímky, které obdržíme při použití vektorů parametrů $(\varphi, 0)$ a $(\varphi + \pi, 0)$. Vidíme, že se jedná pouze o případ, kdy $C = 0$, a ten nenastává příliš často, uvádíme ho zde pouze pro úplnost. V této parametrizaci budeme chtít minimalizovat funkci

$$\Phi(\varphi, C) = \sum_{i=1}^n (x_i \cos \varphi + y_i \sin \varphi + C)^2. \quad (8)$$

Jak je to s existencí řešení? Funkce Φ ze vzorce (8) je zřejmě spojitá na našem parametrickém prostoru. Víme, že spojitě funkce na kompaktním prostoru nabývají svého minima. Jelikož ale C nabývá libovolně velkých hodnot, nemáme náš parametrický prostor kompaktní, takže to tak snadné nebude. Zde přeskočíme poměrně obsáhlý a technicky náročný důkaz o existenci minima funkce ze vztahu (8). Celý je dobře rozepsaný v podkapitole 2.1 v [3]. Pro tuto práci bude dostatečující pouze závěr, který říká, že funkce Φ ze vzorce (8) nabývá svého minima.

Zde si dovolím krátkou odbočku od hlavního tématu. V případě parametrizace (5) nastává jedna zajímavá specialita. Odhadujeme zde tři parametry na jednu (A, B, C) . Předpokládáme, že existuje skutečná přímka, kterou neznáme, ale víme, že bude mít předpis

$$\tilde{A}x + \tilde{B}y + \tilde{C} = 0,$$

kde $\tilde{A}^2 + \tilde{B}^2 = 1$. Dále víme, že existuje n „pravých“ bodů $(\tilde{x}_1, \tilde{y}_1), \dots, (\tilde{x}_n, \tilde{y}_n)$ ležících na této přímce. Jestliže tyto „pravé“ body leží na této „pravé“ přímce, znamená to, že platí následující rovnost

$$\tilde{A}\tilde{x}_i + \tilde{B}\tilde{y}_i + \tilde{C} = 0, \quad i = 1, \dots, n.$$

Každý z pozorovaných bodů (x_i, y_i) je náhodně posazen nedaleko tohoto „pravého“ bodu, tedy platí

$$x_i = \tilde{x}_i + \delta_i, \quad y_i = \tilde{y}_i + \epsilon_i,$$

kde δ_i, ϵ_i jsou nezávislé náhodné veličiny s normálním rozdělením $N(0, \delta^2)$. Pravé hodnoty \tilde{x}_i, \tilde{y}_i neznáme, ale jsou pevné. Tedy s nimi musíme zacházet jako s dalšími parametry. Jedná se o podobnou situaci, jakou jsme diskutovali v předchozí kapitole. Problém nastane, když se budeme chtít podívat, jak se chová výsledný odhad pro rostoucí n . Ve statistice je obvykle zajímavé sledovat, jak se naše odhady chovají při rostoucím počtu pozorování (zejména rostoucím do nekonečna). S rostoucím n narážíme na komplikaci. Postupně přibývají ty „pravé“ body, a tedy uměle navyšujeme počet parametrů.

S tím se v klasické regresi nesetkáváme, například odhad pomocí metody maximální věrohodnosti vůbec nepočítá s rostoucím počtem parametrů, a tedy zcela ztrácí smysl. Tento přístup je vhodnější pro experimenty, kde máme opravdu velké množství dat, pro zlepšení přesnosti výsledků. Ovšem znamená to ve výsledku pro jednotlivá pozorování to, že se nemusí nacházet blízko našeho odhadu. Pro celek jsou však výsledky validní.

Další varianta, jak studovat asymptotické chování odhadů, je nechat pevné n a sledovat chování odhadů pro $\sigma \rightarrow 0$. Tento přístup je lepší, pokud nemáme příliš velké množství dat. Například u rekonstrukce fotografií máme pouze limitovaný počet bodů, které můžeme použít. A co se týče jednotlivých pozorování, tak je typicky budeme mít blízko našeho odhadu, na rozdíl od předchozího přístupu.

Třetí přístup, kterým můžeme nahlížet na tento problém, je jistý druh kompromisu. Budeme koukat na obě limity ($n \rightarrow \infty, \sigma \rightarrow 0$) současně. To je samo o sobě velmi silný předpoklad, bez kterého se v lineárních modelech obejdeme. Více o nelineárním případě se můžeme dočíst v [3, 22].

2.4 Geometrické aproximace

V oblasti analýzy obrázků nebo fotografií se dostáváme k disciplíně geometrické aproximace. U nepovedených (nepřesných) křivek v obrázcích je budeme chtít narovnat, tj. aproximovat přesnou křivkou, například najít body na obrysu této křivky a lineárně je spojit. Vznikne tak po částech lineární aproximace původní křivky. Další možnosti jsou u kulatých či oválných objektů na fotografiích, respektive v obrázcích, můžeme je aproximovat pomocí pravidelné elipsy nebo obecně kuželoseček.

Když se podíváme podrobněji na aproximaci přímkou, dostaneme podobný výsledek jako u lineární regrese. Zvolíme body na obrysu křivky, kterou chceme aproximovat. Tyto body budou mít souřadnice $(x_1, y_1), \dots, (x_n, y_n)$ a budeme chtít najít nejlepší regresní funkci ve tvaru $y = a + bx$. Vidíme nám již dobře známý vztah z lineární regrese. Je zde však jeden důležitý rozdíl, o kterém jsme

hovořili v kapitole 2.2. V tomto případě mohou být obě proměnné x, y zatíženy chybou, takže se dostáváme zpět k problémům s chybami v proměnných.

Již v roce 1877 bylo představeno řešení speciálního případu této úlohy, a sice když předpokládáme stejné rozptyly pro obě proměnné, tedy

$$\sigma_x^2 = \sigma_y^2. \quad (9)$$

Toto řešení je založeno pouze na geometrické interpretaci aproximace přímkou, více se můžeme dočíst v [4]. Podobně jako v kapitole 2.1 použijeme metodu nejmenších čtverců, tentokrát ale počítáme součet ortogonálních vzdáleností regresní přímky od příslušných bodů (x_i, y_i) . Tedy $\Phi = \sum_{i=1}^n d_i^2$, kde d_i je právě tato ortogonální vzdálenost. Pak dostáváme vztah

$$\Phi(a, b) = \frac{1}{1 + b^2} \sum_{i=1}^n (y_i - a - bx_i)^2.$$

Budeme chtít najít minimum této funkce, abychom obdrželi předpis pro nejlepší lineární aproximaci. Potom, co parciální derivaci funkce Φ podle a položíme rovnu nule, dostáváme výsledek $\hat{a} = \bar{y} - b\bar{x}$, kde \bar{x}, \bar{y} jsou výběrové průměry. Provedeme-li to samé, ale s parciální derivací funkce Φ podle b , dostaneme

$$\hat{b} = \frac{s_{yy} - s_{xx} + \sqrt{(s_{yy} - s_{xx})^2 + 4s_{xy}^2}}{2s_{xy}}, \quad (10)$$

kde s_{xx}, s_{xy}, s_{yy} jsou prvky kovarianční matice. Tento výsledek platí pro každý model, kde $s_{xy} \neq 0$. Takto jsme dostali odhady pro a i b a můžeme pomocí nich sestavit naši nejlepší lineární aproximaci $y = \hat{a} + \hat{b}x$. Tento vztah se postupně objevoval hned v několika publikacích a hned s několika drobnými chybami. Nejprve to byl Adcock [2] v roce 1878, o rok později Kummel [13] a nakonec už správně v publikaci Madanskeho [15] až z roku 1959, což je více než 80 let od první zmínky, takže vidíme, že to ani zdaleka není tak triviální, jak by se mohlo zdát na první pohled. V některých publikacích se tato metoda neoznačuje jako geometrická, ale jako ortogonální aproximace právě kvůli ortogonálním vzdálenostem, které chceme počítat. Co se týká přesnosti odhadů parametrů kružnice, nikde nezmiňujeme jejich přesnost ani intervaly spolehlivosti. Jedná se o poměrně složitou a obsáhlou kapitolu a tak jsme ji vynechali, vše je k nalezení v publikacích [[15, 3].

Vzhledem k tomu, že je tato teorie odvozená přímo z geometrie, tak nezáleží na tom, jaký souřadný systém zvolíme. Takže kdybychom například použili transformaci $T : (x, y) \rightarrow (x + c, y + d)$, zůstane naše regresní přímka stejná. Jediná změna, která by nastala, je v samotné rovnici. Ta by pak vypadala takto: $y + d = a + b(x + c)$. Obdobné tvrzení platí pro rotace či škálování. Jak si později ukážeme, tyto vlastnosti budou velmi užitečné.

Jak pracovat s modely bez požadavků na rovnost hodnot σ_x^2, σ_y^2 si řekneme v další kapitole.

2.5 Obecné řešení pro modely s chybami v proměnných

Pojďme se nyní podívat na řešení problému s chybami v proměnných bez předpokladu rovnosti rozptylů obou veličin. Máme tedy předpoklady (3) a (4) a s těmito předpoklady budeme dále pracovat, takže ani jeden z rozptylů našich dvou proměnných x, y není vázán žádnou další podmínkou.

V roce 1879 přišel Kummell [13] s tvrzením, že problém hledání vztahu $y = g(x)$, v kontextu modelu s chybami v proměnných, nemá řešení. Chybí nám podmínka, nebo alespoň nějaký další předpoklad o rozptylech. Tvrdil, že bude stačit dokonce jen znalost vzájemného vztahu mezi nimi.

Jeho tvrzení o neřešitelnosti vzbudilo velký rozruch až do padesátých let dvacátého století³. Postupně mu ale dávalo stále více odborníků za pravdu. Později, v roce 1956, bylo toto tvrzení formálně dokázáno v publikaci Andersona a Rubina [6]. (V důkazu se ukáže, že věrohodnostní funkce není omezená, a tedy maximálně věrohodné odhady nemůžeme získat. O pár let později se ukázalo v práci M. Solariho [19], že veškeré kritické body jsou pouze sedlové body, a nejedná se tedy o maxima).

Kummel přišel s tím, že mu stačí znát vztah mezi rozptyly. Tedy předpokládal, že zná hodnotu poměru κ , kde

$$\kappa = \sigma_x / \sigma_y. \quad (11)$$

Dále tvrdil, že výzkumníci z praxe stejně tento poměr typicky znají ze zkušenosti. Nedávno dokonce bylo použito označení „klasický EIV⁴ model“, který popisoval takový model, pro který platí (3), (4) a právě (11). K čemu ale tento vztah můžeme použít?

Místo počítání vertikálních vzdáleností u metody nejmenších čtverců, můžeme také počítat ortogonální vzdálenosti, nebo dokonce zešikmené vzdálenosti. Právě hodnota κ nám určí, jak moc zešikmené vzdálenosti budeme brát. Směrnice spojnice, kterou budeme chtít použít, je přímo závislá právě na hodnotě κ .

Nyní už náš problém má dobře definované řešení a nejlepší aproximaci přímky dostáváme ze vztahu

$$\Phi(a, b) = \frac{1}{1 + \kappa^2 b^2} \sum_{i=1}^n (y_i - a - bx_i)^2.$$

Tuto funkci budeme chtít minimalizovat pro oba parametry a dostaneme nejlepší regresní přímku ve smyslu EIV modelu. Pro speciální případ $\kappa = 1$ dostáváme řešení původního problému, kde jsme použili ortogonální metodu nejmenších čtverců.

³Madansky v [15] popisuje pobouření fyziků, kteří nechtěli věřit, že takto zdánlivě jednoduchý problém nemá řešení. A popisuje, jak by to fyzikovi mohl vysvětlit statistik.

⁴Error in variables.

Pro $\kappa \neq 1$ dostáváme pro a stejný odhad jako v předchozím případě, tedy $\hat{a} = \bar{y} - b\bar{x}$. Pro b dostaneme trošku jiný vztah než v (10), a sice

$$\hat{b} = \frac{\kappa^2 s_{yy} - s_{xx} + \sqrt{(\kappa^2 s_{yy} - s_{xx})^2 + 4\kappa^2 s_{xy}^2}}{2\kappa^2 s_{xy}}. \quad (12)$$

Pomocí těchto dvou parametrů jsme schopni opět dopočítat přímkou, kterou hledáme. Tato přímka pak bude nejlepší lineární aproximací našich dat ve smyslu EIV modelu.

Zde je na místě ještě krátká poznámka o škálování našich proměnných. EIV model se při vhodné volbě hodnoty κ stává neměnným při operaci posunutí. V případě, že budeme chtít měnit naše souřadnice $x \mapsto \alpha x$ a $y \mapsto \beta y$, oba dva parametry α, β můžeme zahrnout do hodnoty κ následovně $\kappa \mapsto \kappa\alpha/\beta$. Tento poznatek vlastně v reálném světě znamená, že nezáleží na jednotkách, ve kterých měříme naše data. Pro speciální případ, kdy škálujeme $x \mapsto x$ a $y \mapsto \kappa y$, dostaneme novou hodnotu $\kappa \mapsto 1\kappa/\kappa = 1$ a jsme zpět u našeho „jednoduchého“ modelu s ortogonálními vzdálenostmi a příslušnou metodou nejmenších čtverců. Což vlastně znamená, že obecný EIV problém, kde známe $\kappa = \sigma_x/\sigma_y$, můžeme pomocí transformace převést na jednoduchou ortogonální regresi.

2.6 Vlastnosti regresních odhadů

Opět předpokládejme, že platí (3), (4) a (9). Již víme, že za těchto předpokladů umíme najít správnou regresní funkci. Jedná se o ortogonální aproximaci.

Fakt, že naše odhadované parametry a, b nepochází z žádné „pěkné“ skupiny rozdělení, které bychom používali běžně, se vzhledem ke složitosti celého problému dá pochopit a očekávat. Důležitá a možná trochu překvapující vlastnost je ta, že a, b nemají konečné momenty, tedy

$$E(|\hat{a}|) = \infty, \text{ a } E(|\hat{b}|) = \infty,$$

což ústí v další, závažnější problém. Naše střední kvadratická odchylka bude rovněž nekonečná, tj.

$$E([\hat{a} - a]^2) = \infty, \text{ a } E([\hat{b} - b]^2) = \infty.$$

Oba tyto problémy jsou podrobněji diskutované v [4]. S těmito momenty se tedy pracovat reálně nedá, a tak se pracuje s aproximací, která je založená na Taylorově rozvoji. Což samozřejmě vyvolalo hodně rozruchu, neboť aproximace hodnoty, která neexistuje, nemůže dávat dobrý smysl. Anderson dále píše, že v tomto případě by se mělo hovořit o momentech některé aproximace, nikoliv o aproximacích momentů.

Zde nebudeme zabíhat příliš do detailů, ale z publikací [4, 5, 14, 17] plyne, že kvadratická odchylka u odhadu maximální věrohodnosti (\hat{b}_{MLE}) je nekonečná, ale

u odhadů pomocí metody nejmenších čtverců (\hat{b}_{LS}) je konečná. Vypadá to tedy, že metoda nejmenších čtverců je (nekonečně-krát) přesnější. Ale odhad \hat{b}_{MLE} je konzistentní a asymptoticky nestranný, zatímco \hat{b}_{LS} není. Takže to vypadá, že i \hat{b}_{MLE} je v jistém smyslu lepší odhad. Jak to tedy je? Anderson dále ukazuje, že platí následující vztah

$$P\{|\hat{b}_{MLE} - b| > t\} < P\{|\hat{b}_{LS} - b| > t\},$$

pro každé $t > 0$, malé. Velká t neuvažujeme, protože chceme, aby rozdíl mezi odhadem a pravou hodnotou parametru b byl malý. Odtud plyne, že odhad \hat{b}_{MLE} je lepší než odhad \hat{b}_{LS} všude, kde není odchylka od pravé hodnoty b příliš velká. Právě případ s velkým t způsobuje, že kvadratická odchylka odhadu \hat{b}_{MLE} je nekonečná.

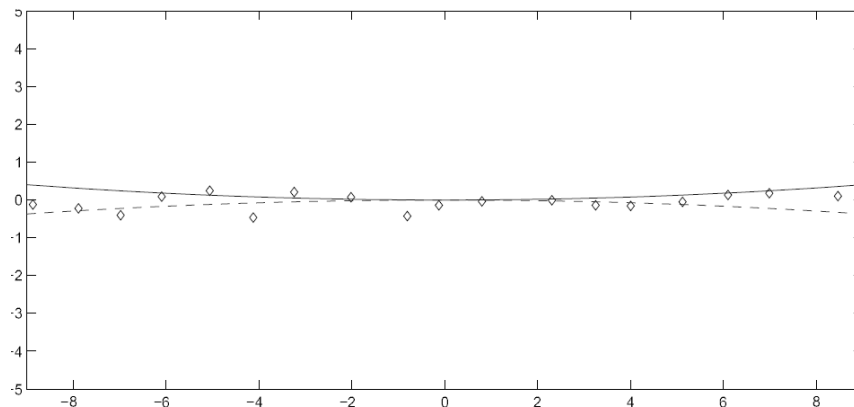
2.7 Nelineární modely

Budeme opět řešit model s chybami v proměnných, ale v tomto případě se nebude jednat o jeho lineární podobu. Nelineární model, je takový model, kdy nelze regresní funkci vyjádřit jako lineární kombinaci hledaných parametrů. Například regresní funkce popisující exponenciální růst: $y = ab^x$ spadá do množiny nelineárních modelů.

Máme tedy stejné problémy, které jsme zmínili v předchozích kapitolách, ale ještě nám bohužel některé přibudou. Když v matematice řešíme jakoukoli úlohu, tak se ptáme na základní tři otázky: existence řešení, jednoznačnost řešení a stabilita řešení. Zde postupně narazíme hned na několik komplikací, které budeme diskutovat.

Nelineární aproximace nemusí mít vždy řešení. U případu, kdy budeme chtít proložit kružnicí několik kolineárních bodů, je zřejmé, že se nám taková aproximace nebude hledat snadno a může se stát, že řešení vůbec nenajdeme. S problémem (ne)existence řešení se setkáváme typicky u nelineárních případů. Můžeme je proložit obloukem kružnice, která má střed relativně daleko (záleží na počtu bodů), nebo podobným obloukem kružnice se středem přesně na druhé straně. Chyba těchto odhadů bude velmi malá, ale rozdíl mezi těmito kružnicemi je obrovský (viz obrázek 5). Důkladněji se k tomu dostaneme v následující kapitole a na závěr celé diplomové práce budeme tuto úlohu simulovat v praktické části.

Dalším problémem je, že když už najdeme nejlepší aproximaci, tak se může stát, že nebude jednoznačně určená. Může zkrátka nastat, že dostaneme více řešení, která budou stejně dobrá. Což samozřejmě vede k závažným problémům v teorii. A když už máme to štěstí a řešíme případ s jednoznačným řešením, pak nevíme nic o odhadnutých parametrech. Neznáme vzorec pro hustotu ani pro momenty, ty dokonce často ani neexistují. Obecně odhadnuté parametry nejsme schopni spočítat přesně ze vzorce, jako tomu tak bylo u lineárních problémů. Jediná možnost, jak dostat jejich hodnotu, je pomocí numerických algoritmů,



Obrázek 5: Náhodně generovaná data podél přímky a dvě kružnice s částí odhadu oblouku těsně u sebe, přitom středy mají zřejmě velmi daleko od sebe.

tedy pouze jako aproximace.

Zde nastávají další problémy. Numerické metody typicky fungují iteračně. Jak je to s konvergencí numerických metod, je další veliké téma, které není předmětem této diplomové práce. Řekněme snad jen, že konvergence numerické metody ke globálnímu minimu není zdaleka samozřejmá. Mnohdy se nám stane, že se „zasekneme“ v lokálním minimu a požadované globální minimum zůstane netknuté. Tomu bychom se určitě rádi vyhnuli. Také může nastat, že máme k dispozici příliš málo kroků iterace k dispozici a algoritmus stále hledá, a nedostane se ani do lokálního, ani do globálního minima, protože to „nestihne“. Zkrátka výběr algoritmu a jeho nastavení je klíčové v nelineárních problémech. Když zvolíme jiný algoritmus, může se stát, že dostaneme jiné výsledky, nebo mají naše odhady jiné charakteristiky.

U lineárních modelů jsme viděli, že mezi EIV aproximacemi a ortogonálními existuje velmi úzký vztah, a že můžeme z jedné na druhou přecházet poměrně snadnou transformací. U nelineárních modelů tomu tak není. Když budeme chtít použít rotaci na lineární funkci, dostaneme opět lineární funkci. Když zkusíme použít operaci rotace na polynom například třetího stupně, dostaneme křivku, která nereprezentuje žádnou funkci. Pro každou nelineární funkci jsme schopni najít takovou rotaci, abychom dostali křivku, která není podle definice funkcí. A to je zřejmě problém. Typicky se setkáváme s polynomy stupňů jedna nebo dvě. Polynomy třetího a vyššího stupně se vyskytují velmi zřídka, a když už na ně narazíme, umíme je poměrně efektivně „rozsekat“ na dílčí segmenty, které potom odhadujeme pomocí polynomů nižšího stupně.

3 Odhady kuželoseček a kvadrik

Kde v praxi můžeme téma odhadu kuželoseček a kvadrik využít, jsme si naznačili v kapitole 1.3, delší výčet případně můžeme najít v [3]. V této kapitole se podíváme na několik teoretických problémů a řekneme si, jak na ně. Začneme pohledem na parametrizaci a na několik základních myšlenek odhadu kuželosečkou či kvadrikou.

V této kapitole bude velmi důležitá znalost důvodů, proč odhadujeme naše data. Když se například jedná o medicínská data, je přesnost odhadu určitě prioritou. Naproti tomu v oblasti fyziky, potřebujeme hlavně rychlé odhady, protože v urychlovači částic se vytvoří milióny částic za den, minimalizace odchylky pak není na prvním místě. Podle toho se bude lišit taktika použitá při samotném výpočtu.

Také se budou často lišit vstupní data. Někdy dostaneme body podél celé kružnice a někdy jen pár bodů v malé části velké kružnice. Opět se podle toho budeme řídit při výpočtu. Je tedy zřejmé, že se jedná o dosti komplexní problém.

3.1 Parametrizace

Mějme napozorované body (x_i, y_i) , chceme jimi proložit kružnici, tak aby součet ortogonálních čtverců byl minimální. Obecná rovnice kružnice vypadá následovně

$$(x - a)^2 + (y - b)^2 = R^2,$$

kde bod $[a, b]$ je střed kružnice a R je její poloměr. Hledaná vzdálenost se počítá jako

$$d_i = \sqrt{(x_i - a)^2 + (y_i - b)^2} - R.$$

Takže budeme minimalizovat výraz

$$\Phi(a, b, R) = \sum_{i=1}^n d_i^2 = \sum_{i=1}^n \left[\sqrt{(x_i - a)^2 + (y_i - b)^2} - R \right]^2. \quad (13)$$

Pomocí několika následujících úprav jsme schopni se zbavit proměnné R a pracovat pouze s funkcí dvou proměnných. Práce s funkcí dvou proměnných místo tří má zřejmou výhodu ve vizualizaci. Funkci se dvěma proměnnými jsme schopni poměrně intuitivně „vidět“, zatímco funkce více proměnných jsou pro představivost či grafické znázornění o mnoho náročnější. Funkce Φ je polynom druhého stupně vzhledem k parametru R , pro libovolné pevné hodnoty parametrů a, b . Pro kvadratický polynom platí, že má jediné globální minimum. Položme

$$r_i = r_i(a, b) = \sqrt{(x_i - a)^2 + (y_i - b)^2},$$

pak hledané minimum je rovno

$$R = \bar{r} = \frac{1}{n} \sum_{i=1}^n r_i.$$

Nyní se vrátíme k původní funkci Φ a dosadíme tento výsledek

$$\begin{aligned}\Phi(a, b) &= \sum_{i=1}^n [r_i - \bar{r}]^2 \\ &= \sum_{i=1}^n \left[\sqrt{(x_i - a)^2 + (y_i - b)^2} - \frac{1}{n} \sum_{j=1}^n \sqrt{(x_j - a)^2 + (y_j - b)^2} \right]^2.\end{aligned}$$

Není to příliš jednoduchý výraz, jak bychom doufali, ale nelze jej již dále upravit. Hledání minima této funkce, je nelineární problém, který rovněž nebude mít příliš jednoduché řešení.

Pohlédneme-li opět na obrázek 5, vidíme tam dva možné odhady naší kružnice. V takovém případě, když máme například několik kolineárních bodů, nastává velký problém. Můžeme se dostat ke kružnicím, které mají libovolně velký průměr. Navíc v každém případě budou tyto kružnice dvě, každá „z jedné strany“ našich bodů. Jak tedy s tímto problémem naložit?

Když nám jde pouze o naměřené body, minimum naší funkce bude velmi podobné pro kružnice na obou stranách. Nezáleží tedy na tom, kterou z těchto dvou kružnic označíme za řešení. Ale přidáme-li později další měření, můžeme zjistit, že jsme vybrali tu na špatné straně, a tedy jsou naše výsledky velmi daleko od správného odhadu.

Jedna z dalších možností, jak popsat kružnici, je pomocí algebraických parametrů. Tento přístup byl představen australským profesorem Vaughem Prattem v roce 1987 v [18]. Popisuje kružnici následující algebraickou rovnicí

$$A(x^2 + y^2) + Bx + Cy + D = 0,$$

kde pochopitelně požadujeme, aby $A \neq 0$. Když převedeme tento algebraický tvar na geometrický, dostáváme rovnici

$$\left(x + \frac{B}{2A}\right)^2 + \left(y + \frac{C}{2A}\right)^2 - \frac{B^2 + C^2 - 4AD}{4A^2} = 0.$$

V případě, že $B^2 + C^2 - 4AD = 0$, dostáváme právě jeden bod. A v případě, že $B^2 + C^2 - 4AD < 0$, dostáváme prázdnou množinu. Abychom zabránili mnohoznačnosti určení, musíme zavést novou podmínku, která zároveň vyloučí obě právě diskutované situace. Podmínka je tedy tvaru

$$B^2 + C^2 - 4AD = 1.$$

Nemáme sice přímo jednoznačnost, ale každou kružnici jsme schopni vyjádřit právě dvěma vektory parametrů (stejně až na znaménko). Čili společně s podmínkou $A > 0$ již dostáváme jednoznačnost určení libovolné kružnice.

Pomocí poměrně jednoduché úpravy jsme schopni přecházet z jedné parametrizace na druhou, stačí použít následující substituci

$$a = -\frac{B}{2A}, \quad b = -\frac{C}{2A}, \quad R^2 = \frac{B^2 + C^2 - 4AD}{4A^2}.$$

Dohromady tedy dostáváme parametrický prostor tvořen prvky (a, b, R) . Tento prostor je zřejmě spojitý, ale vzhledem k tomu, že tyto proměnné mohou nabývat libovolných hodnot, není omezen, a tedy nemůže být kompaktní. Podívejme se nyní na některé základní vlastnosti aproximace kružnicí, a jak je to s těmi třemi důležitými otázkami.

Pro nás bude pravděpodobně nejzajímavější bude následující parametrizace. Zde budeme používat parametry jak pro střed kružnice, tak pro poloměr, ale také pro úhel. Použijeme trigonometrické funkce a dostaneme dvě rovnice transformace polárních souřadnic na kartézské

$$\begin{aligned} x &= a + R \cos \varphi, \\ y &= b + R \sin \varphi, \end{aligned}$$

kde parametr φ se pohybuje v intervalu $[0, 2\pi)$. Takže opět dostaneme dohromady celou kružnici. Rovnice transformace v opačném směru pro parametr $\varphi \in [0, \frac{\pi}{2}]$, by vypadaly následovně

$$\begin{aligned} R &= \sqrt{(x - a)^2 + (y - b)^2}, \\ \varphi &= \arctan\left(\frac{y - b}{x - a}\right). \end{aligned}$$

V případě, že přidáme ještě jednu dimenzi, dostaneme se do oblasti kvadrik. Kvadrika je obecně algebraická plocha 2. stupně. Typicky hovoříme o kvadratické ploše v trojrozměrném prostoru. Například se tedy jedná o elipsoidy, paraboloidy, kuželové plochy, válcové plochy a další. Nás ale bude nejvíce zajímat koule. Podívejme se nyní na možné parametrizace pro kouli. Analytické vyjádření koule dobře známe a vypadá takto

$$(x - a)^2 + (y - b)^2 + (z - c)^2 \leq R^2,$$

kde samozřejmě bod $[a, b, c]$ je středem koule a R je jejím poloměrem. Hledaná vzdálenost se potom počítá jako

$$d_i = \sqrt{(x_i - a)^2 + (y_i - b)^2 + (z_i - c)^2} - R.$$

Což pro nás znamená, že bychom minimalizovali výraz

$$\Phi(a, b, c, R) = \sum_{i=1}^n d_i^2 = \sum_{i=1}^n \left[\sqrt{(x_i - a)^2 + (y_i - b)^2 + (z_i - c)^2} - R \right]^2.$$

Nyní máme kouli vyjádřenou pomocí 4 parametrů. To je příliš mnoho pro jakoukoli vizualizaci. Proto budeme hledat lepší parametrizaci s méně parametry. Použijeme trigonometrické funkce a dostaneme tzv. polární souřadnice

$$\begin{aligned}x &= a + R \sin \theta \cos \varphi, \\y &= b + R \sin \theta \sin \varphi, \\z &= c + R \cos \theta.\end{aligned}$$

V této parametrizaci je R konstanta a stále se jedná o poloměr koule. Pro naše dva parametry pak platí $\theta \in [0, \pi]$, $\varphi \in [0, 2\pi)$. θ je parametrem délky a φ je parametrem šířky. V podstatě jsou to přesně ty údaje, které používáme na popis lokace objektů na zeměkouli. V takovém případě hovoříme o zeměpisné délce a zeměpisné šířce.

V případě parametrizací kvadrik se nebudeme pouštět tak do hloubky. Zde jsme se jen chtěli podívat, jak vypadá rozšíření z kružnice na kouli. Jedná se skutečně jen o mechanické přidání dimenze. Parametry se chovají přesně tak, jak by člověk intuitivně čekal, a nenarazíme na žádný velký problém, který by bylo potřeba diskutovat. Tedy za poměrně málo úsilí se můžeme hned věnovat úlohám odhadu koule. Pokud bychom se ale chtěli pouštět do složitějších kvadrik, bude to obtížnější a obtížnější. Vybrali jsme na tomto místě, ten nejvíce příklad, abychom měli představu, jak se taková transformace z 2D do 3D provádí. Zbytek teorie projdeme už jen pro kuželosečky, zejména pro kružnici, ale vidíme, že rozšíření do třetí dimenze (kružnice \rightarrow koule) není nijak technicky náročné. Teď už se zaměříme na teorii samotných odhadů.

3.2 Teorie

Začneme hned poněkud překvapujícím tvrzením. Pro některá data, nemusí vůbec existovat nejlepší aproximace kružnicí. Jedná se opět o nám již známý příklad s n ($n \geq 3$) kolineárními body. Naše funkce Φ ze vztahu (13) bude mít pro každý vektor parametrů (a, b, R) kladnou hodnotu, nemůžeme ji zcela minimalizovat. Můžeme tuto hodnotu udělat libovolně malou – pomocí kružnic s velkým poloměrem, ale vynulovat tuto funkci neumíme. Pro každou kružnici bude existovat jiná kružnice, která bude mít ještě menší součet ortogonálních čtverců, a tedy ji nahradí jako řešení. Tento postup by se mohl stále opakovat, řešení tedy neexistuje.

Jednoduché a velmi elegantní řešení by v tomto případě bylo použít prostě přímku na proložení těchto dat. Tedy v případě, že máme hledat aproximaci kuželosečkou, nemusí být vždy špatný nápad podívat se na některé přímky, jestli nebudou řešit naši úlohu lépe.

Podívejme se ještě na jiný příklad než na množinu několika kolineárních bodů. Když vezmeme čtveřici bodů $(-X, 0)$, $(X, 0)$, $(0, 1)$, $(0, -1)$, kde $X \gg 1$, dostaneme se do situace, kdy pomocí žádné kružnice nedostaneme lepší řešení než

pomocí přímky. V tomto případě, by nejlepší řešení byla osa x . Dokonce můžeme i poměrně snadno dopočítat přesnou hodnotu naší funkce Φ , kterou se snažíme minimalizovat, v našem případě to bude $\Phi = 2$. Pro každou kuželosečku, použitou jako regresní funkci pro tento příklad dostane $\Phi \geq 2$.

Takových příkladů by se dalo najít více, ale v praxi se s nimi setkáváme jen velmi zřídka, a tak je lepší soustředit se na aproximace kružnicí či jinou kuželosečkou. Dokonce bylo dokázáno, že pokud budeme generovat náhodná data se spojitým rozdělením, je pravděpodobnost, že bude lepší aproximace přímka než kuželosečka rovna nule ([3] - Theorem 8).

V naší algebraické parametrizaci uvažujeme jak kružnice, tak přímky. V případě, že $A = 0$, se totiž jedná o rovnici přímky. Budeme určitě chtít v našich příkladech uvažovat možnost přímky, protože jsme si ukázali v kapitole 1.3, že se tato teorie často používá v rekonstrukci obrázku. Může se stát, že budeme aproximovat přímkou, nebo případný oblouk „rozsekáme“ a jeho části budeme aproximovat přímkou, vznikne tak po částech lineární křivka.

Co se jednoznačnosti týče, tak v ([3] - Section 3.5) je ukázáno několik velmi specifických příkladů, u kterých dochází k nejednoznačnosti řešení. Pro obecná, náhodně generována data se spojitým rozdělením pravděpodobností jsou však tyto příklady irelevantní. My se v této práci budeme zabývat obecným případem, a tedy v těchto konkrétních příkladech pouze přiložíme odkaz na příslušnou literaturu. Dále budeme pokračovat v teorii s myšlenkou, že bude existovat pouze jedno globální minimum funkce Φ .

3.3 Typy aproximace kružnicí

V podstatě veškeré úlohy aproximace kružnicí se počítají na počítačích, a tedy pomocí algoritmů. Tyto algoritmy se rozdělují do dvou hlavních skupin.

První z nich používá iterativní algoritmy, které konvergují k minimu funkce

$$\Phi = \sum_{i=1}^n d_i^2,$$

kde d_i jsou ortogonální vzdálenosti datových bodů od naší aproximace. Tuto skupinu metod označujeme pojmem geometrická aproximace. Jedná se obecně o přesnější přístup. Jelikož ale jde o iterativní algoritmus, musíme se tak potýkat s problematikou spojenou s tímto druhem algoritmu, jako jsou časová náročnost, náročnost na paměť, občasná divergence apod. Později zde v krátkosti uvedeme metodu sovětského fyzika Landaua jako příklad geometrické aproximace.

Nás bude v této práci více zajímat ta druhá skupina metod. Nejprve uvedeme pár základních informací obecně o algebraické aproximaci a pak si ukážeme tři příklady konkrétních metod a porovnáme je. Označuje se jako algebraická aproximace, protože vychází z algebraické rovnice kružnice, tj.

$$A(x^2 + y^2) + Bx + Cy + D = 0.$$

Vezmeme funkci Φ , kterou chceme minimalizovat u geometrických aproximací a zaměníme zde d_i za nějaké jiné hodnoty f_i , typicky se jedná o nějaké jednoduché algebraické výrazy. Potom minimalizujeme funkci $\Phi = \sum_{i=1}^n f_i^2$. Beze sporu největší výhodou metod algebraické aproximace je jejich rychlost a spolehlivost. Navíc přihlédneme-li k rychlosti výpočtu, je i přesnost výsledných odhadů poměrně vysoká.

Označme $\mathbb{A} = (A, B, C, D)^T$ jako 4-rozměrný vektor parametrů. Každá algebraická aproximace tedy minimalizuje výraz

$$\begin{aligned}\Phi(\mathbb{A}) &= \Phi(A, B, C, D) = \frac{1}{n} \sum_{i=1}^n (Az_i + Bx_i + Cy_i + D)^2 \\ &= n^{-1} \mathbb{A}^T (X^T X) \mathbb{A} = \mathbb{A}^T M \mathbb{A},\end{aligned}$$

za podmínky $\mathbb{A}^T N \mathbb{A} = 1$, pro nějakou matici N . V maticovém zápise jsme užili označení $z_i = x_i^2 + y_i^2$ a vztahů

$$M = \frac{1}{n} X^T X = \begin{pmatrix} \bar{z}z & \bar{z}x & \bar{z}y & \bar{z} \\ \bar{z}x & \bar{x}x & \bar{x}y & \bar{x} \\ \bar{z}y & \bar{x}y & \bar{y}y & \bar{y} \\ \bar{z} & \bar{x} & \bar{y} & 1 \end{pmatrix}, \text{ kde } X = \begin{pmatrix} z_1 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ z_n & x_n & y_n & 1 \end{pmatrix}$$

Podle podoby matice N z naší podmínky pak poznáme o jakou metodu algebraické aproximace se jedná.

Jedna z hlavních nevýhod geometrické aproximace je její rychlost. V případě řešení problému z fyziky, kde je potřeba velmi svižně odhadovat velké množství dat, si nemůžeme dovolit čekat. Proto v tomto případě použijeme algebraickou aproximaci i za cenu menší přesnosti. Ale o tom jsme již hovořili na začátku kapitoly 3.

3.3.1 Landauova metoda

Jak již bylo řečeno v předchozí kapitole, jedná se o geometrickou aproximaci, teoreticky by tedy měla být přesnější než ta algebraická. Jelikož je to iterativní metoda, bude do velké míry záležet na zvoleném počátečním řešení. O tom se dozvíme více v dalších kapitolách.

Cílem každé geometrické aproximace, je minimalizovat výraz

$$\Phi(a, b, R) = \sum_{i=1}^n \left[\sqrt{(x_i - a)^2 + (y_i - b)^2} - R \right]^2 = \sum_{i=1}^n d_i^2,$$

kde d_i jsou geometrické, tedy ortogonální vzdálenosti napozorovaných bodů od jejich aproximace na kružnici. Výraz trochu upravíme, ať je lépe vidět, jak uvažoval

Landau. Položme

$$\begin{aligned} r_i &= \sqrt{(x_i - a)^2 + (y_i - b)^2}, \\ u_i &= -\partial r_i / \partial a = (x_i - a) / r_i, \\ v_i &= -\partial r_i / \partial b = (y_i - b) / r_i. \end{aligned} \tag{14}$$

Snadno se pak dopočítá Jakobián jako

$$J = \begin{pmatrix} -u_1 & -v_1 & -1 \\ \vdots & \vdots & \vdots \\ -u_n & -v_n & -1 \end{pmatrix}.$$

Nyní položíme $g = [r_1 - R, \dots, r_n - R]$ a máme

$$J^T g = n \begin{pmatrix} R\bar{u} - \bar{u}r \\ R\bar{v} - \bar{v}r \\ R - \bar{r} \end{pmatrix} = n \begin{pmatrix} R\bar{u} + a - \bar{x} \\ R\bar{v} + b - \bar{y} \\ R - \bar{r} \end{pmatrix}.$$

Použili jsme zkrácený zápis pro výběrové průměry, kde např. $\bar{u}\bar{u} = \frac{1}{n} \sum_i^n u_i u_i$.

Landau říká, že hledání minima funkce Φ , je ekvivalentní s hledáním řešení rovnice $\nabla\Phi = 0$. Což je ekvivalentní s hledáním řešení rovnice $J^T g = 0$. To lze za předpokladu $\bar{x} = \bar{y} = 0$ napsat jako

$$\begin{aligned} a &= -\bar{u}R, \\ b &= -\bar{v}R, \\ R &= \bar{r}. \end{aligned} \tag{15}$$

Konečně se dostáváme k jeho návrhu na iterační proces. Pomocí počátečního řešení napočítáme $\bar{r}, \bar{u}, \bar{v}$ a pak postupujeme následovně:

$$\begin{aligned} R_{\text{new}} &= \bar{r}, \\ a_{\text{new}} &= -\bar{u}R_{\text{new}}, \\ b_{\text{new}} &= -\bar{v}R_{\text{new}}. \end{aligned}$$

Nebudeme se zde pouštět do hloubky, jde skutečně jen o ukázkou toho, jak se chová alespoň jedna geometrická aproximace. Abychom si udělali nějakou úvodní představu a později ji mohli porovnat s výsledky výpočtů z další kapitoly. Více dopodrobna jsou Landauova metoda a také další geometrické aproximace rozepsány v [3].

3.3.2 Kasova metoda

Tato metoda byla poprvé představena v v sedmdesátých letech dvacátého století a celé jméno zní Delogne – Kasova metoda. Jedná se o velmi populární metodu zejména díky její jednoduchosti, rychlosti a efektivnosti.

V případě, že bod (x_i, y_i) leží na naší kružnici, tato metoda najde kružnici přesně a velmi rychle. Při použití některých metod geometrické aproximace by to trvalo několik iterací a ani bychom správné řešení nemuseli dostat. Ze tří metod, na které se podíváme, je to určitě ta nejjednodušší. Chceme minimalizovat výraz

$$\Phi_1(a, b, R) = \sum_{i=1}^n [(x_i - a)^2 + (y_i - b)^2 - R^2]^2.$$

Nahradili jsme tedy d_i pomocí jednoduchého algebraického výrazu $f_i = (x_i - a)^2 + (y_i - b)^2 - R^2$. A chceme minimalizovat funkci $\Phi_1 = \sum f_i^2$.

Někteří autoři v souvislosti s touto metodou hovoří o *algebraické vzdálenosti*, když hovoří o vzdálenosti bodu (x_i, y_i) od kružnice $(x_i - a)^2 + (y_i - b)^2 + R^2 = 0$.

Chceme-li nyní derivovat naši funkci Φ_1 podle parametrů a, b, R , dostaneme nelineární vztah, a tomu se určitě chceme vyhnout. Proto použijeme následující substituci

$$B = -2a, \quad C = -2b, \quad D = a^2 + b^2 - R^2.$$

A dostáváme

$$\Phi_1 = \sum_{i=1}^n (z_i + Bx_i + Cy_i + D)^2,$$

kde $z_i = x_i^2 + y_i^2$. Nyní po derivaci podle parametrů B, C, D dostáváme systém lineárních rovnic, který pomocí algebraických metod umíme dobře řešit

$$\begin{aligned} \frac{1}{n} \sum x_i^2 B + \frac{1}{n} \sum x_i y_i C + \frac{1}{n} x_i D &= \frac{1}{n} \sum x_i z_i, \\ \frac{1}{n} \sum x_i y_i B + \frac{1}{n} \sum y_i^2 C + \frac{1}{n} y_i D &= \frac{1}{n} \sum y_i z_i, \\ \frac{1}{n} \sum x_i B + \frac{1}{n} \sum y_i C + D &= \frac{1}{n} \sum z_i. \end{aligned}$$

Náš na první pohled možná trošku chaotický systém lineárních rovnic má poměrně hezké řešení

$$a = -\frac{B}{2}, \quad b = -\frac{C}{2}, \quad R = \frac{\sqrt{B^2 + C^2 - 4D}}{2}.$$

Tuto metodu lze také zavést i geometricky (dokonce hned několika způsoby, viz [3] - kapitola 5), tedy je zřejmé, že se jedná o metodu invariantní k operacím otočení či posunutí.

Jde o pravděpodobně tu nejrychlejší z metod algebraické aproximace. Bohužel, jak tomu tak bývá, se tato rychlost odráží v přesnosti našich výsledků a právě zde Kasova metoda pokulhává. Po krátké analýze lze ukázat, že Kasova metoda pro body, které nejsou kolineární, najde právě jedno řešení. Plyne to z důkazu

pozitivní definitnosti systému lineárních rovnic, který jsme odvodili výše. V případě, že se jedná o kolineární body, nám Kasova metoda nabídne více než jedno řešení.

Tato metoda funguje velmi slušně v případech, kdy jsou data rozmístěna kolem kružnice, či její poloviny. Jakmile budeme mít data kolem malého oblouku, Kasova metoda bude generovat menší kružnice a kvalita výsledku se rapidně zhoršuje. Pokud budeme řešit příklady s neúplnými oblouky, opět narazíme na problém s přesností výsledků.

Pro úplnost dodáme, že v případě Kasovy metody má matice N z předchozí kapitoly tuto podobu

$$N = K = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

3.3.3 Prattova metoda

Tato metoda vznikla jako vylepšení Kasovy metody a vychází z algebraické parametrizace

$$A(x^2 + y^2) + Bx + Cy + D = 0,$$

a tedy budeme chtít minimalizovat výraz

$$\Phi_2 = \frac{1}{B^2 + C^2 - 4AD} \sum_{i=1}^n (Az_i + Bx_i + Cy_i + D)^2.$$

Jelikož pro každou kružnici platí $B^2 + C^2 - 4AD > 0$, promění se náš minimalizační problém v jiný, mnohem jednodušší

$$\Phi_3(A, B, C, D) = \sum_{i=1}^n (Az_i + Bx_i + Cy_i + D)^2,$$

za podmínky, že $B^2 + C^2 - 4AD = 1$. Vidíme, že za platnosti této podmínky budeme v otázkách o nejlepší aproximaci uvažovat jak kružnice, tak přímky. To je jeden z těch důležitých rozdílů mezi Kasovou a Prattovou metodou, význam tohoto rozdílu jsme popsali v předchozí kapitole.

Vaughan Pratt na řešení minimalizačního problému použil algoritmus založen, na singulárním rozkladu. Jedná se tedy o využití lineární algebry, zejména pak vlastních vektorů. Tato metoda je přibližně 2-3x pomalejší než Kasova. Ale například geometrické aproximace jsou až 10x pomalejší!

Pro $\sigma \rightarrow 0$ se naše odhady parametrů (a, b, R) blíží skutečným hodnotám těchto parametrů, hovoříme tedy o konzistentní metodě. Další důležitá vlastnost této metody, je invariance k operacím otočení či posunutí.

V případě Prattovy metody má matice N z předchozí kapitoly následující podobu

$$N = P = \begin{pmatrix} 0 & 0 & 0 & -2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -2 & 0 & 0 & 0 \end{pmatrix}.$$

3.3.4 Taubinova metoda

Poslední metoda, kterou v této práci zmíníme, je metoda Gabriela Taubina představena v jeho publikaci: *Estimation Of Planar Curves, Surfaces And Non-planar Space Curves Defined By Implicit Equations, With Applications To Edge And Range Image Segmentation*. Jedná se opět o metodu algebraické aproximace. Je to velmi podobná metoda té Prattově. Hlavní rozdíl je, že se dá snadno zobecnit na elipsy a další algebraické křivky, které by se nám mohly hodit.

Odvození Taubinovy metody je již poněkud složitější, a tak zde přiložíme pouze odkaz na ([3] - Section 5.9) a raději vyčteme pár vlastností této metody. Taubinova metoda je podobná Prattově nejen ve funkci, již minimalizuje, ale i poměr „cena/výkon“ je podobný. Kasova metoda zaostává v tomto měřítku za oběma metodami. Opět se jedná o metodu invariantní k operacím posunutí a otočení.

V nápovědě často využívaného softwaru Matlab se můžeme dokonce dočíst, že je tato metoda stabilnější než Kasova metoda, a o trochu rychlejší než Prattova metoda.

V případě Taubinovy metody má matice N z předchozí kapitoly tuto podobu

$$N = P = \begin{pmatrix} 4\bar{z} & 2\bar{x} & 2\bar{y} & 0 \\ 2\bar{x} & 1 & 0 & 0 \\ 2\bar{y} & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

3.4 Kombinace metod

Hlavní nevýhoda geometrických aproximací je v tom, že jsou silně závislé na počátečním odhadu. Jakmile není počáteční řešení dobré, nemůžeme očekávat přesné řešení geometrické aproximace. Proto se typicky obě metody kombinují. Pomocí algebraické aproximace nalezneme poměrně rychle relativně dobrý odhad a dosadíme ho do geometrické aproximace jako počáteční řešení, které nám pomůže k větší přesnosti. Podobně se používá například průměr jako počáteční řešení pro složité algoritmy počítající střední hodnotu. Čím lepší (přesnější) počáteční řešení máme, tím lépe bude fungovat náš algoritmus. Toto tvrzení platí obecně v numerických metodách, nejen v našem případě. Proto budeme každý algoritmus opakovat hned několikrát a s různými počátečními řešeními, šance na selhání při hledání globálního minima se tak zdatelně redukuje.

Chernov a Lesort v publikaci [9] diskutují o vhodných kombinacích algebraických a geometrických aproximací. Závěr je takový, že při odhadování velkých oblouků nebo celých kružnic, jsou všechny možné kombinace úspěšné na 100 %. Jakmile se ale jedná o menší oblouky, Kasova metoda se nedoporučuje jako počáteční řešení pro žádnou geometrickou aproximaci. Naopak modernější metody, jako jsou Prattova a Taubinova, jsou natolik dobré, že už je geometrické aproximace příliš nevylepší. Proto je většinou stačí použít samostatně.

Je zřejmé, že Kasova metoda je nejjednodušší. Znamená to, že se bude počítat nejrychleji a zároveň není příliš náročná na paměť ani sílu softwaru, na kterém je počítána. Pochopitelně to má i svou daň. Výsledky Kasovy metody jsou z pravidla mnohem méně přesné než výsledky Prattovy či Taubinovy metody.

Tyto metody jsou z kategorie algebraických aproximací. Ty jsou typicky výrazně horší v přesnosti výsledků. Když srovnáme Prattovu a Taubinovu metodu, vidíme skoro ihned, že jsou výrazně přesnější než Kasova. Dokonce jsou tak přesné, že se jejich výsledky blíží geometrickým aproximacím. Kdybychom měli uvést pořadí těchto metod v přesnosti, bylo by následující: Kasova metoda, Prattova metoda, Taubinova metoda, geometrické aproximace. Dokonce pro velké oblouky je Taubinova metoda svou přesností blíže ke geometrickým aproximacím než k metodě Prattově. To je informace jen pro úplnost, ve skutečnosti jsou ty rozdíly natolik malé, že je v podstatě jedno, jakou metodu vybereme. Více pak v tabulkách níže, které vyhodnocují napočítané příklady.

3.5 Další rozšíření

Tématem diplomové práce jsou odhady kuželoseček a kvadrik. My jsme se až na jednu krátkou pasáž v teoretické části věnovali pouze odhadům kružnice. Kružnice, jak známo, je speciální případ elipsy, jedná se tedy také o kuželosečku. Vybrali jsme kružnice ze dvou prostých důvodů. Za prvé jsme nechtěli čtenáře zahltnout příliš náročnou teoretickou částí, již teď v ní narazíme na některé poměrně složité myšlenky se složitými výrazy, a to máme jen jednu hodnotu pro poloměr. A také jsme chtěli, aby byly všechny naše obrázky přehledné a bylo na nich na první pohled vidět, že se jedná o kružnice. V praktické části se přesvědčíme, jak moc se tento plán vydařil.

Kružnice je takovou elipsou, která má shodné obě poloosy. V případě některých odhadů parametrů uvedených v teoretické části by se veškeré výrazy staly mnohem složitějšími a již by nebylo snadné se v nich vyznat. Rozšíření na elipsy z hlediska výpočtů není nijak obtížné, stačí do definovaných funkcí v softwaru *R* zadat místo poloměru kružnice dva parametry (pro obě poloosy) a výsledek i jeho forma bude obdobná. Přímou v balíčku, který používáme v praktické části, existuje například funkce s názvem *EllipseFitByTaubin*, která veškeré naše výpočty zvládne také. Problémy, jak již bylo naznačeno, nastávají v teorii, kde se skutečně jedná o složité výrazy, a práce by tak byla několikanásobně delší.

Co se týče kvadrik, ty jsme vynechali prakticky úplně. Rozšíření do třetí

dimenze jsme v krátkosti diskutovali v podkapitole o parametrizaci. Rozšíření z kružnice na kulovou plochu je velmi jednoduché. Naopak rozšíření z elipsy na elipsoid je už poměrně náročné a z tohoto důvodu jsme upozadili tyto aplikace celé teorie. V literatuře [3] se můžeme dočíst o odhadu koule. V této práci jsme se tedy soustředili na odhady kružnic.

4 Praktické úlohy

Dostáváme se do nejdůležitější kapitoly z celé práce. Praktická část, která je popsána v této kapitole, by nám měla potvrdit, všechny doposud zmíněné teoretické předpoklady. Úlohy jsou zpracované ve statistickém softwaru R a použitý kód je přiložen k práci k nahlédnutí či reprodukci v podobě scriptu.

Pro naše výpočty je klíčový balíček funkcí *conicfit*, který napsali společně Nikolai Chernov a Jose Gama. Obsahuje přesně funkce, které se nám hodí k výše zmíněným metodám. Pro každou funkci použitou ve výpočtu nám vrátí vektor tří hodnot. První dvě hodnoty udávají souřadnice středu odhadnuté kružnice a poslední hodnota je její poloměr. K vykreslení veškerých grafů jsme použili balíček funkcí *ggplot2*. Všechny nezbytné knihovny, použité ve scriptu, jsou naimportovány na prvních řádcích, tak aby script fungoval každému čtenáři. Pro přehlednost kódu je script dělaný pro jednu metodu. Všechny ostatní metody jsou počítány analogickým kódem, pro stejná vstupní data.

Všechny praktické příklady budeme ukazovat na odhadu kružnicí. Rozšíření na elipsy, paraboly, koule apod. je také možné s funkcemi ze stejného balíčku. Protože je na kružnicích nejlépe vidět, jak si které metody počínaly, zvolili jsme tuto variantu. Postup bude pro každý příklad stejný. Nejprve vygenerujeme N bodů na kružnici. Potom je náhodně posuneme a budeme sledovat, jak se jednotlivé metody přiblíží původní kružnici. Budeme uvažovat několik různých zadání a pro každé zadání provedeme 500 simulací. Důvod je prostý, chceme mít jistotu, že se skutečně jedná o trend, a ne o jedno/dvě nešťastná pozorování.

Jak porovnat přesnost výsledků? Zavedeme si hned tři jednoduché metriky, které budeme porovnávat. Asi nejzajímavější pro nás bude sledovat vzdálenost odhadnuté kružnice od té původní. Porovnávat tedy budeme souřadnice středů kružnic a hodnotu poloměrů. Tady budeme uvažovat hodnotu rozdílu obou veličin, tedy

$$\begin{aligned}\|\cdot\|_S &= \sqrt{(a - \hat{a})^2 + (b - \hat{b})^2}, \\ \|\cdot\|_R &= |R - \hat{R}|.\end{aligned}$$

První metrika s označením $\|\cdot\|_S$ se v našich úlohách v podstatě napočítá sama. Jako střed našich kružnic jsme zvolili bod počátku souřadnicového systému, tedy vzdálenost středů budeme pozorovat přímo na hodnotách odhadnutých parametrů \hat{a}, \hat{b} .

Třetí ukazatel, který nás bude zajímat, bude souviset hlavně s kvalitou dané aproximace. Budeme počítat pro každý vygenerovaný bod jeho geometrickou vzdálenost od odhadnuté kružnice. Použitý vzorec má tuto podobu

$$d_i^2 = \left[\sqrt{(\hat{a} - x_i)^2 + (\hat{b} - y_i)^2} - R \right]^2.$$

$$\| \cdot \|_{RMSE} = \sqrt{\frac{\sum_{i=1}^n d_i^2}{N}}.$$

Jedná se o odmocninu střední kvadratické chyby pro dvourozměrná data a ve výsledných tabulkách bude označována jako *RMSE* z anglického *Root Mean Square Error*. Důležité je připomenout, že se snažíme odhadnout původní kružnici. Takže tato třetí metrika neříká přímo, jak dobře jsme vyřešili zadanou úlohu. Dává nám pouze informaci o tom, jak dobře jsme odhadli náhodně posunuté body. Je tedy potřeba mít stále na mysli, na jaké hodnoty se díváme a co vlastně znamenají. Pomocí kombinace těchto tří metrik si již umíme udělat dobrý názor na to, jak která metoda funguje, za jakých parametrů.

Postupně se podíváme na data generovaná kolem celé kružnice, poté na data generovaná pouze kolem části kružnice. Bude nás zajímat oblouk pro $\varphi \in [0, \pi]$ a také $\varphi \in [0, \frac{\pi}{2}]$, kde φ vyjadřuje úhel, pod kterým budeme generovat oblouk. V případě celé kružnice je $\varphi \in [0, 360]$ apod. Na závěr se podíváme na některé speciální úlohy. Budeme chtít ukázat, že teorie vyřčená v předchozích kapitolách kolem několika kolineárních bodů byla uvedena správně. A také se podíváme na příklad, kde máme 4 body o souřadnicích $(-X, 0)$, $(X, 0)$, $(0, 1)$, $(0, -1)$, kde $X \gg 1$. O takové úloze jsme hovořili v teoretické části v podkapitole 3.2 a uvedli jsme, že pomocí žádné kružnice nedostaneme lepší řešení než pomocí přímky. Úlohu zadáme hned několika způsoby a přesvědčíme se o pravdivosti tvrzení z teoretické části.

Všechny tyto úlohy budeme počítat s různými hodnotami pro směrodatnou odchylku náhodné chyby generovaných dat. Výsledky budou ilustrovány obrázky a přiložena bude tabulka zapsaných výsledků podle našich zavedených metrik. Poslední graf bude pro vybrané úlohy heatmapa, ve které bude zaznamenán výsledek napočítaných RMSE hodnot pro různý počet generovaných bodů a pro různě nastavenou směrodatnou odchylku náhodných chyb.

Všechny naše úlohy se budou počítat analogicky. V prvním kroku určíme teoretickou kružnici. Ve druhém kroku vygenerujeme na této kružnici několik ($N = 25$) bodů a pomocí náhodné chyby z normálního rozdělení (s různými směrodatnými odchylkami) tyto body posuneme. V dalším kroku se budeme dívat na veškeré doposud diskutované metody a jejich odhad původního objektu. V případě geometrických aproximací, které se počítají iteračně, jsme zvolili maximální počet iterací roven 50 a toleranci rovnu $1e - 06$. Výsledné odhady poté vykreslíme pomocí *ggplot* funkce a porovnáme nejdříve jen vizuálně a později i podle právě zavedených metrik.

Aproximace budeme provádět čtyřmi metodami. Nejprve provedeme tři algebraické aproximace (Kasovu, Prattovu a Taubinovu). Podíváme se také na jednu geometrickou aproximaci (Landauovu). Tu ale použijme hned několikrát, ať vidíme, jak fungují jednotlivé kombinace algebraické a geometrické aproximace. Nejprve napočítáme Landauovu metodu bez počátečního řešení a poté použijeme jako počáteční řešení postupně všechny tři výsledky našich algebraických odhadů. Tedy sloupec s názvem $Landau(K)$ bude označovat metodu geometrické aproximace typu Landau, kde jsme jako počáteční řešení vzali výsledek Kasovy aproximace.

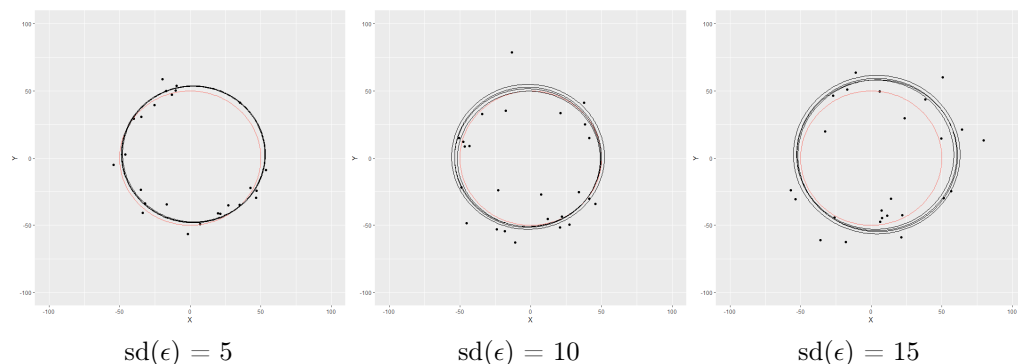
4.1 Odhadování kružnice

Cílem úlohy je nalézt odhady dané kružnice pomocí určených algebraických a geometrických metod aproximace. Původní kružnice se středem v bodě $S = [0, 0]$ a poloměrem $R = 50$ je reprezentována $N = 25$ body v celém prostoru, tedy platí $\varphi \in [0, 2\pi]$. Tyto body byly posunuty o náhodnou chybu e , pro kterou platí $e \sim N(0, sd(\epsilon))$. Směrodatná odchylka náhodné chyby bude postupně rovna hodnotám 5, 10, 15 a 20.

V našem výpočtu nejprve nastavíme vstupní parametry

- úhel $\varphi \in [0, 2\pi]$
- počet opakování $i = 500$
- počet generovaných bodů $N = 25$
- směrodatnou odchylku náhodné chyby postupně $sd(\epsilon) = 5, 10, 15, 20$

A poté můžeme pustit simulaci. V balíčku *conicfit* najdeme vhodné funkce a aplikujeme je na naše data. Výsledky pro každou iteraci zapíšeme do příslušného sloupce a podíváme se na výsledky. Na obrázku 6 vidíme, jak vypadají odhadnuté kružnice pro všechny metody, o kterých jsme mluvili v teoretické části této práce.



Obrázek 6: Odhady daných metod pro úlohu s 25 náhodně generovanými body, kde $\varphi \in [0, 2\pi]$ a směrodatná odchylka náhodných chyb nabývá různých hodnot. Červeně teoretická kružnice, černě odhady daných metod.

Červenou barvou je znázorněna teoretická kružnice. Ta má rovnici

$$x^2 + y^2 = 50^2.$$

Pro nízké směrodatné odchylky náhodné chyby vidíme, že se výsledky všech metod velmi podobají. Proto to na první pohled vypadá, že je odhadnuta jen jedna kružnice. Ve skutečnosti jsou tam ale všechny určené metody. To je závěr, který plyne ze všech úloh, které budeme počítat.

Na obrázku 7 je tabulka napočítaných hodnot třetí metriky pro úlohu, kde $\text{sd}(\epsilon) = 5$.

Tabulka vznikne tak, že při každé iteraci napočítáme RMSE pro každou metodu a její hodnotu zaneseme do sloupce se správným názvem. Tedy pro sloupec s názvem *Kasa_RMSE* se jedná o vektor o velikosti 500 s hodnotami RMSE pro každou iteraci. Software *R* nám pak umožňuje pomocí funkce *summary* nahlédnout na základní charakteristiky tohoto vektoru. Nejzajímavější pro nás budou průměrné a prostřední hodnoty jednotlivých vektorů, tedy řádky *Mean*, *Median*. Zajisté je ale dobré prozkoumat i ostatní hodnoty, jako například *Min*, *Max*, ke kterým se budeme vracet v další podkapitole. Celkově nám proto příkaz *summary* dává ucelenou představu o výkonu každé metody ve smyslu třetí zavedené metriky. Na prvním řádku je výsledek RMSE pro původní kružnici. Další řádky pak ukazují výsledky použitých metod.

Z tabulky na obrázku 7 je zřejmé, že tato úloha (pro případ $\text{sd}(\epsilon) = 5$) je pro všechny metody poměrně jednoduchá a není zde příliš prostoru pro chybu. Je vidět, že všechny metody si počínají velmi podobně a navíc jsou jejich výsledky velmi blízké teoretické kružnici (vidíme porovnáním výsledků RMSE pro jednotlivé metody a původní kružnici).

Čím vyšší je směrodatná odchylka náhodné chyby, tím se naše výsledky zhoršují. Až v případě $\text{sd}(\epsilon) = 20$ je vidět rozdíl mezi algebraickou a geometrickou


```

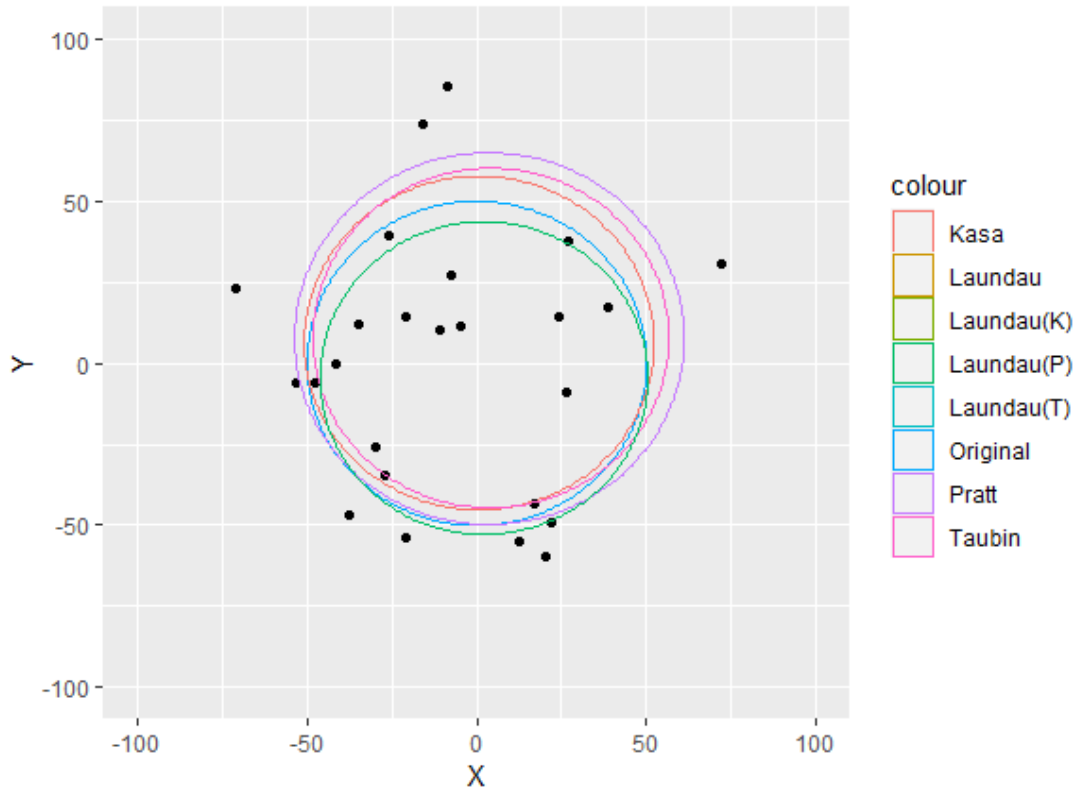
> summary(Original_RMSE)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 2.842  4.229  4.643  4.674  5.144  6.797
> summary(Algebraic_RMSE)
  Kasa_RMSE      Pratt_RMSE      Taubin_RMSE
Min.   :2.845   Min.   :2.852   Min.   :2.844
1st Qu.:4.238   1st Qu.:4.266   1st Qu.:4.236
Median :4.653   Median :4.689   Median :4.652
Mean   :4.689   Mean   :4.728   Mean   :4.687
3rd Qu.:5.161   3rd Qu.:5.208   3rd Qu.:5.159
Max.   :6.831   Max.   :6.955   Max.   :6.824
> summary(Geometric_RMSE)
  Landau_RMSE  Landau_Kasa_RMSE  Landau_Pratt_RMSE  Landau_Taubin_RMSE
Min.   :2.842   Min.   :2.842   Min.   :2.842   Min.   :2.842
1st Qu.:4.229   1st Qu.:4.229   1st Qu.:4.229   1st Qu.:4.229
Median :4.643   Median :4.643   Median :4.643   Median :4.643
Mean   :4.674   Mean   :4.674   Mean   :4.674   Mean   :4.674
3rd Qu.:5.144   3rd Qu.:5.144   3rd Qu.:5.144   3rd Qu.:5.144
Max.   :6.797   Max.   :6.797   Max.   :6.797   Max.   :6.797

```

Obrázek 7: Tabulka hodnot metriky RMSE jednotlivých metod pro úlohu s 25 náhodně generovanými body, kde $\varphi \in [0, 2\pi]$ a $\text{sd}(\epsilon) = 5$.

aproximací. Landauova metoda zde, podle předpokladu, poráží všechny algebraické aproximace a vychází nejlépe. Za povšimnutí stojí fakt, že veškeré číselné charakteristiky všech variant Landauovy metod jsou stejné. Znamená to, že tato úloha byla jednoduchá a tak dostane Landauova metoda stejný výsledek, bez ohledu na zvolené počáteční řešení.

Na obrázku 8 vidíme všechny metody najednou. Původní kružnice je tentokrát modrou barvou a ostatní typy odhadů jsou popsány v legendě grafu. Podívejme se nyní na hodnoty zavedených metrik. U geometrické metody vidíme opět jen jednu kružnici, neboť jsou odhady totožné pro všechna zvolená počáteční řešení. Zřejmě pro úlohy, kdy je $\varphi \in [0, 2\pi]$, mohou takové výsledky nastat.



Obrázek 8: Odhady daných metod pro úlohu s 25 náhodně generovanými body, kde $\varphi \in [0, 2\pi]$ a $\text{sd}(\epsilon) = 20$.

```
> summary(Original_RMSE)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 10.26 15.36  16.98  17.07 18.76  24.71
> summary(Algebraic_RMSE)
  Kasa_RMSE      Pratt_RMSE      Taubin_RMSE
Min.   :10.57   Min.   :10.95   Min.   :10.52
1st Qu.:15.72   1st Qu.:16.58   1st Qu.:15.67
Median :17.47   Median :18.65   Median :17.51
Mean   :17.62   Mean   :18.82   Mean   :17.66
3rd Qu.:19.42   3rd Qu.:20.88   3rd Qu.:19.43
Max.   :25.88   Max.   :29.03   Max.   :27.14
> summary(Geometric_RMSE)
  Landau_RMSE  Landau_Kasa_RMSE  Landau_Pratt_RMSE  Landau_Taubin_RMSE
Min.   :10.26   Min.   :10.26   Min.   :10.26   Min.   :10.26
1st Qu.:15.36   1st Qu.:15.36   1st Qu.:15.36   1st Qu.:15.36
Median :16.98   Median :16.98   Median :16.98   Median :16.98
Mean   :17.07   Mean   :17.07   Mean   :17.07   Mean   :17.07
3rd Qu.:18.76   3rd Qu.:18.76   3rd Qu.:18.76   3rd Qu.:18.76
Max.   :24.71   Max.   :24.71   Max.   :24.71   Max.   :24.71
```

Obrázek 9: Tabulka hodnot metriky RMSE jednotlivých metod pro úlohu s 25 náhodně generovanými body, kde $\varphi \in [0, 2\pi]$ a $\text{sd}(\epsilon) = 20$.

Na obrázku 9 vidíme, jak podobné jsou si výsledky pro různé metody. Nedá se přímo říct, že by geometrické aproximace měly kvalitnější odhady. Co se týče napočítaných vzdáleností generovaných bodů od odhadované kružnice, průměry a mediány všech metod se pohybují kolem čísla 17. Možná je trochu překvapivý závěr, že zde geometrické úlohy nemají jednoznačně navrch.

Podívejme se na další tabulku (Obrázek 10), která znázorňuje hodnoty odhadů středu kružnice a jejího poloměru. Co se týče polohy středu odhadovaných kružnic, opět zde nemáme jednoznačně nejlepší a nejhorší metody. Všechny metody odhadly střed velmi slušně. U odhadu poloměru kružnice už je ale vidět znatelnější rozdíl. Zde mají geometrické aproximace lepší výsledek. Poněkud zajímavý výsledek je, že Landauova metoda bez počátečního řešení má stejný výsledek jako Landauova metoda, do které jsme „vložiteli“ výsledek Taubinovy metody jako počáteční řešení. Rozdíl mezi výsledky těchto dvou metod není vidět, dokud se nepokusíme odhadovat složitější úlohy.

```
> summary(result_a)
  a_Kasa      a_Pratt      a_Taubin      a_Landau      a_Landau_Taubin
Min.   :-20.2229 Min.   :-22.3617 Min.   :-23.4236 Min.   :-23.4120 Min.   :-23.4169
1st Qu.: -5.3012 1st Qu.: -5.6249 1st Qu.: -5.6202 1st Qu.: -5.6820 1st Qu.: -5.6819
Median : -1.0328 Median : -0.5596 Median : -0.4592 Median : -0.4715 Median : -0.4715
Mean   : -0.3652 Mean   : -0.2933 Mean   : -0.2905 Mean   : -0.4578 Mean   : -0.4421
3rd Qu.:  4.3602 3rd Qu.:  4.7020 3rd Qu.:  4.7150 3rd Qu.:  4.7214 3rd Qu.:  4.7215
Max.   : 18.6940 Max.   : 24.7635 Max.   : 25.7516 Max.   : 21.4048 Max.   : 21.4048

> summary(result_b)
  b_Kasa      b_Pratt      b_Taubin      b_Landau      b_Landau_Taubin
Min.   :-25.1726 Min.   :-29.06079 Min.   :-29.41317 Min.   :-31.8270 Min.   :-31.8281
1st Qu.: -4.5882 1st Qu.: -5.43567 1st Qu.: -5.45875 1st Qu.: -4.7200 1st Qu.: -4.7200
Median :  0.2652 Median :  0.01294 Median : -0.01454 Median : -0.1929 Median : -0.1875
Mean   : -0.2635 Mean   : -0.33819 Mean   : -0.34553 Mean   : -0.2000 Mean   : -0.2122
3rd Qu.:  4.1665 3rd Qu.:  4.92798 3rd Qu.:  5.18646 3rd Qu.:  4.5828 3rd Qu.:  4.5828
Max.   : 22.5692 Max.   : 26.71973 Max.   : 27.61420 Max.   : 20.7613 Max.   : 20.7616

> summary(result_r)
  r_Kasa      r_Pratt      r_Taubin      r_Landau      r_Landau_Taubin
Min.   :45.12 Min.   :49.36 Min.   :45.71 Min.   :42.52 Min.   :42.52
1st Qu.:54.35 1st Qu.:58.77 1st Qu.:54.98 1st Qu.:52.37 1st Qu.:52.37
Median :57.35 Median :62.15 Median :57.80 Median :55.14 Median :55.14
Mean   :57.19 Mean   :62.04 Mean   :57.80 Mean   :55.00 Mean   :55.00
3rd Qu.:59.81 3rd Qu.:65.21 3rd Qu.:60.59 3rd Qu.:57.76 3rd Qu.:57.76
Max.   :67.77 Max.   :73.96 Max.   :69.14 Max.   :66.13 Max.   :66.13
```

Obrázek 10: Tabulka hodnot odhadů souřadnic středu kružnice $[a, b]$ a poloměru R jednotlivých metod pro úlohu s 25 náhodně generovanými body, kde $\varphi \in [0, 2\pi]$ a $\text{sd}(\epsilon) = 20$.

Důležité je podotknout, že se zde nejedná přímo o hodnoty našich metrik. Skutečně se koukáme na odhady parametrů kružnice. Naše první metrika se počítá jako

$$\|\cdot\|_S = \sqrt{(a - \hat{a})^2 + (b - \hat{b})^2} = \sqrt{(0 - \hat{a})^2 + (0 - \hat{b})^2} = \sqrt{\hat{a}^2 + \hat{b}^2}.$$

A po výpočtu v R jsme dostali následující výsledek pro Kasovu metodu

```

> sqrt(mean(a_Kasa)^2 + mean(b_Kasa)^2)
[1] 0.1690975
> sqrt(mean(a_Pratt)^2 + mean(b_Pratt)^2)
[1] 0.2297077
> sqrt(mean(a_Taubin)^2 + mean(b_Taubin)^2)
[1] 0.2426096
> sqrt(mean(a_Landau)^2 + mean(b_Landau)^2)
[1] 0.1969128
> sqrt(mean(a_Landau_Taubin)^2 + mean(b_Landau_Taubin)^2)
[1] 0.2314787

```

Obrázek 11: Základní charakteristiky vektoru metriky $\|\cdot\|_S$ pro dané metody.

Při výpočtu hodnoty metriky $\|\cdot\|_S$ jsme počítali pro jednoduchost s průměrnou hodnotou parametrů a, b ze všech iterací. Formálně správně by se mělo jednat o průměrnou hodnotu napočítaných metrik pro každou iteraci, ale pro přehled nám stačí hodnoty na obrázku 11. Vidíme, že Kasova metoda poměrně překvapivě poráží zbytek algebraických metod. Dokonce vychází lépe než obě varianty Landauovy metody. Landauova metoda bez počátečního řešení opět funguje lépe, ve smyslu první metriky, než Landauova metoda se zadaným počátečním řešením. Místo počítání celé metriky nám pohled na samotné hodnoty odhadovaných parametrů kružnice bude dále stačit, abychom si udělali představu o kvalitě odhadů. Byl to také jeden z důvodů, proč jsme volili střed teoretické kružnice v bodě $[0, 0]$. Další údaj, který nám pomůže porovnat zvolené metody je směrodatná odchylka odhadů těchto parametrů.

Na obrázku 12 vidíme, že Kasova metoda má nejnižší směrodatnou odchylku ze všech metod u všech tří odhadovaných parametrů. To je závěr, který bychom nejspíše ani nečekali. Proto je určitě dobré, podívat se na odhady z několika různých pohledů.

```

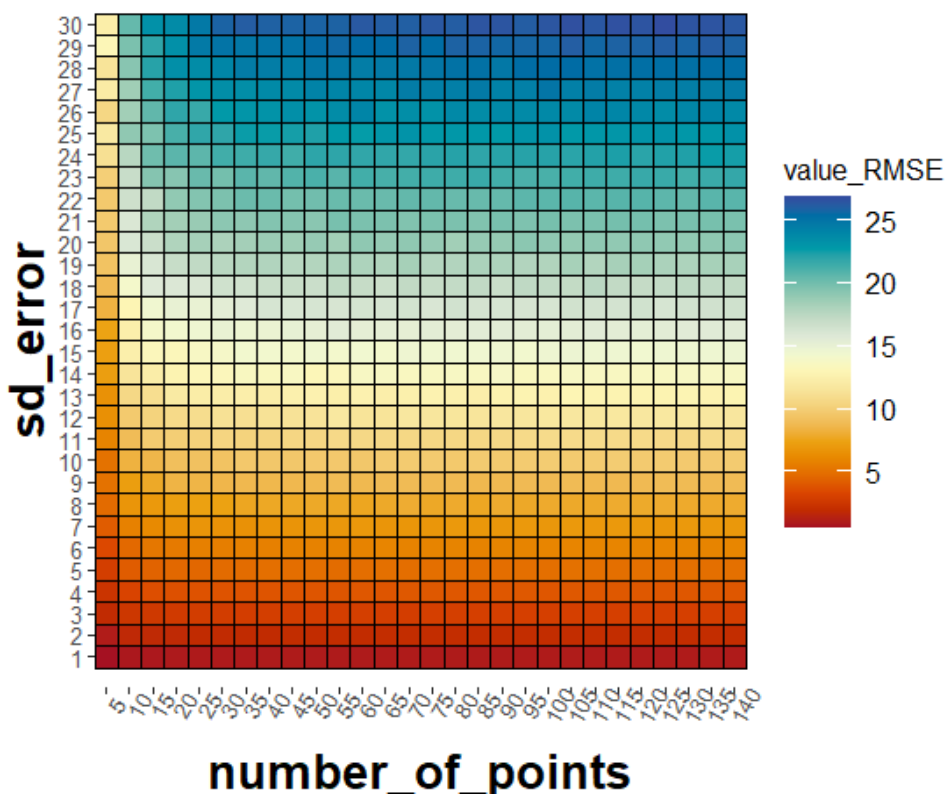
sd_Kasa
sd.a_Kasa. sd.b_Kasa. sd.r_Kasa.
 6.713901  7.463507  4.001808
sd_Taubin
sd.a_Taubin. sd.b_Taubin. sd.r_Taubin.
 7.790976   8.726476   4.233828
sd_Landau
sd.a_Landau. sd.b_Landau. sd.r_Landau.
 7.283256   8.133724   4.160903
sd_Landau_T
sd.a_Landau_Taubin. sd.b_Landau_Taubin. sd.r_Landau_Taubin.
.                7.28559                8.14132                4.165267

```

Obrázek 12: Směrodatné odchylky odhadnutých parametrů pro zvolené metody.

Otázka, na kterou bychom rádi znali odpověď, zní, jak moc záleží na směrodatné odchylce náhodné chyby a na počtu zvolených bodů? Na obrázku 13 máme

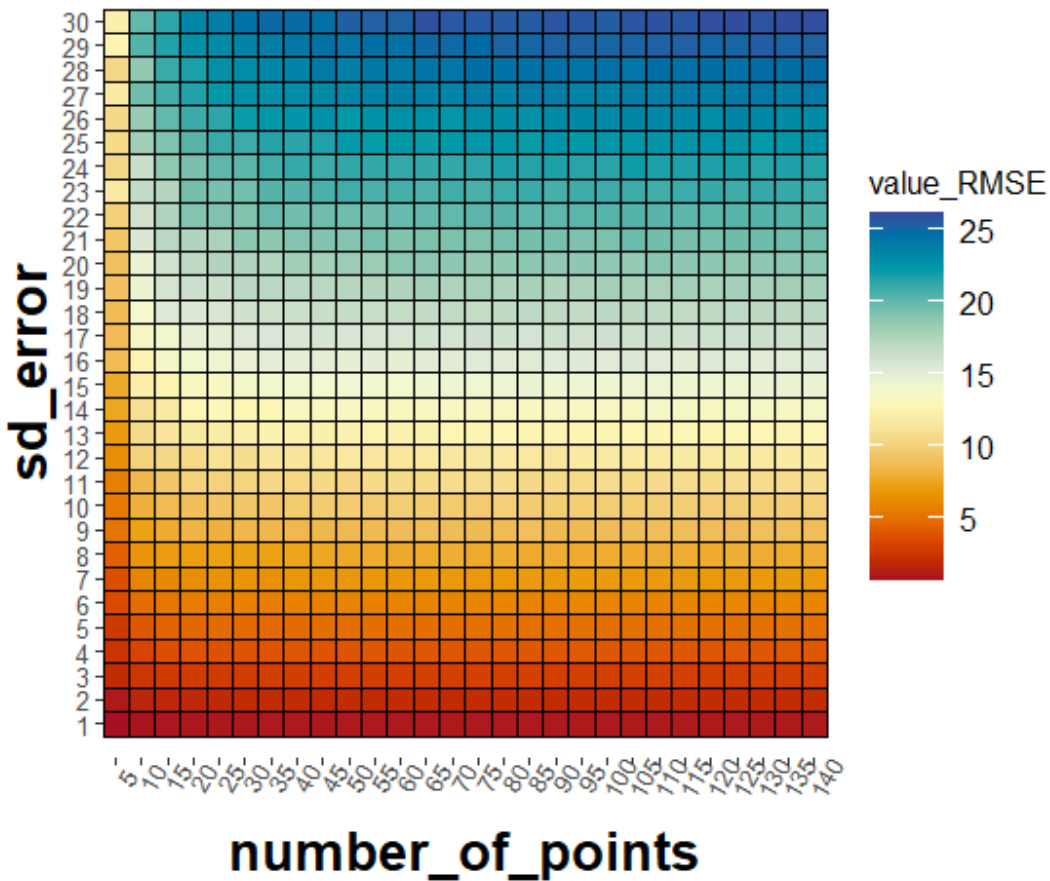
částečnou odpověď. Oba diskutované parametry jsou vyneseny na osy x, y a výsledek je vykreslen různými barvami podle své hodnoty. Jedná se o hodnotu třetí metriky, tedy chyby odhadnuté kružnice od generovaných bodů, a na škále vedle grafu vidíme, co která barva znamená.



Obrázek 13: Heatmapa pro Taubinovu metodu, kde $\varphi \in [0, 2\pi]$. Barevně je vyznačena hodnota metriky RMSE.

Pro nízkou směrodatnou odchylku chyby (<5) je skoro jedno, kolik bodů generujeme, výsledek bude vždy dobrý (ve smyslu třetí metriky). Podobné tvrzení platí pro nízký počet bodů (<5), zde je také vidět, že příliš nezáleží na velikosti směrodatné odchylky chyby a výsledek bude vždy relativně dobrý. Další závěr, který můžeme vyvodit, je zcela intuitivní, a sice, že při vyšší směrodatné odchylce náhodné chyby, se nám bude zhoršovat odhad. Pro druhý parametr nic takového na první pohled neplatí. Průměrná hodnota RMSE metriky nezáleží na počtu generovaných bodů. Je to způsobeno tím, že výsledky jsou počítány jako odhad bodů, nikoliv teoretické kružnice.

Pro úplnost zde přidáváme ještě heatmapu pro Landauovu metodu. Opět poněkud překvapivý závěr je ten, že způsob, jakým ovlivňují tyto dva parametry výsledek odhadu, je stejný pro oba typy metod, tedy algebraické i geometrické.

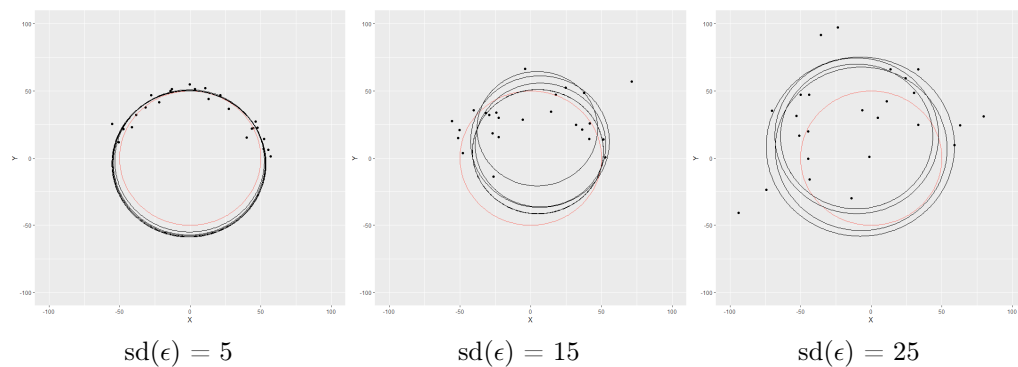


Obrázek 14: Heatmapa pro Landauovu metodu, kde $\varphi \in [0, 2\pi]$. Barevně je vyznačena hodnota metriky RMSE.

Celá úloha je postavena tak, abychom měli první pohled na aproximace jako takové. 25 bodů kolem celé kružnice je poměrně jednoduchá úloha, a tedy si s ní dobře poradily všechny metody. Zajímavější bude, co se stane při generování dat jen pod určitým úhlem.

4.2 Odhadování půlkružnice

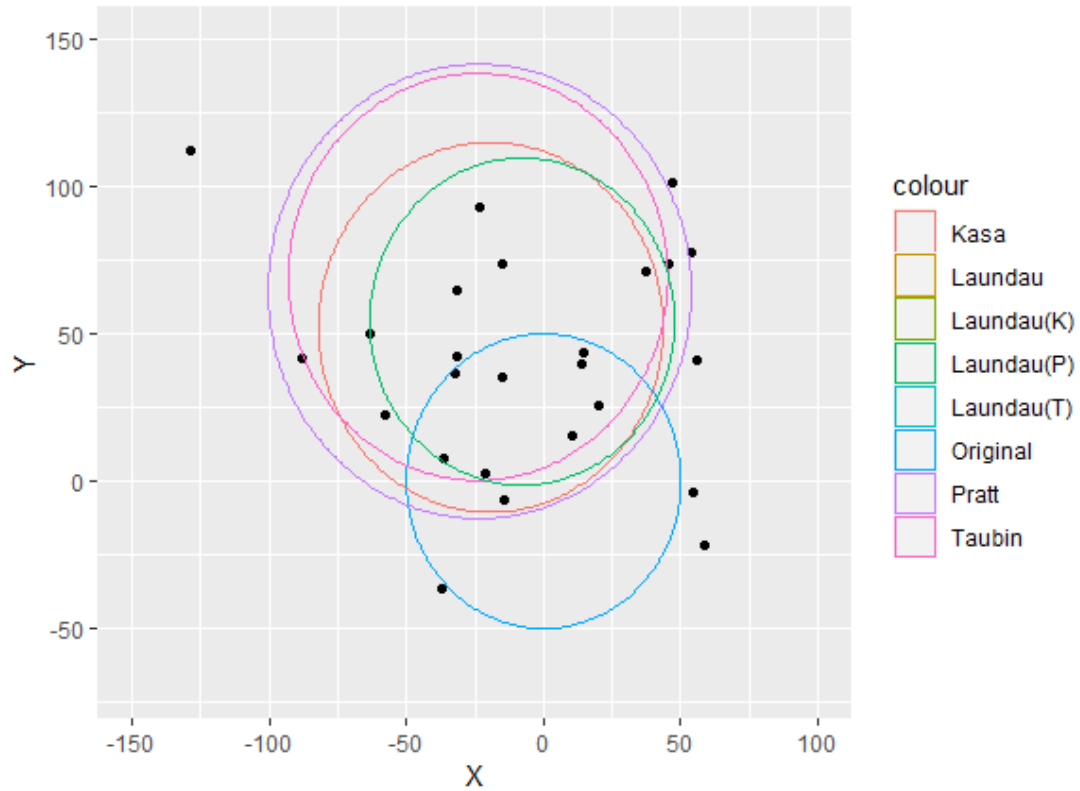
Cílem úlohy je nalézt odhady dané kružnice pomocí určených algebraických a geometrických metod aproximace. Původní kružnice se středem v bodě $S = [0, 0]$ a poloměrem $R = 50$ je reprezentována $N = 25$ body, které jsou pod úhlem $\varphi \in [0, \pi]$. Tyto body byly posunuty o náhodnou chybu ϵ . Dále platí, že $\epsilon \sim N(0, \text{sd}(\epsilon))$. Směrodatná odchylka náhodné chyby bude postupně rovna hodnotám 5, 15, 25 a 35.



Obrázek 15: Odhady daných metod pro úlohu s 25 náhodně generovanými body, kde $\varphi \in [0, \pi]$ a směrodatná odchylka náhodných chyb nabývá různých hodnot. Červeně teoretická kružnice, černě odhady daných metod

Nejprve se podíváme na výsledky pro tři různé směrodatné odchylky náhodné chyby na obrázku 15 a pak půjdeme více do detailu. Výsledek je v podstatě identický s předchozí úlohou. Pro nízkou směrodatnou odchylku máme dobré výsledky a opět jsou skoro stejné pro všechny metody. Jakmile se zvyšuje směrodatná odchylka chyby, vidíme, že si geometrické aproximace vedou lépe.

Pro případ, kdy máme $sd(\epsilon) = 35$ už vidíme i poměrně zajímavý obrázek, jak se chová, která metoda. Podle obrázku 16 vidíme, že už ani samotné datové body nepřipomínají kružnici, a tedy by se dalo očekávat, že se výsledky různých metod budou velmi lišit.



Obrázek 16: Odhady daných metod pro úlohu s 25 náhodně generovanými body, kde $\varphi \in [0, \pi]$ a $\text{sd}(\epsilon) = 35$.

Když půjdeme do detailu tohoto konkrétního příkladu, je zřejmé, že geometrické aproximace fungují o trochu lépe. Může nastat situace, kdy ani geometrická metoda pro velký rozptyl nemusí fungovat. Na obrázku vidíme, že jsou všechny odhady špatné. Použité metody odhadují body poměrně daleko od teoretické kružnice, je to způsobeno právě vysokou směrodatnou odchylkou pro náhodné chyby.


```

> summary(Original_RMSE)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 13.20  22.19  24.41  24.62  27.17  36.75
> summary(Algebraic_RMSE)
  Kasa_RMSE      Pratt_RMSE      Taubin_RMSE
Min.   :13.36  Min.   :14.20  Min.   :13.43
1st Qu.:23.43  1st Qu.:24.92  1st Qu.:23.07
Median :26.26  Median :28.28  Median :25.72
Mean   :26.30  Mean   :28.39  Mean   :25.97
3rd Qu.:29.14  3rd Qu.:31.53  3rd Qu.:28.55
Max.   :40.78  Max.   :47.61  Max.   :42.00
> summary(Geometric_RMSE)
  Landau_RMSE  Landau_Kasa_RMSE  Landau_Pratt_RMSE  Landau_Taubin_RMSE
Min.   :13.20  Min.   :13.20  Min.   :13.20  Min.   :13.20
1st Qu.:22.19  1st Qu.:22.19  1st Qu.:22.19  1st Qu.:22.19
Median :24.46  Median :24.44  Median :24.39  Median :24.41
Mean   :24.72  Mean   :24.65  Mean   :24.59  Mean   :24.62
3rd Qu.:27.27  3rd Qu.:27.12  3rd Qu.:27.11  3rd Qu.:27.17
Max.   :35.14  Max.   :35.29  Max.   :35.37  Max.   :36.75

```

Obrázek 17: Tabulka hodnot RMSE jednotlivých metod pro úlohu s 25 náhodně generovanými body, kde $\varphi \in [0, \pi]$, $\text{sd}(\epsilon) = 35$.

Jak jsme ale již zmínili, nemůžeme brát hodnotu třetí metriky jako dogmatické označení lepšího a horšího odhadu. Je potřeba tuto hodnotu dát do souvislosti s parametry odhadnuté kružnice.

```

> summary(result_a)
  a_Kasa      a_Pratt      a_Taubin      a_Landau      a_Landau_Taubin
Min.   :-45.2110  Min.   : -72.812  Min.   : -136.722  Min.   : -72.8555  Min.   : -129.701
1st Qu.: -8.8715  1st Qu.: -10.477  1st Qu.: -10.823  1st Qu.: -9.7401  1st Qu.: -9.802
Median : -0.5306  Median : -0.935  Median : -1.246  Median : -0.9556  Median : -0.983
Mean   : -0.8241  Mean   :  9.232  Mean   : 12.929  Mean   : -0.6250  Mean   : 12.938
3rd Qu.:  7.3467  3rd Qu.:  9.255  3rd Qu.: 10.824  3rd Qu.:  9.9460  3rd Qu.: 10.751
Max.   : 34.8840  Max.   :4850.522  Max.   :6443.115  Max.   : 39.2980  Max.   :6443.115
> summary(result_b)
  b_Kasa      b_Pratt      b_Taubin      b_Landau      b_Landau_Taubin
Min.   : -14.71  Min.   : -305.54  Min.   : -1005.334  Min.   : -57.84  Min.   : -1004.67
1st Qu.: 19.53  1st Qu.:  11.32  1st Qu.:  7.479  1st Qu.: 13.36  1st Qu.:  9.86
Median : 28.16  Median :  25.39  Median :  24.161  Median : 26.49  Median :  25.59
Mean   : 28.10  Mean   :  68.19  Mean   :  80.263  Mean   : 25.83  Mean   :  80.62
3rd Qu.: 37.26  3rd Qu.:  40.27  3rd Qu.:  40.935  3rd Qu.: 38.42  3rd Qu.:  39.30
Max.   : 80.65  Max.   :22755.50  Max.   :30189.333  Max.   :107.98  Max.   :30189.33
> summary(result_r)
  r_Kasa      r_Pratt      r_Taubin      r_Landau      r_Landau_Taubin
Min.   :49.60  Min.   :  55.10  Min.   :  50.41  Min.   : 45.06  Min.   :  45.06
1st Qu.:59.06  1st Qu.:  68.22  1st Qu.:  61.27  1st Qu.: 55.55  1st Qu.:  56.58
Median :63.73  Median :  74.11  Median :  67.24  Median : 60.22  Median :  61.53
Mean   :63.62  Mean   : 123.64  Mean   : 138.04  Mean   : 61.71  Mean   : 133.53
3rd Qu.:67.64  3rd Qu.:  80.10  3rd Qu.:  73.31  3rd Qu.: 65.00  3rd Qu.:  67.69
Max.   :83.70  Max.   :23237.66  Max.   :30840.12  Max.   :115.33  Max.   :30840.12

```

Obrázek 18: Tabulka hodnot odhadů souřadnic středu kružnice $[a, b]$ a poloměru R jednotlivých metod pro úlohu s 25 náhodně generovanými body, kde $\varphi \in [0, \pi]$, $\text{sd}(\epsilon) = 35$.

Na první pohled je zajímavé, že všechny metody se rozhodly použít větší poloměr kružnice, než má kružnice původní. Pokud budeme hledět opět na hodnotu

průměru a mediánu, vidíme, že střed odhadovaných kružnic má velmi dobře odhadnutou x -ovou souřadnici, ale to samé neplatí pro souřadnici y -ovou. V případě, že bychom generovali data pro úhel $\varphi \in [\frac{\pi}{2}, \frac{3\pi}{2}]$, dostali bychom výsledek obrácený. Tedy dobrý odhad y -ové souřadnice, ale výrazně horší odhad x -ové souřadnice. A opět platí, že středy odhadovaných kružnic pro různé metody jsou blízko sebe. To je trend, který nastává u většiny úloh, co počítáme. Když už je střed špatně odhadnutý, tak je odhadnutý různými metodami „stejně špatně“. Zde bych se chtěl pozastavit nad některými extrémními hodnotami v tabulce. Když se zaměříme na hodnoty *Max*, *Min*, vidíme některé opravdu výrazně odchylené hodnoty. Obrázek 16 je pouze jednou ze zmíněných 500 simulací, ze kterých se počítají naše metricky. Protože směrodatná odchylka náhodné chyby rovna 35 už je opravdu velká v poměru s poloměrem původní kružnice, může se stát, že některé simulace nám „uletí“ a budou vypadat úplně jinak. Proto pak mohou nastat takové extrémní hodnoty. Proto je zde opět nutné napočítat směrodatné odchylky pro odhadované parametry kružnice. V obrázku 19 stojí za povšimnutí hodnoty u Taubinovy metody a Landauovy metody, ve které jsme tuto Taubinovu použili jako počáteční řešení. Vidíme, že špatné výsledky Taubinovy metody táhnou velmi silně výsledky Landauovy metody dolů. To je přesně případ, o kterém jsme hovořili v teoretické části, kdy špatně zvolené počáteční řešení geometrických metod znamená špatný celkový výsledek.

```

sd_Kasa
sd.a_Kasa. sd.b_Kasa. sd.r_Kasa.
  12.54924   14.28528   6.591827
sd_Taubin
sd.a_Taubin. sd.b_Taubin. sd.r_Taubin.
  21.25337   244.662   237.1639
sd_Landau
sd.a_Landau. sd.b_Landau. sd.r_Landau.
  14.63799   23.75294   9.873794
sd_Landau_T
sd.a_Landau_Taubin. sd.b_Landau_Taubin. sd.r_Landau_Taubin.
  20.91718   245.042   237.6262

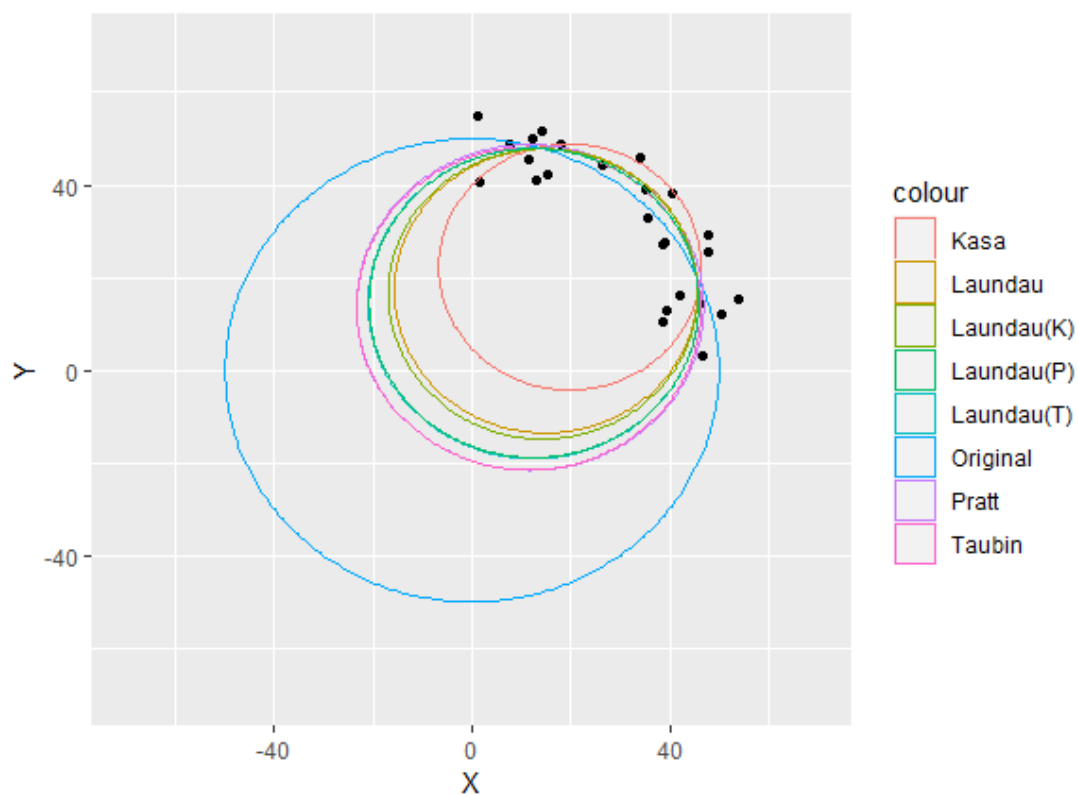
```

Obrázek 19: Směrodatná odchylka odhadnutých parametrů pro zvolené metody.

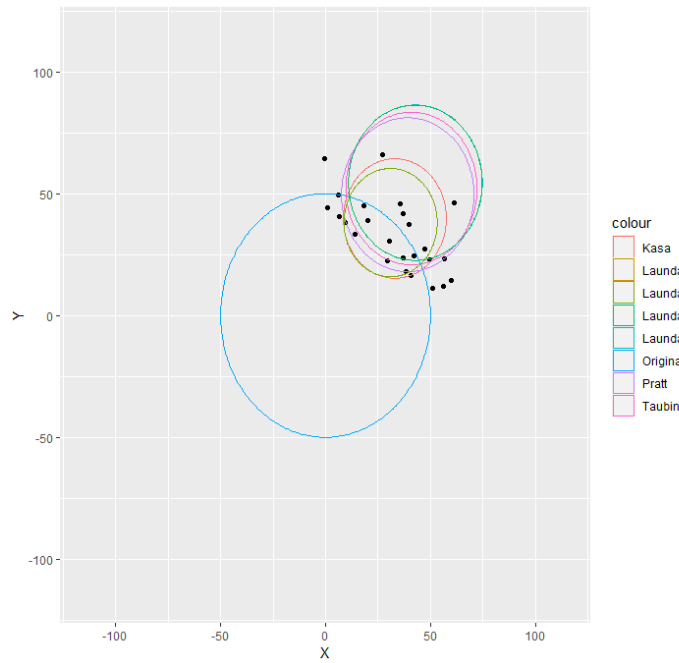
4.3 Odhadování čtvrtkružnice

Cílem úlohy je nalézt odhady dané kružnice pomocí určených algebraických a geometrických metod aproximace. Původní kružnice se středem v bodě $S = [0, 0]$ a poloměrem $R = 50$ je reprezentována $N = 25$ body, které jsou pod úhlem $\varphi \in [0, \frac{\pi}{2}]$. Tyto body byly posunuty o náhodnou chybu ϵ . Dále platí, že $\epsilon \sim N(0, sd(\epsilon))$. Směrodatná odchylka náhodné chyby bude postupně rovna hodnotám 5, 10 a 15.

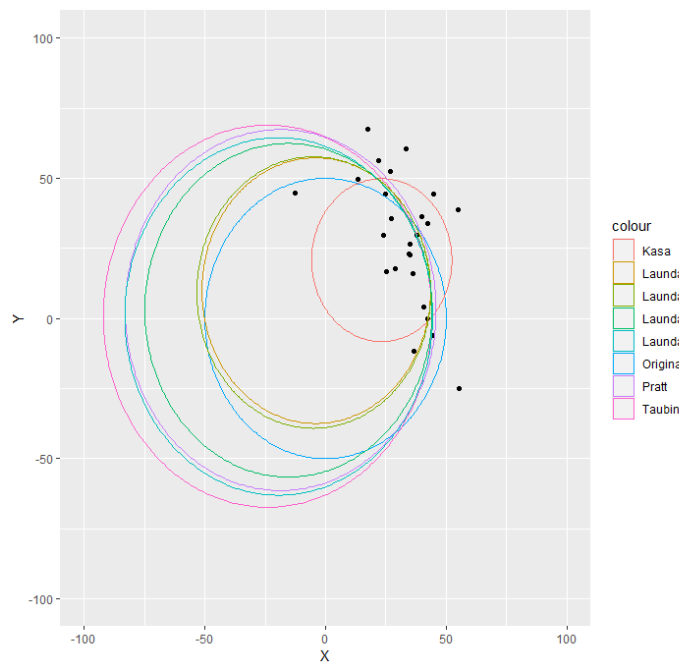
V této úloze potvrdíme vyšší kvalitu geometrických aproximací jednou pro vždy. V těchto úlohách jsou všechny obrázky hezky symetrické. Jen se nám pro některé metody zvětšuje průměr kružnice a úměrně posouvá střed. Ale všechny metody mají (opět) stejný „nápad“.



Obrázek 20: Odhady daných metod pro úlohu s 25 náhodně generovanými body, kde $\varphi \in [0, \frac{\pi}{2}]$ a $\text{sd}(\epsilon) = 5$.



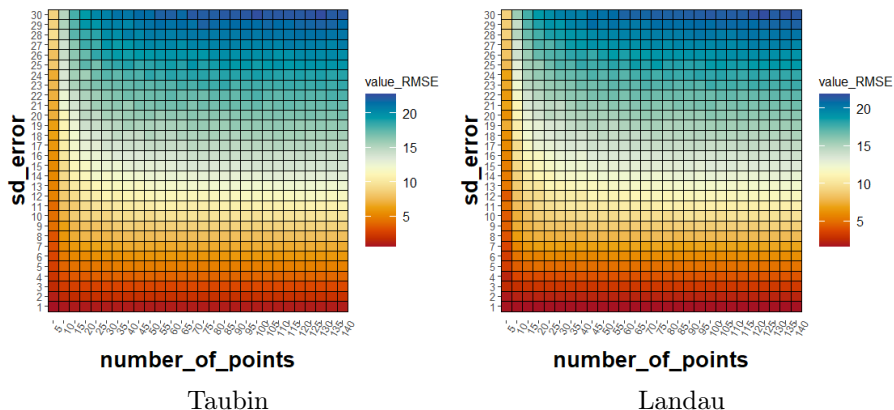
Obrázek 21: Odhady daných metod pro úlohu s 25 náhodně generovanými body, kde $\varphi \in [0, \frac{\pi}{2}]$ a $\text{sd}(\epsilon) = 10$.



Obrázek 22: Odhady daných metod pro úlohu s 25 náhodně generovanými body, kde $\varphi \in [0, \frac{\pi}{2}]$ a $\text{sd}(\epsilon) = 15$.

Jakmile se věnujeme pouze takhle malému úhlu z kružnice, výsledky se liší případ od případu a je prakticky nemožné udělat obecný závěr. Zde tedy pro představu ukážeme, jak mohou jednotlivé úlohy vypadat v závislosti na směrodatné odchylce náhodné chyby generovaných bodů.

Zůstaneme ještě u $\varphi \in [0, \frac{\pi}{2}]$ a podíváme se opět na heatmapu pro dvě metody.



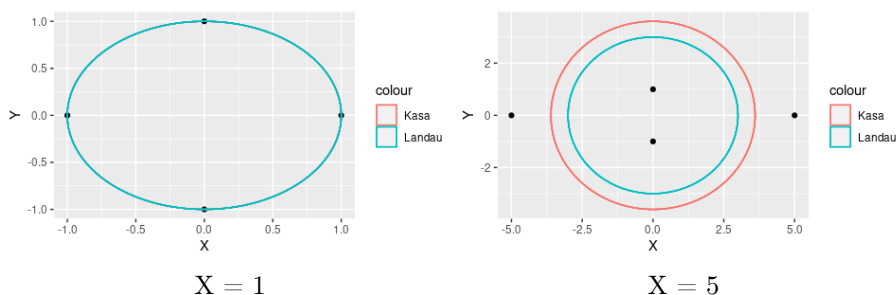
Obrázek 23: Heatmapy pro Taubinovu a Landauovu metodu, kde $\varphi \in [0, \frac{\pi}{2}]$. Barevně je vyznačena hodnota metriky RMSE.

I v případě malého oblouku kružnice se obě metody chovají prakticky stejně. Závislost výsledné hodnoty třetí metriky na počtu generovaných bodů a směrodatné odchylce jejich chyby vypadá stejná v obou případech. Je to stejný závěr, jaký jsme dostali u odhadu celé kružnice. Tady rozdíl mezi metodami skutečně vidět není. Museli bychom se opět věnovat mnohem složitějším úlohám.

4.4 Speciální úlohy

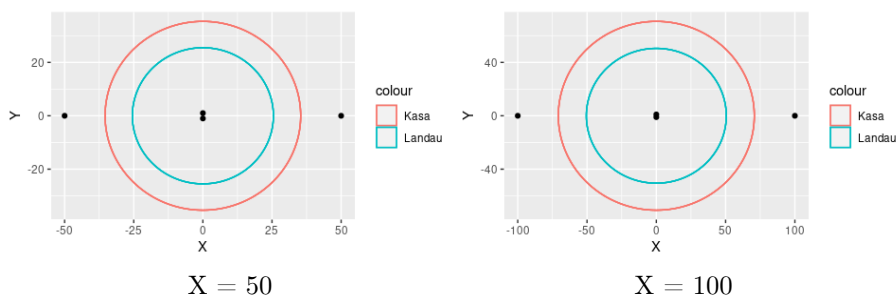
Nyní se podíváme na dva speciální příklady, o kterých jsme nejdříve mluvili v teoretické části. První z nich bude skupina 4 bodů.

Cílem úlohy je nalézt odhady bodů o souřadnicích $(-X, 0)$, $(X, 0)$, $(0, 1)$, $(0, -1)$, kde $X \gg 1$. Tyto body byly opět posunuty o náhodnou chybu. Pro tu platí $\epsilon \sim N(0, \text{sd}(\epsilon))$, kde X bude postupně nabývat hodnot 1, 5, 50 a 100 a $\text{sd}(\epsilon)$ bude postupně nabývat hodnot 3 a 10.



Obrázek 24: Odhady Kasovy a Landauovy metody pro body $(X, 0)$, $(-X, 0)$, $(0, 1)$, $(0, -1)$ a směrodatná odchylka náhodných chyb rovna 1 a 5.

Nejprve se podíváme, jak vypadají odhady přímo pro zadané body. Vidíme, že pro $X = 1$ (což není přímo součástí této úlohy, neboť chceme splnit podmínku $X \gg 1$) odhadnou obě úlohy velmi spolehlivě a rychle správnou kružnici, která odhaduje body s nulovou chybou. Se zvětšujícím se X pochopitelně chyba narůstá. Vidíme zde zajímavý trend, Landauova metoda odhaduje tyto body kružnicí s menším poloměrem než Taubinova metoda. V obou případech v obrázku 24 bude střed kružnice pro všechny odhady stejný. Pokud půjdeme s X do vyšších čísel, budeme dostávat pokaždé už stejný obrázek, jen v jiném měřítku, viz obrázek 25.

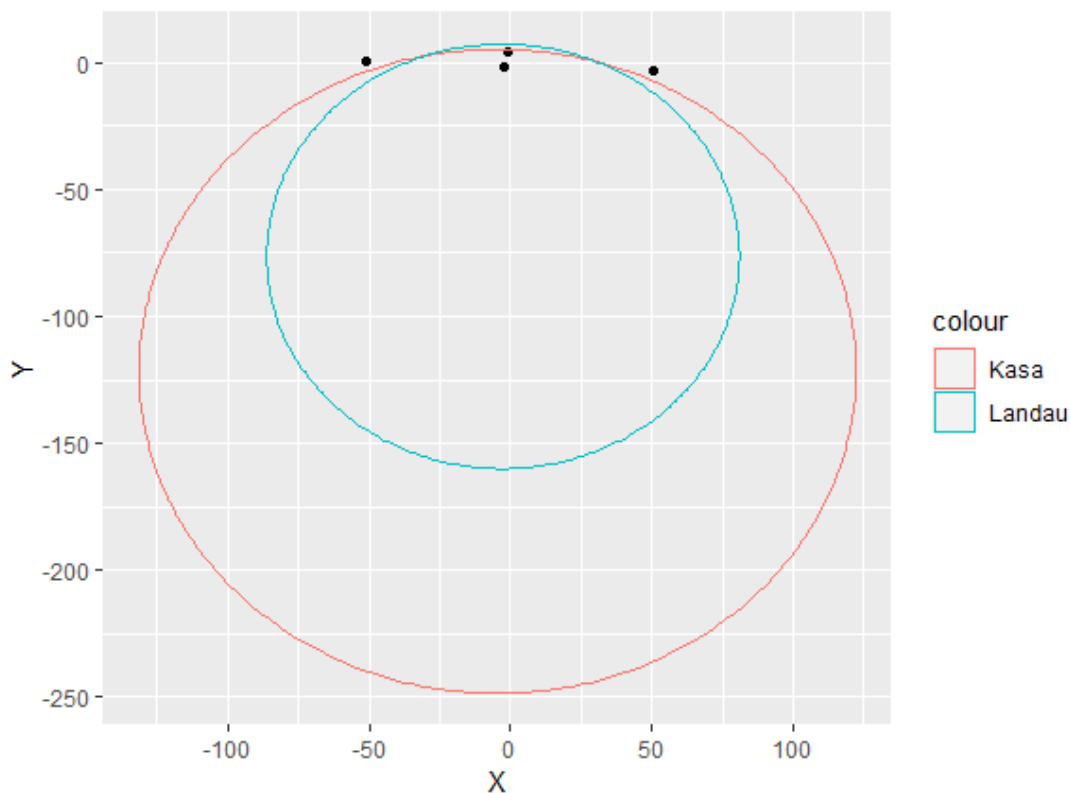


Obrázek 25: Odhady Kasovy a Landauovy metody pro body $(X, 0)$, $(-X, 0)$, $(0, 1)$, $(0, -1)$ a směrodatná odchylka náhodných chyb rovna 50 a 100.

Zde neodhadujeme žádný původní objekt posunutých bodů, ale samotné body. Proto je vzdálenost bodů od odhadnuté kružnice klíčový ukazatel kvality odhadu. A zde narážíme na výsledek, který bychom možná nečekali. Kasova metoda zde vychází lépe. Připomeňme, že Kasova metoda je jedna z nejslabších algebraických metod. A algebraické metody vychází o něco hůře, než metody geometrické.

V teoretické části jsme hovořili o tomto případě v souvislosti s jednoznačností řešení úlohy. Když totiž posuneme naše body o náhodnou chybu, dostaneme

se do situace, kdy určené metody začnou odhadovat generované body pomocí kružnic s velkým poloměrem. Zafixujeme tedy hodnotu pro $X = 50$ v zadání úlohy a podíváme se, co to udělá s řešeními pro různé směrodatné odchylky náhodné chyby.



Obrázek 26: Odhady daných metod pro body $(X, 0)$, $(-X, 0)$, $(0, 1)$, $(0, -1)$, kde $\text{sd}(\epsilon) = 3$.

V obrázku 26 vidíme, že již při volbě $\text{sd}(\epsilon) = 3$ dostáváme přesně výsledek, který jsme podle teorie měli dostat. V případě, že by obě kružnice měly své středy „na druhé straně“ generovaných bodů, celková kvalita odhadu bude velmi podobná. V této úloze záleží čistě na tom, kterou ze simulací vybereme pro obrázek, ale velmi snadno se mohly obě naše kružnice nacházet „nad“ generovanými body.

```

> summary(results_Kasa)
  Kasa_RMSE      Kasa_R
Min.   :0.001614  Min.   : 137.5
1st Qu.:0.513240  1st Qu.: 364.5
Median :1.109817  Median : 627.1
Mean   :1.306554  Mean   : 3176.2
3rd Qu.:1.930728  3rd Qu.: 1448.8
Max.   :4.591501  Max.   :434627.6
> summary(results_Landau)
  Landau_RMSE      Landa_R
Min.   : 3.800    Min.   :49.56
1st Qu.: 6.085    1st Qu.:79.53
Median : 6.705    Median :82.17
Mean   : 6.790    Mean   :82.32
3rd Qu.: 7.494    3rd Qu.:85.32
Max.   :11.669    Max.   :94.92

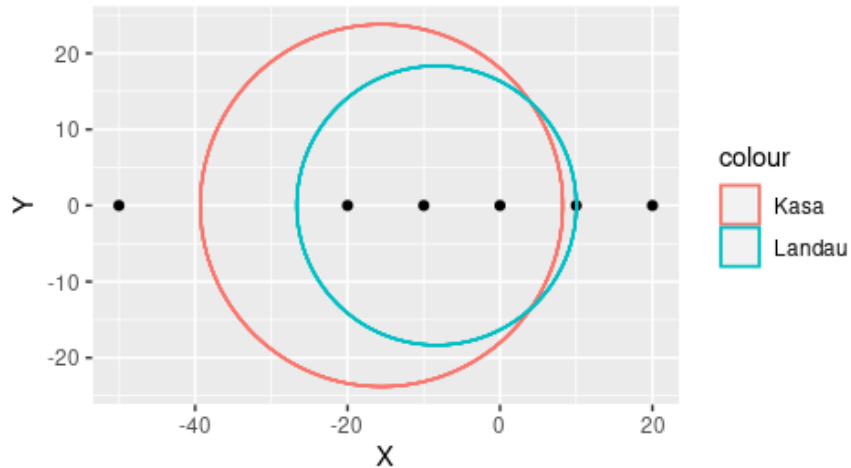
```

Obrázek 27: Výsledky daných metod pro kolineární body, kde $\text{sd}(\epsilon) = 3$.

Opět vidíme v tabulce na obrázku 27, že Kasova metoda používá pro svůj odhad kružnici s výrazně větším poloměrem. Velmi zajímavým výsledkem, který stojí za povšimnutí, je mnohonásobně lepší výsledek Kasovy metody ve smyslu RMSE. Kdybychom chtěli úlohu řešit, ještě pro větší hodnotu X , výsledky by byly analogické. Lišilo by se jen měřítko. Opět by Landauova metoda měla výrazně menší kružnici a Kasova metoda by odhadovala body pomocí kružnice s velkým poloměrem. Z důvodu přehlednosti obrázku jsme se uchýlili k hodnotě $X = 50$.

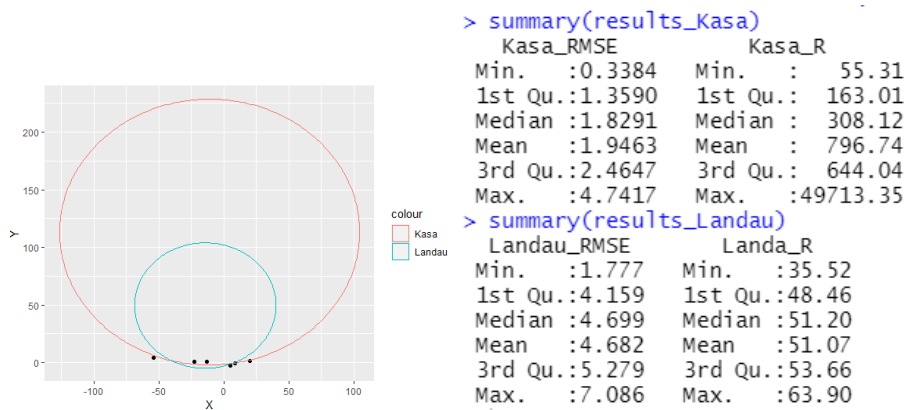
Přejdeme nyní na poslední úlohu celé práce. V teoretické části jsme opakovaně hovořili o příkladu, kdy máme několik kolineárních bodů a chceme je odhadovat. Cílem úlohy je nalézt odhady kružnicí kolineárních bodů. Podíváme se nejprve, jak by úloha vypadala pro čistě kolineární body, tedy bez posunutí o náhodnou chybu. A poté tyto body opět posuneme o náhodnou chybu. Pro tu platí $\epsilon \sim N(0, \text{sd}(\epsilon))$, kde $\text{sd}(\epsilon)$ bude postupně nabývat hodnot 3 a 10.

Na obrázku 28 je znázorněn výsledek takové simulace. Je zřejmé, že nejlepším odhadem těchto bodů by byla samotná osa X . V případě odhadů kuželosečkami dostáváme velmi podobné kružnice pro obě zvolené metody. Opět se liší hlavně ve velikosti poloměru, kdy Kasova metoda volila větší kružnici. A proto RMSE hodnota pro Landauovu metodu vyjde dokonce $3x$ nižší.



Obrázek 28: Odhady daných metod pro kolineární body, kde $sd(\epsilon) = 3$.

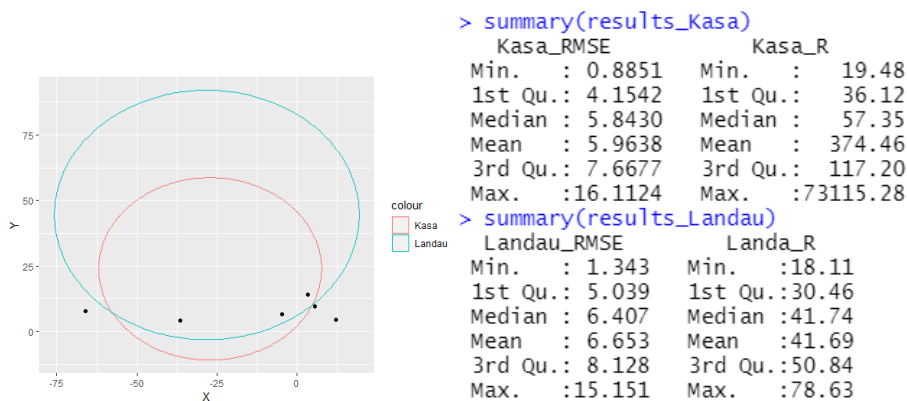
Co se ale stane, posuneme-li generované body o náhodnou chybu? Nejprve se podíváme na simulaci, kde směrodatná odchylka náhodné chyby je rovna 3.



Obrázek 29: Odhady a výsledky daných metod pro kolineární body, kde $sd(\epsilon) = 3$.

Z výsledků metriky RMSE vidíme, že jsou si oba odhady velmi podobné. Na řádce *Min* je vidět, že Kasova metoda si v některých iteracích počínala o mnoho lépe, ale řádky *Mean*, *Median* hovoří o velké podobnosti výsledků obou metod. Dokonce je vidět, že ani poloměry odhadnutých kružnic se tolik neliší. Z extrémních hodnot vidíme, že typicky byl poloměr Kasovy odhadnuté kružnice vyšší.

Na dalším obrázku máme již poslední simulaci v této práci. Jedná se o stejnou úlohu jako v předchozím případě (tedy několik kolineárních bodů), ale tentokrát je hodnota směrodatné odchylky náhodné chyby rovna 10.



Obrázek 30: Odhady a výsledky daných metod pro kolineární body, kde $\text{sd}(\epsilon) = 10$.

Určené metody nám opět odhadly generované body pomocí kružnic s relativně velikým poloměrem. Tentokrát to ale podle tabulky vypadá, že Landauova metoda odhaduje body pomocí větších kružnic. Když se pozorně podíváme na tabulku RMSE metriky vidíme, že Landauova metoda funguje z hlediska této metriky o mnoho lépe.

V takových úlohách, jako jsme právě řešili, nám příliš nezáleží na tom, z jaké strany máme výsledný odhad. Problém je ale v praktických úlohách, kdy se nám může snadno stát, že přidáme jeden nebo dva nové body. V takovém případě už bude poznat, zda jsme zvolili kružnici na správné straně. V případě, že tomu tak není, se dostaneme i k velmi špatným výsledkům, které neodpovídají realitě. Je tedy skutečně potřeba dobře znát úlohu, kterou máme řešit, a co znamenají dvě velmi podobná řešení.

4.5 Souhrn

Půjdeme-li poctivě do detailu jednotlivých tabulek, dostaneme závěr, který nás nijak nepřekvapí. I v úlohách, kde vyjdou velmi dobré odhady, jsou stále geometrické aproximace o malinko přesnější. Je to závěr, který nám potvrzuje vyřčenou teorii v předchozích kapitolách. Jediný problém je s časovou náročností při složitějších úlohách. Je tedy potřeba vhodně volit metodu, podle toho jak rychlé a jak přesné výsledky potřebujeme.

Chceme-li se podívat na časovou náročnost jednotlivých metod a porovnat je, potřebujeme složitější úlohy. V našem případě, kdy se snažíme odhadnout několik málo bodů, porovnáváme velmi dobré a rychlé výpočty. Hlavní rozdíl je ale dobře viditelný i v těchto úlohách. Celkový čas pro našich 500 iterací mají geometrické úlohy přibližně dvojnásobný. Odhadnout 25 bodů generovaných podél celé kružnice se směrodatnou odchylkou chyby rovnou 5 trvá Kasově metodě 0,3

s, zatímco Landauově metodě to trvá 0,75 s. Taubinova a Prattova metoda dokonce počítají rychleji než metoda Kasova. Tento výsledek je zarážející, protože jsme řekli, že Kasova metoda je nejjednodušší a její velká výhoda je výpočetní rychlost. Nicméně je potřeba říci, že jestli výpočet trvá 0,3 s nebo 0,25 s pro nás není signifikantním rozdílem, a můžeme tedy bez problému používat Prattovu či Taubinovu metodu, která má lepší výsledky, i když je trochu pomalejší. Ve složitějších výpočtech už bychom potvrdili skutečnost z teoretické části a sice, že Kasova metoda je nejrychlejší. Bohužel jsme pro takové úlohy v této práci neměli prostor.⁵

V úlohách, které jsme řešili v kapitole 4, nám na rychlosti nezáleží. Pro nízkou hodnotu směrodatné odchylky náhodné chyby vycházejí všechny metody velmi podobně, a proto bychom volili metody algebraické. Pro úlohy s vyšší směrodatnou odchylkou náhodné chyby již vycházejí geometrické aproximace lépe a v našem případě nás pomalejší metoda nezdržuje.

Nejzajímavější jsou úlohy v podkapitole 4.4. Napočítané simulace nám potvrdily, že v obou diskutovaných příkladech nám naše metody odhadu kuželosečkami selhávají. V teoretické části jsme prohlásili, že nejlepší možná aproximace je v obou případech lineární. A že žádná z metod ji nedokáže porazit. Přestože se jedná o příklady, na které v praxi typicky nenarazíme, uvedli jsme je zde alespoň pro zajímavost. Pokud se tedy budeme v praxi rozhodovat, jakou metodu použít, neměli bychom se nechat ovlivnit závěry z těchto dvou úloh. Příklady, uvedené zde jsou skutečně jen ilustrační a pravděpodobnost jejich výskytu při reálném výzkumu je minimální. Co se týče časové náročnosti u této úlohy, není zde praktický žádný rozdíl. Obě simulace jsme provedli 500x, ale jelikož se jedná o malý počet bodů a relativně jednoduchou úlohu, dostáváme výsledek velmi svižně, v řádu setin sekundy. Nejdelší z celého procesu bylo kreslení grafů pomocí funkce *ggplot*.

Důležité je zdůraznit, že u příkladů, ve kterých odhadujeme pouze část kružnice, například pro úhel $\varphi \in [0, \frac{\pi}{2}]$, ale i pro mnohem menší části kružnice, se dostáváme do problémů. V takových případech se odhady různých metod značně liší. Je tedy potřeba dávat si na to pozor a hlavně vědět, jaký poměr kvality a rychlosti hledáme.

Podobný závěr plyne pro úlohy, kde odhadujeme větší počet bodů. Pokud máme úlohu, kde odhadujeme 3 body, je zřejmé, že budou výsledky dobré. Ale pokud se nám zvyšuje počet generovaných bodů, typicky se nám také zvyšuje chybovost.

⁵Výpočty byly prováděny na notebooku značky Acer s 64bitovým operačním systémem s pamětí RAM 8GB a s procesorem Intel(R) Core(TM) i5-7200U CPU 2.50GHz 2.71 GHz. Časové údaje jsou uvedeny pro celý výpočet, tedy pro 500 iterací v každé simulaci.

5 Závěr

Seznámili jsme se s metodami odhadu kuželoseček a kvadrik. Ukázali jsme, kdy a proč celý nápad vznikl a jaké zajímavé praktické využití tyto metody mohou mít. Rozebrali jsme teorii kolem lineární i nelineární regrese a řekli si, jak pracovat s úlohou, kde máme vysvětlovanou i vysvětlující proměnnou zatíženou chybou. Podívali jsme se na různé parametrizace pro kružnici a kouli a naznačili, k čemu je jaká parametrizace dobrá a v čem má slabiny. Na závěr teoretické části jsme si představili konkrétní metody určené k odhadování kuželoseček a předběžně naznačili, kde bude vhodné použít jakou z nich.

V praktické části jsme provedli několik simulací, ve kterých jsme využili poznatků z teoretické části práce a přesvědčili jsme se, že vyřčené skutečnosti opravdu platí. Simulovali jsme nejprve body kolem celé teoretické kružnice, poté jsme se podívali pouze na některé části oblouku a opět je odhadovali pomocí uvedených metod. A na závěr jsme se podívali na speciální příklady, které měly velmi zajímavé a neobvyklé výsledky.

Výsledky veškerých simulací jsou doplněny o obrázky a tabulky výsledků. Závěrem celé práce je srovnání metod pro různé typy úloh. Výsledky jsou řádně okomentovány ve čtvrté kapitole. Zjistili jsme, že v jednoduchých úlohách, příliš nebude záležet na volbě metody pro odhad kružnice. Rozdíl mezi metodami ukáží až složitější úlohy, jako je například odhadování trajektorií částic vzniklých v urychlovači částic. Tam již bude záležet, na každém zlomku vteřiny, a tedy se budeme chtít spíše přiklonit k algebraickým metodám. Zatím co v úlohách například z medicínského prostředí, kde se nám jedná primárně o přesnost výsledků, bychom sáhli spíše po geometrické aproximační metodě.

Reference

- [1] Kunderová P., Hron K., *Základy počtu pravděpodobnosti a metod matematické statistiky*, 2. dopl. vydání. Olomouc: Univerzita Palackého v Olomouci, (2015). Skripta.
- [2] Adcock R. J., *A Problem in Least Squares*, The Analyst, vol. 5, no. 2, (1878), pp. 53–54.
- [3] Chernov N., *Circular and linear regression: fitting circles and lines by least squares*, Taylor & Francis, (2011).
- [4] Anderson T. W., *Estimation of linear functional relationships: Approximate distributions and connections with simultaneous equations in econometrics.*, Journal of the Royal Statistical Society: Series B (Methodological) 38.1 (1976): 1-20.
- [5] Anderson T. W., *Estimating linear statistical relationships*, The Annals of Statistics 12.1 (1984): 1-45.
- [6] Anderson, T. W., Rubin H., *Statistical inference in*, Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1., University of California Press, (1956).
- [7] Biggerstaff R. H., *Three variations in dental arch form estimated by a quadratic equation*, Journal of Dental Research, Volume 51 Issue 5, (1972).
- [8] Carroll R. J., Ruppert D., Stefansky L. A., Crainiceanu C.M., *Measurement error in nonlinear models: a modern perspective*, Chapman and Hall/CRC, (2006).
- [9] Chernov N., Lesort C., *Least squares fitting of circles*, Journal of Mathematical Imaging and Vision 23.3 (2005): 239-252.
- [10] Chernov N., Sapirstein P. N., *Fitting circles to data with correlated noise*, Computational statistics & data analysis 52.12 (2008): 5328-5337.
- [11] Freeman P. R. *Thom's survey of the Avebury ring*, Journal for the History of Astronomy 8.2 (1977): 134-136.
- [12] Halir R., Flusser J., *Numerically stable direct least squares fitting of ellipses*, Proc. 6th International Conference in Central Europe on Computer Graphics and Visualization. WSCG. Vol. 98. Citeseer, (1998).
- [13] Kummell C. H., *Reduction of observation equations which contain more than one observed quantity*, The Analyst (1879): 97-105.

- [14] Kunitomo N., *Asymptotic expansions of the distributions of estimators in a linear functional relationship and simultaneous equations*, Journal of the American Statistical Association 75.371 (1980): 693-700.
- [15] Madansky A., *The fitting of straight lines when both variables are subject to error*, Journal of the American Statistical Association 54.285 (1959): 173-205.
- [16] Ososkov G. A., *JINR technical report P10-83-187*, Dubna, (1983).
- [17] Patefield W. M., *On the validity of approximate distributions arising in fitting a linear functional relationship*, Journal of Statistical Computation and Simulation 5.1 (1976): 43-60.
- [18] Pratt V., *Direct least-squares fitting of algebraic surfaces*, ACM SIGGRAPH computer graphics 21.4 (1987): 145-152.
- [19] Solari M. E., *The "maximum likelihood solution" of the problem of estimating a linear functional relationship*, Journal of the Royal Statistical Society: Series B (Methodological) 31.2 (1969): 372-375.
- [20] Wald A., *The fitting of straight lines if both variables are subject to error*, The Annals of Mathematical Statistics 11.3 (1940): 284-300.
- [21] Whalen M. E., *Ceramic vessel size estimation from sherds: An experiment and a case study*, Journal of Field Archaeology 25.2 (1998): 219-227.
- [22] Wolter K. M., Fuller W. A., *Estimation of Nonlinear Errors-in-Variables Models*, Ann. Statist. 10 (2) 539 - 548, June, (1982).
- [23] Chernov N., Lesort C., *Statistical efficiency of curve fitting algorithms* Computational statistics & data analysis 47.4 (2004): 713-728.
- [24] Al-Sharadqah A., Chernov N., *Error analysis for circle fitting algorithms*, Electronic Journal of Statistics 3 (2009): 886-911.
- [25] Amazon E-commerce company, *Amazon Machine Learning Developer Guide* (2016)