

Rekonstrukce rozdělení pravděpodobnosti z odhadů zobecněných momentů

Diplomová práce

Studijní program:

N2612 Elektrotechnika a informatika

Studijní obor:

Informační technologie

Autor práce:

Bc. Martin Špetlík

Vedoucí práce:

doc. Mgr. Jan Březina, Ph.D.

Ústav nových technologií a aplikované informatiky





Zadání diplomové práce

Rekonstrukce rozdělení pravděpodobnosti z odhadů zobecněných momentů

Jméno a příjmení: **Bc. Martin Špetlík**
Osobní číslo: M18000153
Studijní program: N2612 Elektrotechnika a informatika
Studijní obor: Informační technologie
Zadávací katedra: Ústav nových technologií a aplikované informatiky
Akademický rok: **2019/2020**

Zásady pro vypracování:

1. Aplikujte do cenového funkcionálu pro metodu maximální entropie vhodnou regularizaci.
2. Pro výsledný funkcionál najděte vhodné předpokmínění a optimalizační metodu.
3. Ověřte robustnost tohoto přístupu a optimalizujte výpočet na testovací sadě hustot, viz [1].
4. Porovnejte tento přístup s přímou aproximací distribuční funkce pomocí spline funkcí (viz [2]) a případně s dalšími přístupy.
5. Zobecněte metodu pro případ bivariálních hustot.

Rozsah grafických prací:
Rozsah pracovní zprávy:
Forma zpracování práce:
Jazyk práce:

dle potřeby
40-50 stran
tištěná/elektronická
Čeština



Seznam odborné literatury:

- [1] Farmer, Jenny, and Donald Jacobs. High Throughput Nonparametric Probability Density Estimation. PLOS ONE 13, no. 5 (May 11, 2018): e0196937. <https://doi.org/10.1371/journal.pone.0196937>.
- [2] Giles, Michael B., Tigran Nagapetyan, and Klaus Ritter. Multilevel Monte Carlo Approximation of Distribution Functions and Densities. SIAM/ASA Journal on Uncertainty Quantification 3, no. 1 (January 2015): 267-95. <https://doi.org/10.1137/140960086>.

Vedoucí práce:

doc. Mgr. Jan Březina, Ph.D.
Ústav nových technologií a aplikované informatiky

Datum zadání práce:

9. října 2019

Předpokládaný termín odevzdání:

18. května 2020

prof. Ing. Zdeněk Plíva, Ph.D.
děkan

L.S.

Ing. Josef Novák, Ph.D.
vedoucí ústavu

Prohlášení

Prohlašuji, že svou diplomovou práci jsem vypracoval samostatně jako původní dílo s použitím uvedené literatury a na základě konzultací s vedoucím mé diplomové práce a konzultantem.

Jsem si vědom toho, že na mou diplomovou práci se plně vztahuje zákon č. 121/2000 Sb., o právu autorském, zejména § 60 – školní dílo.

Beru na vědomí, že Technická univerzita v Liberci nezasahuje do mých autorských práv užitím mé diplomové práce pro vnitřní potřebu Technické univerzity v Liberci.

Užiji-li diplomovou práci nebo poskytnu-li licenci k jejímu využití, jsem si vědom povinnosti informovat o této skutečnosti Technickou univerzitu v Liberci; v tomto případě má Technická univerzita v Liberci právo ode mne požadovat úhradu nákladů, které vynaložila na vytvoření díla, až do jejich skutečné výše.

Současně čestně prohlašuji, že text elektronické podoby práce vložený do IS/STAG se shoduje s textem tištěné podoby práce.

Beru na vědomí, že má diplomová práce bude zveřejněna Technickou univerzitou v Liberci v souladu s § 47b zákona č. 111/1998 Sb., o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách), ve znění pozdějších předpisů.

Jsem si vědom následků, které podle zákona o vysokých školách mohou vyplývat z porušení tohoto prohlášení.

27. května 2020

Bc. Martin Špetlík

Poděkování

Rád bych tímto poděkoval doc. Mgr. Janu Březinovi, Ph.D. za vstřícnost, cenné rady a mnoho času, který mi věnoval v průběhu zpracování této diplomové práce.

Abstrakt

Cílem této diplomové práce je zdokonalení algoritmu pro rekonstrukci rozdělení pravděpodobnosti. Aproximace hustoty pravděpodobnosti je provedena pomocí metody maximální entropie. Na základě odhadnutých zobecněných momentů je z množiny všech přípustných řešení vybráno to, které má maximální Shannonovu entropii. Pro výpočet momentů jsou použity Legendreovy polynomy. Při řešení úlohy může docházet k přeučení, které se projevuje zvlněním hustoty pravděpodobnosti. Tento efekt se zvyšuje s rostoucí chybou v odhadu momentů.

K potlačení zvlnění je do původního funkcionálu přidána regularizace, která spočívá v penalizaci druhé derivace logaritmu funkce hustoty pravděpodobnosti. Nový přístup je porovnán s tím předchozím na šesti vytipovaných rozděleních. Pro srovnání je použita Kullback-Leibler divergence. Při aplikaci regularizace dochází ke zlepšení aproximace hustoty pravděpodobnosti.

Výsledky původní i regularizované metody maximální entropie jsou porovnány s přímou interpolací distribuční funkce pomocí spline funkcí. V tomto případě jsou všechny přístupy provedeny nad daty získanými klasickou metodou Monte Carlo a víceúrovňovou metodou Monte Carlo. Také z tohoto srovnání vychází nejlépe metoda s regularizací.

Původní algoritmus metody maximální entropie je rozšířen na rekonstrukce bivariálních rozdělení pravděpodobnosti.

Klíčová slova: rekonstrukce rozdělení pravděpodobnosti, metoda maximální entropie, regularizace, předpodmínění, spline interpolace distribuční funkce, bivariální rozdělení pravděpodobnosti

Abstract

The aim of this diploma thesis is to improve the algorithm for the reconstruction of the probability distribution. The approximation of the probability density function is performed using the maximum entropy method. Based on the estimated generalized moments, the one with the maximum Shannon entropy is selected from the set of all admissible solutions. Legendre polynomials are used to calculate moments. During the solution, an overfitting may occur, which is manifested by a ripple in the probability density function. This effect increases with increasing error in moments estimates.

To suppress the ripple, regularization is added to the original functional, which is based on penalizing the second derivative of the logarithm of the probability density function. The new approach is compared with the previous one on six selected probability distributions. The Kullback-Leibler divergence is used for comparison. The approximation is improved when applying regularization. The results of the original and regularized method of maximum entropy are compared with the direct interpolation of the distribution function using spline functions. In this case, all approaches are performed using data obtained by the Monte Carlo method and the multilevel Monte Carlo method. Also in this comparison, the method with regularization works best.

The original algorithm of the maximum entropy method is extended to reconstructions of bivariate probability distributions.

Keywords: reconstruction of probability distribution, maximum entropy method, regularization, preconditioning, distribution function spline interpolation, bivariate probability distributions

Obsah

Seznam zkratk	9
1 Úvod	10
2 Metoda maximální entropie (MEM)	13
2.1 Řešený problém	13
2.1.1 Shannonova entropie	14
2.1.2 Kullback-Leibler (KL) divergence	14
2.1.3 Vyjádření v podobě rozdělení exponenciálního typu	15
2.1.4 Algoritmus MEM	16
2.2 Numerické řešení	17
2.2.1 Výchozí předpokládání	17
2.2.2 Metoda numerické optimalizace	19
2.2.3 Volba momentových funkcí	20
3 Předpokládání	22
3.1 Příčiny nepozitivně definitní kovarianční matice	22
3.2 Posun vlastních čísel kovarianční matice	23
3.3 Redukce dimenze kovarianční matice pomocí PCA	24
3.3.1 Analýza hlavních komponent	24
3.3.2 PCA v předpokládání MEM	25
4 Srovnávací rozdělení pravděpodobnosti	28
4.1 Normální rozdělení	28
4.2 Lognormální rozdělení	29
4.3 Rozdělení two-gaussians	29
4.4 Rozdělení five-fingers	29
4.5 Cauchyho rozdělení	30
4.6 Nespojité rozdělení	30
5 Analýza KL divergence MEM	32
5.1 Způsob výpočtu odhadů	32
5.2 Chyba aproximace	33
5.3 Chyba odhadu	35

6	Regularizace	40
6.1	Tikhonovská regularizace	41
6.2	TSVD	41
6.3	Regularizace aproximace PDF	42
6.4	Metoda maximální entropie s regularizací	43
6.5	Volba regularizačního parametru	45
6.5.1	Křížová validace	45
6.5.2	LOOCV	46
6.5.3	LOOCV s metodou maximální entropie	46
6.6	Testy na srovnávacích rozděleních	48
6.6.1	Použití výchozího předpokládání	48
6.6.2	Použití předpokládání pomocí PCA	50
7	Rekonstrukce rozdělení pomocí spline funkcí	54
7.1	Spline interpolace	55
7.2	Interpolace distribuční funkce	56
7.2.1	Aplikace metody Monte Carlo	56
7.2.2	Aplikace MLMC	57
7.3	Porovnání spline interpolace s MEM	58
7.3.1	Postup výpočtu	58
7.3.2	Porovnání výsledků	59
8	Rozšíření MEM na bivarietní rozdělení pravděpodobnosti	63
8.1	Metoda maximální entropie pro bivarietní rozdělení	63
8.2	Otestování MEM na bivarietních rozděleních	64
8.2.1	Výsledky testů	65
9	Závěr	67

Seznam obrázků

2.1	Vlastní čísla Hessiany matice $\mathbf{H}(\hat{\boldsymbol{\lambda}})$ bez použití předpokládání	17
2.2	Vlastní čísla Hessiany matice $\mathbf{H}(\hat{\boldsymbol{\lambda}})$ s použitým předpokládání	19
2.3	Legendreovy polynomy	21
3.1	Vlastní čísla Hessiany matice $\mathbf{H}(\hat{\boldsymbol{\lambda}})$ s předpokládáním využívajícím posun vlastních čísel kovarianční matice	24
3.2	Porovnání vlastních čísel Hessiany matice $\mathbf{H}(\hat{\boldsymbol{\lambda}})$ s výchozím předpokládáním a předpokládáním využívajícím PCA	27
4.1	Normální rozdělení	28
4.2	Lognormální rozdělení	28
4.3	Rozdělení two-gaussians	29
4.4	Rozdělení five-fingers	29
4.5	Cauchyho rozdělení	30
4.6	Nespojitě rozdělení	30
5.1	Rekonstrukce rozdělení pravděpodobnosti s přesnými momenty	34
5.2	Průběh KL divergence s rostoucím počtem momentů R	35
5.3	Rozdělení pravděpodobnosti pro zašuměné momenty	37
5.4	Průběh chyby odhadu KL divergence s rostoucím σ	38
5.5	Vývoj iterací algoritmu pro různý počet momentových funkcí R	39
5.6	Vývoj iterací algoritmu pro různé chyby momentů σ	39
6.1	Průběh CV a KL divergence při hledání α_{opt}	48
6.2	Porovnání rekonstrukce rozdělení pravděpodobnosti s použitím regularizace a bez ní	49
6.3	Rekonstrukce rozdělení pravděpodobnosti s regularizací a bez ní, použito je předpokládání pomocí PCA	52
6.4	KL divergence pro různé chyby momentů σ a předpokládání pomocí PCA	53
6.5	Iterace algoritmu pro různé chyby momentů σ a předpokládání pomocí PCA	53
7.1	Porovnání výsledků rekonstrukce rozdělení pravděpodobnosti pomocí metody maximální entropie a spline interpolací	62
8.1	Rekonstruovaná bivariální rozdělení s přesnými momenty	65

8.2	Rekonstruovaná bivarietní rozdělení se zašuměnými momenty	66
8.3	Bivarietní rozdělení - KL divergence a počty iterací numerického řešiče v závislosti na chybě momentů	66

Seznam zkratek

MC	metoda Monte Carlo
MEM	metoda maximální entropie
MLE	maximálně věrohodný odhad
MLMC	víceúrovňová metoda Monte Carlo
MSE	střední kvadratická chyba
PCA	analýza hlavních komponent
PDF	funkce hustoty pravděpodobnosti

1 Úvod

Jedním ze základních problémů statistické analýzy dat je rekonstrukce rozdělení pravděpodobnosti. Jedná se o inverzní problém, jehož řešení spočívá v nalezení funkce hustoty pravděpodobnosti (PDF) z dostupných dat. Inverzní problémy se neomezují pouze na určení rozdělení pravděpodobnosti. Jejich hlavním cílem je obecně nalezení parametrů zkoumaného modelu na základě jeho odpovědi. Tento přístup se uplatňuje v mnoha oblastech, mezi ně se řadí předpovídání počasí, dálkový průzkum země, zobrazovací metody v lékařství a mnoho dalších, viz [59].

V praktických inženýrských úlohách se lze setkat s tím, že zkoumanou veličinu nelze přímo změřit. Důvodem může být nedostupnost oblasti, pokud se jedná například o děje probíhající pod zemským povrchem nebo ve vesmíru. Rovněž nelze přímo zkoumat procesy, které mohou v přírodě trvat stovky a více let. V těchto případech se vytváří počítačová simulace experimentu, která nahrazuje skutečný děj matematickým modelem. Příkladem jsou stochastické počítačové simulace (viz [28]), jejichž výsledky závisí na náhodných parametrech. Pro stanovení dobrých odhadů sledovaných parametrů je nutné počítačovou simulaci opakovat.

Pro tyto účely lze použít metodu Monte Carlo (MC) (viz [19]). Postup může být takový, že jsou nejprve vygenerována vzájemně nezávislá náhodná data s nimiž jsou realizovány stejné počítačové simulace. Následně jsou vypočteny průměry výsledků jednotlivých realizací. Tímto způsobem lze určit nestranný odhad středních hodnot momentů pozorované náhodné veličiny X . Nevýhodou MC je vysoká výpočetní cena nutná ke snížení rozptylu odhadovaných parametrů, je nezbytné provést mnoho realizací simulace (viz [19, s. 2]). Z tohoto důvodu se v posledních letech začala uplatňovat také víceúrovňová metoda Monte Carlo (MLMC). Ta na rozdíl od MC provádí simulace na několika úrovních, kde každá úroveň aproximuje X s různou přesností. Algoritmus MLMC provádí nejvíce simulací na nejhrubší úrovni, která poskytuje nejhorší aproximaci X , ovšem má nejmenší výpočetní cenu. Simulací na nejjemnější úrovni, které odpovídají aproximaci X pomocí MC a jsou výpočetně nejnáročnější, je provedeno nejméně. Tímto způsobem lze dosáhnout stejné přesnosti výsledků jako v případě MC, ale s nižší celkovou výpočetní cenou. Kompletní matematický popis poskytuje M. B. Giles [19]. Metodou Monte Carlo se také zabývala předcházející bakalářská práce [50].

Pomocí MLMC nelze přímo určit rozdělení pravděpodobnosti. Na to je třeba použít jiné metody. Rozdělit se dají do dvou skupin, a to na parametrické a neparametrické metody (viz [18]). V případě parametrických aproximací PDF se předpokládá předchozí znalost tvaru a některých charakteristik rozdělení, ze kterého by měla pocházet dostupná data. Následně jsou na základě dat určeny chybějící parametry,

často se jedná o střední hodnotu nebo rozptyl. Tento způsob může být efektivní pouze pokud je výchozí znalost o rozdělení stanovena správně. Oproti tomu neparametrické metody rekonstrukce hustoty pravděpodobnosti nepracují s předchozí znalostí parametrů rozdělení, ale přímo hledají aproximaci PDF. Mezi tyto metody se řadí dobře známý histogram, jádrové odhady hustot (KDE), metoda maximální entropie, přímá aproximace polynomem a další [52].

V rámci této práce je pozornost zaměřena na neparametrické metody. Konkrétně se jedná o aplikaci metody maximální entropie a interpolaci distribuční funkce pomocí spline funkcí.

Řešení uvedeného inverzního problému rekonstrukce hustoty pravděpodobnosti není jednoznačné. Metoda maximální entropie (MEM)[32] je variační princip, který z množiny rozdělení pravděpodobnosti vybírá to, které má největší Shannonovu entropii a vyhovuje stanoveným omezujícím podmínkám. Těmi jsou v řešené úloze tzv. zobecněné momenty. Ať už jsou tyto momenty určeny pomocí MLMC nebo jinak, vždy se jedná o odhady, které nebudou zcela přesně odpovídat momentům cílového rozdělení. Z toho důvodu může docházet k tzv. přeučení (viz [3, s. 67]). To znamená, že výsledné rozdělení příliš odpovídá testovacím datům a ztrácí se jeho obecnost.

Potlačení přeučení lze dosáhnout přidáním vhodné regularizace do funkcionálu MEM. Regularizace je široká škála metod, její původní smysl spočíval v zajištění toho, aby byl řešený inverzní problém tzv. dobře postulovaný. S rozvojem strojového učení a zpracování dat se regularizace stala jakýmsi synonymem pro zabránění přeučení trénovaného modelu.

Druhou v této práci použitou metodou rekonstrukce rozdělení je interpolace distribuční funkce pomocí spline funkcí. Konkrétně se jedná o algoritmus, který představil M. B. Giles [20]. Distribuční funkce je stanovena metodami Monte Carlo s pomocí indikátorové funkce a B-spline bázových funkcí. Tento způsob přímé interpolace je možné použít s MC. V případě MLMC se naráží na problémy spojené s nespojitostí indikátorové funkce. Přesto i zde existuje možnost řešení v podobě použití hladké aproximace indikátorové funkce.

V mnoha úlohách je zkoumáno více veličin, to vede na myšlenku rekonstrukce multivarietních rozdělení pravděpodobnosti. V rámci této práce je zkoumána možnost rozšíření metody maximální entropie na bivarietní rozdělení pravděpodobnosti.

Vybrané postupy jsou implementovány za použití programovacího jazyka Python.

Následuje popis obsahu jednotlivých kapitol této práce. V kapitole 2 jsou shrnuty předchozí poznatky o řešeném problému. Je zde podrobně popsána metoda maximální entropie. A také je vysvětlen pojem KL divergence. Pomocí ní jsou v dalším průběhu práce porovnávány hustoty pravděpodobnosti. Nachází se zde rovněž vyjádření hledané PDF ve tvaru rozdělení exponenciálního typu. V části 2.2 je popsán postup použitého numerického řešení a volba momentových funkcí pro výpočet zobecněných momentů.

V další kapitole 3 je stručně popsán význam předpodmínění a zavedeny jsou dvě alternativní předpodmínění MEM. V části 3.2 je to posun vlastních čísel kovarianční matice a v části 3.3 je to využití analýzy hlavních komponent.

Kapitola 4 představuje srovnávací rozdělení pravděpodobnosti včetně stručného popisu některých jejich důležitých vlastností. Na těchto rozděleních jsou v dalších částech této práce testovány uvedené přístupy.

Následuje kapitola 5 obsahující analýzu chyby výsledných rozdělení rekonstruovaných pomocí metody maximální entropie. Zkoumány jsou zde složky KL divergence, kterými jsou chyba aproximace (část 5.2) a chyba odhadu (část 5.3), obě jsou porovnány s teoretickými předpoklady.

Kapitola 6 se zabývá regularizací. Její princip je vysvětlen na dvou často se vyskytujících příkladech. A to je Tikhonovská regularizace a TSVD metoda. V části 6.3 je popsán vývoj regularizací v úlohách rekonstrukce PDF. Následující část 6.4 ukazuje modifikované řešení MEM s regularizací. Část 6.5 popisuje metodu, která je použita pro určení optimálního regularizačního parametru. Poslední sekce této kapitoly ukazuje výsledky regularizované MEM na srovnávacích rozděleních.

Další způsob rekonstrukce rozdělení je uveden v kapitole 7. Jedná se o interpolaci pomocí spline funkcí. Nachází se zde definice spline funkcí a jejich vyjádření v podobě B-spline bázových funkcí (část 7.1). V části 7.2 je popsán algoritmus pro výpočet distribuční funkce pomocí MC a MLMC. Následuje porovnání tohoto přístupu s MEM (část 7.3).

V poslední kapitole 8 je rozšířena aproximace PDF pomocí metody maximální entropie na bivarietní rozdělení. Uvedeny jsou rozdíly oproti původní MEM. Zavedeny jsou dvě bivarietní rozdělení, na kterých je upravená MEM otestována.

2 Metoda maximální entropie (MEM)

Pro určení distribuční funkce a funkce hustoty pravděpodobnosti lze využít celou řadu přístupů. Jedním z široce používaných je v současné době metoda maximální entropie. Ta pro dané omezující podmínky, které reprezentují veškerou znalost o rekonstruované náhodné veličině, stanoví ze všech přípustných rozdělení pravděpodobnosti to, které má největší entropii a tedy obsahuje minimum přidané informace.

V této kapitole je představena metoda maximální entropie. Osvětleny jsou pojmy Shannonova entropie a Kullback-Leibler divergence. Použito je vyjádření PDF ve tvaru rozdělení exponenciálního typu, aproximace PDF v tomto tvaru analyzoval již A. Barron [4], ovšem bez vazby na numerickou realizaci. Dále je zde uveden obecný algoritmus řešení MEM a konkrétní postup implementovaného numerického řešení. V poslední části této kapitoly jsou popsány momentové funkce, které se s MEM používají. Postup řešení 2.1.4 byl částečně implementován v předcházející bakalářské práci [50] a následně byl zefektivněn a rozšířen J. Březinou až do podoby, která je zde prezentována.

2.1 Řešený problém

Cílem je ze známých hodnot tzv. zobecněných statistických momentů a funkcí na jejich výpočet určit funkci hustoty pravděpodobnosti.

Nechť X je reálná spojitá náhodná veličina, $\rho(X)$ je její funkce hustoty pravděpodobnosti, která je nezáporná na konečném intervalu Ω . Pak platí následující

$$\begin{aligned} \int_{\Omega} \rho(x) dx &= 1, \\ \int_{\Omega} \phi_r(x) \rho(x) dx &= \mu_r, \quad r = 1, \dots, R \end{aligned} \tag{2.1}$$

kde $\{\mu_r\}_{r=1}^R$ jsou hodnoty zobecněných momentů a $\{\phi_r\}_{r=1}^R$ jsou lineárně nezávislé funkce (viz [4, s. 1350]) použité pro výpočet těchto zobecněných momentů, $\rho(X)$ je v tomto případě neznámá hustota pravděpodobnosti.

$\phi_1 = 1$, $\phi_r \in C^R(\Omega)$, $r = 2, \dots, R$ tvoří bázi lineárního prostoru S_R

$$\mathcal{V}_R = \text{span}\{\phi_r, r = 1, \dots, R\}.$$

Jestliže existuje řešení rovnic 2.1 (tj. hodnoty momentů odpovídají nějaké funkci hustoty pravděpodobnosti), pak má soustava rovnic 2.1 nekonečně mnoho řešení.

Nejlepším nestranným odhadem je podle E. Jaynese [32, s. 623] takové řešení, které má maximální entropii.

2.1.1 Shannonova entropie

V kontextu této práce se entropií resp. Shannonovou entropií [47] označuje míra nejistoty obsažená v reálné spojité náhodné veličině X , a to na základě pravděpodobnostního rozdělení této veličiny. Jedná se o ukazatel toho, jak nejasná je náhodná veličina, která má danou distribuci. Čím více informací o X je k dispozici, tím nižší je její entropie.

Definice 1. Nechť X je spojitá náhodná veličina s hustotou pravděpodobnosti $\rho(x)$, která má nosič $\text{supp}(\rho) \subseteq \Omega$, pro konečný interval $\Omega \subset \mathbf{R}$. Pak Shannonova entropie je funkcional

$$H(\rho) = \mathbb{E}_\rho[-\ln(\rho(X))] = - \int_{\Omega} \rho(x) \ln(\rho(x)) dx. \quad (2.2)$$

2.1.2 Kullback-Leibler (KL) divergence

Se Shannonovou entropií úzce souvisí KL divergence, někdy také označována jako relativní entropie. Jedná se o ukazatel odlišnosti dvou funkcí hustoty pravděpodobnosti.

Definice 2. Kullback-Leibler divergence dvou funkcí hustoty pravděpodobnosti $p(x)$ a $q(x)$ spojité náhodné veličiny X se definuje jako:

$$D(p||q) = \int_{\Omega} p(x) \ln \frac{p(x)}{q(x)} dx. \quad (2.3)$$

Poznámka. KL divergence není vzdálenost, protože nejsou splněny podmínky symetrie a trojúhelníkové nerovnosti. Přesto má celou řadu užitečných vlastností:

- $D(p||q)$ je nezáporná
- $D(p||q) = 0$ tehdy a jen tehdy, jestliže $p = q$
- $D(p||q)$ se chová jako druhá mocnina L_2 normy logaritmů hustot p a q , viz [4, sekce 3]

Kullback-Leibler divergence byla zavedena jako zobecnění Shannonovy entropie (viz [38]). V důsledku lze například namísto maximalizace entropie, minimalizovat KL divergenci vůči známé referenční hustotě. Platí zde, že hustota s maximální entropií je zvláštní případ hustoty s minimální relativní entropií, kde referenční hustota je konstanta (viz [58, s. 356]).

2.1.3 Vyjádření v podobě rozdělení exponenciálního typu

V této části je pomocí variačního počtu vypočtena optimální funkce hustoty pravděpodobnosti ρ , která je následně vyjádřena ve formě rozdělení exponenciálního typu (exponential family). Dále je uvedena tzv. Pythagorova identita Kullback-Leibler divergence a popsány jsou rovněž způsoby jak lze snížit celkovou KL divergenci.

Nalezení PDF s maximální entropií má podobu hledání globálního maxima funkcionálu

$$H(\rho) = - \int_{\Omega} \rho(x) \ln(\rho(x)) dx \quad \text{za podmínek 2.1.} \quad (2.4)$$

K jeho určení jsou použity Lagrangeovy multiplikátory

$$\mathcal{L}(\rho, \boldsymbol{\lambda}) = \int_{\Omega} \rho(x) \ln(\rho(x)) dx - \lambda_0 \left[\int_{\Omega} \rho(x) - 1 \right] - \sum_{r=1}^R \lambda_r \left[\int_{\Omega} \phi_r(x) \rho(x) dx - \mu_r \right].$$

Pomocí variace podle ρ je nalezena rovnice pro stacionární bod

$$\ln(\rho(x)) + 1 - \lambda_0 - \sum_{r=1}^R \lambda_r \phi_r(x) = 0,$$

odtud je pak vyjádřena funkce hustoty pravděpodobnosti

$$\rho(x) = \exp \left[\lambda_0 - 1 + \sum_{r=1}^R \lambda_r \phi_r(x) \right].$$

Totéž lze zapsat ve formě třídy hustot exponenciálního typu, viz [4, 39]

$$\rho(x) = \exp \left[\sum_{r=1}^R \lambda_r \phi_r(x) - \psi(\boldsymbol{\lambda}) \right],$$

kde

$$\psi(\boldsymbol{\lambda}) = \lambda_0 - 1 = \ln \int \exp \left[\sum_{r=1}^R \lambda_r \phi_r(x) \right] dx,$$

$\psi(\boldsymbol{\lambda})$ je tzv. log-partitní funkce (logaritmus normalizačního faktoru), jejíž významem je, aby integrál $\rho(x)$ byl roven 1.

Jestliže $\lambda_0 = 1$, pak řešením 2.4 je

$$\rho(x) = \exp \left[\sum_{r=1}^R \lambda_r \phi_r(x) \right],$$

přičemž Lagrangeovy multiplikátory $\{\lambda_r\}_{r=1}^R$ jsou získány vyřešením soustavy nelineárních rovnic

$$\int_{\Omega} \phi_r(x) \exp \left[\sum_{r=1}^R \lambda_r \phi_r(x) \right] dx = \mu_r, \quad r = 1, \dots, R. \quad (2.5)$$

Vzhledem k tomu, že je $\rho(x)$ ve tvaru rozdělení exponenciálního typu, tak jsou zaručené některé užitečné vlastnosti. Například log-partitní funkce je nekonečně diferencovatelná a navíc je konvexní. Z toho plyne, že i množina všech λ je konvexní množina (viz [55, s. 40]). Hodnoty λ lze jednoznačně určit, protože nelze uváznout v lokálním minimu.

KL divergenci 2.1.2 je možné rozdělit na dvě části, které odpovídají chybě aproximace (bias) a chybě odhadu (variance).

Věta 2.1.1. *Pro všechny hustoty ρ_R z rozdělení exponenciálního typu platí tzv. Pythagorova identita KL divergence*

$$D(\rho \|\hat{\rho}_R) = D(\rho \|\rho_R) + D(\rho_R \|\hat{\rho}_R), \quad (2.6)$$

kde $D(\rho \|\rho_R)$ je chyba aproximace (někdy označováno jako chyba oříznutí (truncation error)), $D(\rho_R \|\hat{\rho}_R)$ je chyba odhadu. ρ_R se označuje rovněž jako informační projekce.

Důkaz věty 2.1.1 viz [4, důkaz lemma 3]. V kontextu MEM lze 2.6 interpretovat tak, že ρ je přesná PDF, ρ_R je řešení MEM s přesnými momenty μ_1, \dots, μ_R , $\hat{\rho}_R$ je hustota určená pomocí MEM pro odhadnuté momenty $\hat{\mu}_1, \dots, \hat{\mu}_R$.

Obě části KL divergence lze teoreticky eliminovat. Chyba aproximace klesá se stoupajícím počtem momentů R . Chyba odhadu klesá se zmenšující se chybou momentů μ_k .

$$\begin{aligned} D(\rho \|\rho_R) &\rightarrow 0 \text{ když } R \rightarrow \infty, \\ D(\rho_R \|\hat{\rho}_R) &\rightarrow 0 \text{ když } \hat{\mu}_k \rightarrow \mu_k. \end{aligned} \quad (2.7)$$

V praxi samozřejmě není možné tyto chyby zcela potlačit.

2.1.4 Algoritmus MEM

Nejsložitějším úkolem v rámci MEM je nalezení řešení soustavy nelineárních rovnic 2.5. V této části je uveden obecný algoritmus řešení minimalizačního problému MEM 2.4. Pro další použití je zaveden vektorový zápis: $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_R)$, $\boldsymbol{\phi} = (\phi_1, \dots, \phi_R)$, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_R)$, kde $\phi_1 = \mu_1 = 1$.

K tomu, aby bylo možné určit hustotu $\rho(x) = e^{\boldsymbol{\lambda} \cdot \boldsymbol{\phi}}$, je třeba nejprve nalézt hodnoty $\boldsymbol{\lambda}$. Vyjádření rovnice 2.5 ve vektorové podobě představuje soustavu nelineárních rovnic, jejíž řešení je $\boldsymbol{\lambda}$:

$$\mathbf{G}(\boldsymbol{\lambda}) = \int_{\Omega} \boldsymbol{\phi} e^{\boldsymbol{\lambda} \cdot \boldsymbol{\phi}} dx - \boldsymbol{\mu} = 0. \quad (2.8)$$

Odpovídající Hessova matice pro $\boldsymbol{\lambda}$:

$$\mathbf{H}(\boldsymbol{\lambda}) = \int_{\Omega} \boldsymbol{\phi} \otimes \boldsymbol{\phi} e^{\boldsymbol{\lambda} \cdot \boldsymbol{\phi}} dx,$$

\mathbf{H} je symetrická a pozitivně definitní. Představený problém lze vyjádřit jako minimalizaci funkcionalu

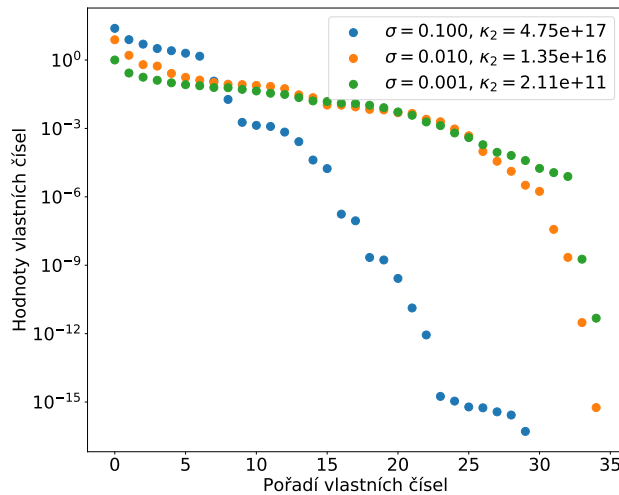
$$F(\boldsymbol{\lambda}) = \int_{\Omega} e^{\boldsymbol{\lambda} \cdot \boldsymbol{\phi}} - \boldsymbol{\lambda} \cdot \boldsymbol{\mu} \quad (2.9)$$

s podmínkou optimality 2.8.

2.2 Numerické řešení

Obsahem této části je představení postupu numerického výpočtu včetně zvoleného předpokládání a numerické metody optimalizace. Obecně neexistuje pro uvedenou úlohu analytické řešení, a proto je nutné použít některou z numerických metod. V rámci tohoto numerického řešení se za μ považují momenty určené s největší dosažitelnou přesností. Statistické odhady momentů jsou označeny $\hat{\mu}$.

Pro realizaci algoritmu 2.1.4 lze použít Newtonovu metodu. V takovém případě je limitující špatná podmíněnost $\mathbf{H}(\hat{\lambda})$, která je pro zkoumané případy dominantně způsobena chybami v odhadu $\hat{\mu}$. Na obrázku 2.1 jsou pro ilustraci uvedeny hodnoty vlastních čísel Hessovy matice $\mathbf{H}(\hat{\lambda})$ v posledním kroku Newtonovy metody. Zobrazeny jsou tři případy pro různé úrovně chyby momentů σ (dále bude podrobněji vysvětleno), přidáno je číslo podmíněnosti κ_2 matice $\mathbf{H}(\hat{\lambda})$. Je zřejmé, že ve všech případech je $\mathbf{H}(\hat{\lambda})$ velmi špatně podmíněná.



Obrázek 2.1: Vlastní čísla Hessovy matice $\mathbf{H}(\hat{\lambda})$ bez použití předpokládání, σ - úroveň chyby momentů, κ_2 - číslo podmíněnosti $\mathbf{H}(\hat{\lambda})$

2.2.1 Výchozí předpokládání

S cílem zlepšit podmíněnost Hessovy matice a zajistit stabilitu a dostatečnou robustnost řešení J. Březina zavedl předpokládání. Pro řešení je použita taková báze \mathcal{V}_R momentových funkcí, pro které je $\mathbf{H}(\hat{\lambda})$ blízká k jednotkové matici. V jiných případech je řešený problém špatně podmíněný. Důležitým poznatkem je, že by v tomto případě měla být $\mathbf{H}(\hat{\lambda})$ blízká tzv. necentrované kovarianční matici momentů $\phi(\hat{X})$.

Centrování $\mathbf{H}(\hat{\lambda})$ se provádí pomocí involutorní matice \mathbf{M} (platí, že $\mathbf{M}^{-1} = \mathbf{M}$)

$$\mathbf{M} = \begin{pmatrix} -1 & 0 & \dots & 0 \\ -\hat{\mu}_2 & 1 & \dots & 0 \\ \vdots & 0 & \ddots & \vdots \\ -\hat{\mu}_R & 0 & \dots & 1 \end{pmatrix}.$$

Výsledná kovarianční matice $\text{Cov } \phi(\hat{X}) = \mathbf{M}\mathbf{H}(\hat{\lambda})\mathbf{M}^T$

$$\text{Cov } \phi(\hat{X}) = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}.$$

Tuto metodu centrování lze použít díky tomu, že vždy platí $\mu_0 = 1$. Stejně jako $\phi(\hat{X})$ tak i kovarianční matice může být odhadnuta pomocí metod Monte Carlo. V takovém případě může být ovšem výsledek zatížen chybou, která způsobuje, že výsledná matice má nenulové i jiné prvky než jen ten první.

Implementace předpokládání

Zde je uveden podrobný postup implementace uvedeného předpokládání. Nejprve je proveden odhad necentrální kovarianční matice $\langle \mathbf{H} \rangle$. Následně je vypočítána aproximace centrované kovarianční matice: $\hat{\mathbf{C}} = \mathbf{M}\langle \mathbf{H} \rangle\mathbf{M}^T$. Poté je $\hat{\mathbf{C}}$ rozložena na vlastní čísla a vlastní vektory $\hat{\mathbf{C}} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^T$.

Odhad $\langle \mathbf{H} \rangle$ je zatížen chybou σ . V důsledku, v závislosti na velikosti σ a počtu momentů R , dochází k tomu, že $\langle \mathbf{H} \rangle$ nemusí být pozitivně definitní. Tedy v $\mathbf{\Lambda}$ se vyskytují nulová nebo záporná vlastní čísla. S cílem zajistit pozitivní definitnost $\langle \mathbf{H} \rangle$ jsou odstraněna nekladná vlastní čísla z $\mathbf{\Lambda}$ spolu s odpovídajícími vlastními vektory v \mathbf{P} . Odstraněna jsou rovněž všechna vlastní čísla jejichž hodnota je menší než σ .

Matice $\mathbf{\Lambda}'$ obsahuje ponechaná vlastní čísla a matice \mathbf{Q} obsahuje ve sloupcích zbývající vlastní vektory. \mathbf{Q} je stále ortogonální matice, platí $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$. Pak

$$\langle \mathbf{H} \rangle \approx \mathbf{M}\mathbf{Q}\mathbf{\Lambda}'\mathbf{Q}^T\mathbf{M}^T$$

Nezbytné je také zredukovat množinu momentových funkcí ϕ , stejně jako byly oříznuty $\mathbf{\Lambda}$ a \mathbf{P} . Pro pozitivně definitní $\mathbf{\Lambda}'$ je proveden RQ rozklad

$$\tilde{\phi} = \mathbf{L}\phi, \quad \mathbf{L} = \mathbf{R}\mathbf{\Lambda}'^{-\frac{1}{2}}\mathbf{Q}^T\mathbf{M},$$

kde $\tilde{\phi}$ jsou zbývající momentové funkce, \mathbf{R} je libovolná ortogonální matice. Nakonec se ukazuje, že matice $\hat{\mathbf{H}}$ by měla skutečně odpovídat jednotkové matici:

$$\begin{aligned} \hat{\mathbf{H}} &= \mathbb{E}_{\hat{\rho}}[\tilde{\phi} \otimes \tilde{\phi}] = \mathbf{L}\mathbb{E}_{\hat{\rho}}[\phi \otimes \phi]\mathbf{L}^T \approx \mathbf{L}\langle \mathbf{H} \rangle\mathbf{L}^T = \mathbf{L}\mathbf{M}\mathbf{P}\mathbf{\Lambda}\mathbf{P}^T\mathbf{M}^T\mathbf{L}^T \\ &\approx \mathbf{L}\mathbf{M}\mathbf{Q}\mathbf{\Lambda}'^{-\frac{1}{2}}\mathbf{\Lambda}'^{-\frac{1}{2}}\mathbf{Q}^T\mathbf{M}^T\mathbf{L}^T = \mathbf{I}. \end{aligned}$$

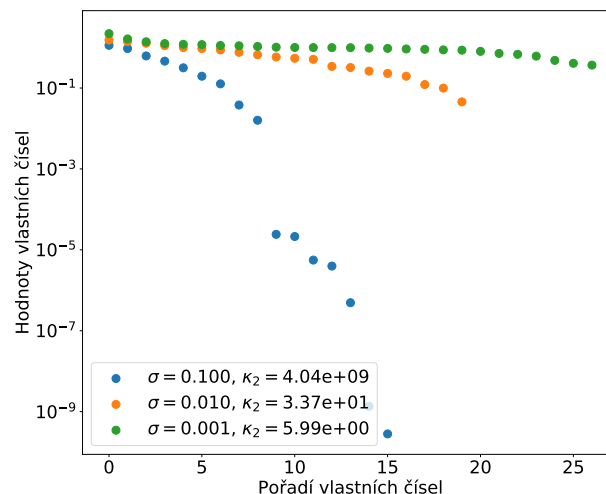
2.2.2 Metoda numerické optimalizace

Pro numerickou minimalizaci je použita metoda trust-region. Jedná se o iterativní optimalizační metodu, variantu Newtonovy metody, která aproximuje nákladovou funkci modelem, který je minimalizovaný v okolí (trust-region) aktuální iterace. Obecný popis tohoto přístupu lze nalézt například v [44, kapitola 4]. Samotný algoritmus metody trust-region nebyl pro účely této práce přímo programován, ale byla použita již hotová implementace v jazyce Python. Konkrétně se jedná o `trust-exact` metodu z knihovny SciPy, autoři navrhli algoritmus podle postupů v [12, s. 169-200]. V této implementaci se využívá přímý lineární řešič pro jednotlivé iterace. Pro výpočet hodnot funkcionálu $F(\boldsymbol{\lambda})$, gradientu $\mathbf{G}(\boldsymbol{\lambda})$ a Hessovy matice $\mathbf{H}(\boldsymbol{\lambda})$ byly naprogramovány samostatné funkce s nimiž následně pracuje metoda `trust-exact`.

Z algoritmu 2.1.4 je patrné, že se bylo nutné vypořádat s numerickou integrací. Použití adaptivní kvadratury je výpočetně náročné a výrazně to zpomalovalo řešení úlohy. Z toho důvodu bylo zvoleno použití pevných kvadraturních bodů. Pro získání bodů a vah je použita metoda `numpy.polynomial.legendre.leggauss`, která poskytuje Gauss-Legendreovy kvadraturní body a váhy s nimiž lze spolehlivě integrovat polynomy až do stupně 100. Při vykonávání minimalizace jsou kvadraturní body a váhy aktualizované v každé iteraci, vždy před určením hodnoty funkcionálu.

Výchozí parametry metody trust-region byly nastaveny takto: tolerance: 10^{-5} , počáteční hodnoty: $\boldsymbol{\lambda} = [1, 0, \dots, 0]$. Počet iterací je implicitně nastaven na 30.

S předpokmáněním a za použití popsané minimalizační metody lze výrazně zlepšit číslo podmíněnosti konečné Hessovy matice (obrázek 2.2). Ukazuje se, že pro velkou chybu momentů σ je stále číslo podmíněnosti κ_2 příliš velké. Dalším možným předpokmáněním se věnuje kapitola 3.



Obrázek 2.2: Vlastní čísla Hessovy matice $\mathbf{H}(\hat{\boldsymbol{\lambda}})$ s použitým předpokmáněním, σ - úroveň chyby momentů, κ_2 - číslo podmíněnosti $\mathbf{H}(\hat{\boldsymbol{\lambda}})$.

2.2.3 Volba momentových funkcí

Vhodné zvolení funkcí ϕ hraje důležitou roli ve smyslu stability a konvergence numerického řešení. To platí o to více v případech, kdy nejsou μ určeny dostatečně přesně.

V literatuře [4, 6, 25] se objevují ve spojitosti s MEM a volbou vhodných funkcí pro výpočet momentů zejména monomiály, Fourierovy funkce (trigonometrické funkce), Legendreovy polynomy nebo spline funkce. Srovnání výsledků MEM s monomiály, Fourierovými funkcemi a Legendreovými polynomy popsal W. Gou [25].

Monomiály

Použité jsou v podobě centrálních momentů:

$$\phi_1(x) = \mu_1 \quad (2.10)$$

$$\phi_{r+1}(x) = (x - \mu_2)^r, \text{ pro } r = 1, \dots, R - 1. \quad (2.11)$$

Tyto funkce nejsou vhodnou volbou. Při použití Newtonovy metody roste číslo podmíněnosti Hessovy matice exponenciálně s R (viz [6, s. 23]). Rovněž případná chyba momentů se může s mocninou zvýrazňovat (to platí pro případy, ve kterých nejsou data standardizována).

Fourierovy funkce

Další variantou je použití Fourierových funkcí:

$$\phi_1(x) = 1 \quad (2.12)$$

$$\phi_{2r}(x) = \sin(ty) \quad (2.13)$$

$$\phi_{2r+1}(x) = \cos(ty), \quad (2.14)$$

kde t se volí jako $\lfloor (r + 1)/2 \rfloor$, $r \geq 1$, $y = x - \mu_2$.

Tyto funkce patří mezi dobře využitelné s MEM. Podle Gou [25] poskytují Fourierovy funkce lepší výsledky pro šikmá rozdělení. Tento autor uvádí, že Fourierovy funkce jsou stabilnější než monomiály a Legendreovy polynomy.

Legendreovy polynomy

Pro tuto práci byly vybrány Legendreovy polynomy

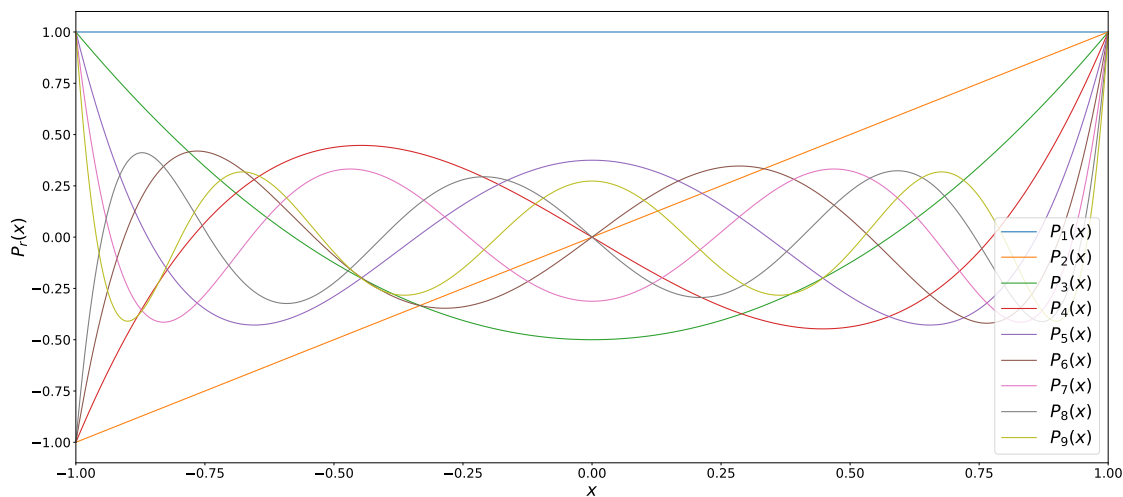
$$P_{n+1}(x) = \sum_{k=0}^n \binom{n}{k} \binom{n+k}{k} \left(\frac{x-1}{2}\right)^k.$$

Tyto polynomy mají některé užitečné vlastnosti:

- jsou vzájemně ortogonální
- $P_1(x) = 1$, $\int_{-1}^1 P_r(x) dx = 0, r > 1$

- Hessova matice v bázi Legendreových polynomů je dobře podmíněná, její číslo podmíněnosti roste lineárně (viz [6, s. 23]).

Legendreovy polynomy jsou typem Fourierových řad zapsaných pomocí ortogonálních polynomů (viz [1, s. 162]). Při použití předpokladů 2.2.1 jsou uvedené momentové funkce přibližně ekvivalentní a prakticky nezáleží na tom, jaká z těchto výchozích bází se použije.



Obrázek 2.3: Legendreovy polynomy

3 Předpodmínění

Předpodmíněním se obecně rozumí technika sloužící ke zlepšení konvergence a stability numerického řešení. Jedním z prvních, kdo se touto technikou zabýval byl D. J. Evans [15], který popsal vztah čísla podmíněnosti matice koeficientů a řešení soustavy lineárních rovnic. Také ukázal, jak lze pomocí předpodmínění zmenšit číslo podmíněnosti Hessovy matice a zlepšit tak rychlost konvergence.

V průběhu metody maximální entropie se provádí řešení soustavy nelineárních rovnic, viz část 2.1.4. Zvláště pro takovéto nelineární systémy je nalezení vhodného předpodmínění obtížné. Je třeba najít rovnováhu mezi tím, aby na jedné straně řešení předpodmíněného systému nemělo přílišnou výpočetní cenu, a bylo tak dostatečně jednoduché. A na druhou stranu, aby předpodmínění přineslo lepší konvergenci řešení. Neexistuje obecné pravidlo na to, jak určit správné předpodmínění (viz [44, s. 120]).

Snahou je docílit malého čísla podmíněnosti Hessovy matice, která odpovídá necentrované kovarianční matici (viz část 2.1.4). Jako předpodmínění byla provedena transformace funkcí pro výpočet momentů ϕ do takové báze, aby výchozí kovarianční matice a výsledná Hessova matice byly blízko jednotkové matici. To je přístup, se kterým se lze setkat i u jiných autorů, jako je např. Yasunori Aoki [2]. Zároveň pro zajištění pozitivní definitnosti odhadu kovarianční matice $\hat{\mathbf{C}}$ bylo přistoupeno k odstranění vlastních čísel menších než zvolená tolerance σ . Toto předpodmínění funguje ve většině případů dobře, ale pro odhady kovarianční matice, které mají velkou chybu dochází k selhání numerického řešiče (viz část 2.2.2). Z toho důvodu jsou otestovány i jiné způsoby toho, jak vyřešit problém kovarianční matice, která není pozitivně definitní.

V této kapitole jsou nejdříve uvedeny dvě časté příčiny, které mohou vést k nepozitivně definitní kovarianční matici. Představena jsou také řešení tohoto problému, a to na příkladu předpodmínění MEM.

3.1 Příčiny nepozitivně definitní kovarianční matice

Ve statistické analýze může vlivem chyb odhadů docházet k výpočtu nepozitivně definitní kovarianční matice. A. G. Holton [29, část 7.3.8] popisuje dva nejčastější důvody proč k tomu dochází. Pro jejich popis je použit náhodný vektor $\mathbf{X} = (X_1, X_2, \dots, X_n)$, $n \in \mathbf{N}$, Σ je kovarianční matice \mathbf{X} .

Prvním důvodem je, že alespoň jedna z náhodných veličin je lineární kombinací ostatních. Tento problém se označuje také jako multikolinearita neboli „téměř“ singularita. Dochází k ní když jsou náhodné veličiny lineárně závislé. Determinant Σ není přímo nula, ale je velmi malý. Důvodem mohou být zaokrouhlovací chyby nebo chyby v datech způsobené málo přesnými odhady. V důsledku jsou některá vlastní čísla matice Σ menší než nula. Tento problém se dá detekovat pomocí rozkladu na vlastní čísla a vektory. Druhou příčinou může být nedostatek dat, a to za situace kdy je počet pozorování menší než počet náhodných veličin n . Pak může vlivem zaokrouhlovacích chyb docházet k tomu, že jsou některá vlastní čísla záporná. V případě úlohy řešené v této práci je nepozitivní definitnost kovarianční matice způsobena použitím MLMC odhadů, při aplikaci MC k tomuto nedochází.

K zajištění pozitivní definitnosti Σ jsou navrženy dva přístupy. Prvním je přičtení konstanty q k vlastním číslům Σ . Tím je zajištěno, že jsou všechna vlastní čísla větší než 0 resp. σ . Druhou možností je redukce dimenze Σ pomocí analýzy hlavních komponent.

Následuje podrobnější popis obou přístupů k vylepšení předpokládání na řešené úloze MEM. Ke každé z variant jsou zobrazeny vlastní čísla a číslo podmíněnosti výsledné Hessiany matice. Zachováno je značení z části 2.2.1.

3.2 Posun vlastních čísel kovarianční matice

Zajištění pozitivní definitnosti kovarianční matice $\hat{\mathbf{C}}$ lze provést přičtením vhodné konstanty q k vlastním číslům Λ matice $\hat{\mathbf{C}}$ (viz [29, část 7.3.8]). Tento postup lze nahradit přičtením hodnoty q k prvkům na diagonále matice $\hat{\mathbf{C}}$, to s sebou přináší výpočetní úsporu.

Platí, že jestliže je λ vlastním číslem $\hat{\mathbf{C}}$ s odpovídajícím vlastním vektorem \mathbf{v} , pak

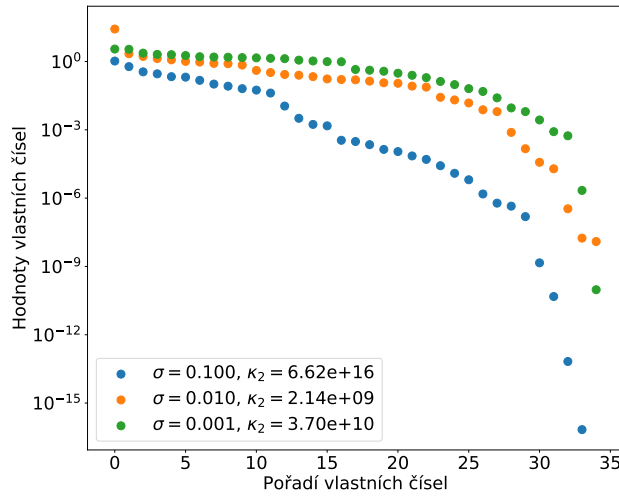
$$\begin{aligned}\hat{\mathbf{C}}\mathbf{v} &= \lambda\mathbf{v} \\ (\hat{\mathbf{C}} + q\mathbf{I})\mathbf{v} &= (\lambda + q)\mathbf{v}.\end{aligned}$$

Hodnota q je zvolena tak, aby nejmenší vlastní číslo bylo rovno hodnotě tolerance resp. hodnotě směrodatné odchylky Gaussovského šumu σ (viz dále sekce 5.1)

$$q = \sigma - \min([\min(\Lambda), 0]).$$

Výhodou tohoto přístupu je, že nemusí docházet k rozkladu na vlastní čísla a vektory. Na druhou stranu se v předpokládání MEM provádí stále RQ rozklad, který má přibližně stejnou výpočetní cenu.

Na obrázku 3.1 jsou zobrazeny vlastní čísla výsledné Hessiany matice $\mathbf{H}(\hat{\boldsymbol{\lambda}})$ pro tři různé hodnoty σ . Oproti variantě bez předpokládání (obrázek 2.1) došlo ke snížení čísla podmíněnosti κ_2 matice $\mathbf{H}(\hat{\boldsymbol{\lambda}})$, a to pro $\sigma = 0.01$ a $\sigma = 0.001$. Přesto jsou tyto hodnoty stále příliš velké a nedaří se dosahovat konvergence numerického řešiče pro MEM. V případě snižování σ řešič začíná konvergovat. Nicméně tento přístup není dostatečně robustní.



Obrázek 3.1: Vlastní čísla Hessovy matice $\mathbf{H}(\hat{\lambda})$ s předpokládáním využívajícím posun vlastních čísel kovarianční matice, σ - úroveň chyby momentů, κ_2 - číslo podmíněnosti $\mathbf{H}(\hat{\lambda})$.

3.3 Redukce dimenze kovarianční matice pomocí PCA

Druhou zmíněnou možností je redukování dimenze kovarianční matice pomocí analýzy hlavních komponent (PCA), podrobně o PCA pojednává I. T. Jolliffe [33]. Jedná se o techniku lineární transformace, jejímž cílem je eliminovat lineární závislosti mezi jednotlivými náhodnými veličinami a současně zachovat co nejvíce informace, která byla obsažena v původních datech. Výsledkem jsou nové veličiny, které se nazývají hlavní komponenty. Mezi nimi již neexistuje žádný lineární vztah. Každá hlavní komponenta reprezentuje lineární kombinaci původních veličin. A je charakterizována svým rozptylem. Hlavní komponenty jsou v PCA seřazeny sestupně podle jejich důležitosti, od největšího rozptylu po nejmenší. Tato metoda se používá také s cílem vizualizovat vícerozměrná data nebo najít skryté veličiny, které přináší dodatečnou informaci o zkoumaném problému.

Dále je v této části popsán obecný princip analýzy hlavních komponent. Následuje představení implementace PCA jako předpokládání pro metodu maximální entropie. Je také uvedeno, z jakých důvodů může být tato metoda vhodná pro řešení problém.

3.3.1 Analýza hlavních komponent

Analýza hlavních komponent se provádí nad náhodným vektorem, který v tomto případě tvoří R momentových funkcí $\phi(\hat{X}) = (\phi_1(\hat{X}), \dots, \phi_R(\hat{X}))$, kde \hat{X} je centrována aproximace náhodné veličiny X .

Vzhledem k tomu, že tato metoda pracuje především s rozptyly, tak k jejímu použití je třeba předpokládat, že pro všechny $\{\phi_r(\hat{X})\}_{r=1}^R$ existují konečné druhé momenty.

První hlavní komponenta β_1 je vektor obsahující R konstant:

$$\beta_1^T \phi = \sum_{j=1}^R \beta_{1j} \phi_j,$$

Zde náhodná veličina $\beta_1^T \phi$ je lineární funkce složek vektoru ϕ , které mají maximální rozptyl, a také platí

$$\|\beta_1\|_2 = \left(\sum_{i=1}^R \beta_{1,i}^2 \right)^{\frac{1}{2}} = 1.$$

Každá další hlavní komponenta

$$\beta_k^T \phi = \sum_{j=1}^R \beta_{kj} \phi_j,$$

je nekorelovaná se všemi předchozími hlavními komponentami, má velikost 1 a zároveň má maximální rozptyl.

V praxi se pro získání hlavních komponent kovarianční matice používá zhusta singulární rozklad (SVD), a to z důvodu výpočetní efektivity. Avšak často aplikovanou variantou je rozklad na vlastní čísla a vektory, který je použit i v této práci. Vlastní vektory reprezentují hlavní komponenty a vlastní čísla udávají jejich velikost.

3.3.2 PCA v předpodmínění MEM

Následuje zavedení PCA jako části předpodmínění MEM. Navazuje se zde na výchozí algoritmus předpodmínění uvedený v sekci 2.2.1. Pro připomenutí, řešení je hledáno pro bázi momentových funkcí \mathcal{V}_R , ve které je $\mathbf{H}(\hat{\lambda})$ blízka k jednotkové matici a k tzv. necentrované kovarianční matici momentů $\phi(\hat{X})$. Matice $\mathbf{H}(\hat{\lambda})$ je rozložena na vlastní čísla a vlastní vektory. Následně jsou odstraněna vlastní čísla, která se nachází pod úrovní tolerance σ . Odstraněny jsou také odpovídající vlastní vektory. A zredukována je rovněž množina momentových funkcí.

Cílem je zde pokusit se zlepšit počet odstraněných momentových funkcí. V některých případech docházelo u výchozího předpodmínění k tomu, že bylo odstraněno příliš momentových funkcí, ačkoliv by to nebylo nutné pro zajištění pozitivní definitnosti kovarianční matice ani pro stabilitu numerického řešiče. V případě velké chyby momentů docházelo naopak k poměrně malému oříznutí, které nevedlo ke zlepšení konvergence numerického řešení.

Při respektování značení z částí 2.2.1 je navržený algoritmus následující. Po provedení odhadu necentrální kovarianční matice $\hat{\mathbf{C}}$ (která nemusí být pozitivně definitní) je tato matice rozložena na vlastní čísla a vlastní vektory $\hat{\mathbf{C}} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^T$, kde vlastní vektory \mathbf{P} jsou v tuto chvíli chápány jako hlavní komponenty a vlastní čísla λ_i na diagonále matice $\mathbf{\Lambda}$ reprezentují velikosti odpovídajících komponent.

Hodnoty λ_i spolu s hlavními komponentami v \mathbf{P} jsou sestupně seřazeny podle velikosti. Vyhodnocen je příspěvek jednotlivých hlavních komponent k celkové informaci o rozptylu původních momentových funkcí

$$\mathbf{l}_i = \sum_{k=1}^i \frac{\lambda_k}{\sum_j \lambda_j}, i = 1, \dots, R,$$

kde \mathbf{l} je vektor kumulativních součtů příspěvků velikostí hlavních komponent na celkovém rozptylu.

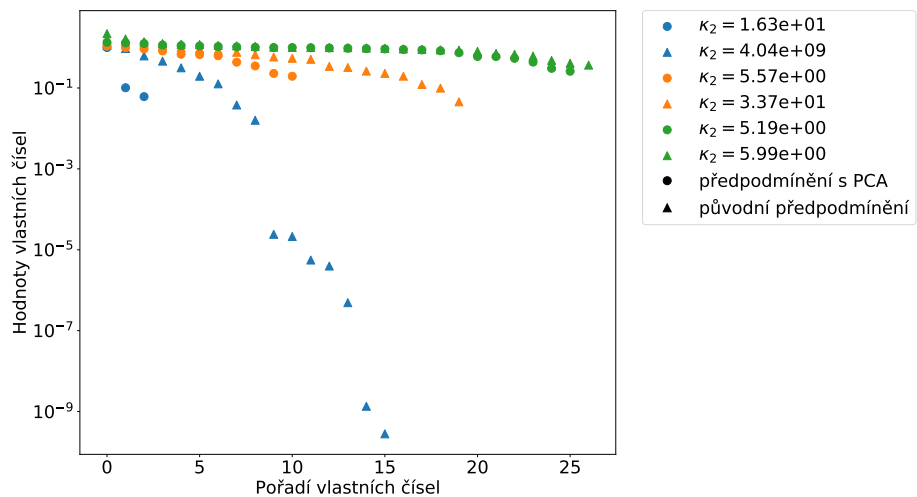
Poté jsou navrženy dvě kritéria na jejichž základě jsou vytvořeny oříznuté matice vlastních vektorů \mathbf{Q} a vlastních čísel $\mathbf{\Lambda}'$:

1. $\mathbf{l}_k > 1 + \sigma^2, k \neq R$. V tomto případě \mathbf{Q} vznikne z \mathbf{P} ponecháním prvních k komponent. Stejným způsobem je oříznuté také $\mathbf{\Lambda}$ na $\mathbf{\Lambda}'$. Zde prvních k komponent obsahuje přibližně 100% informace o rozptylu původních momentových funkcí.
2. $\mathbf{l}_k \leq 1 + \sigma^2, \forall k \in \{1, \dots, R\}$. V této situaci jsou všechny komponenty v \mathbf{P} a $\mathbf{\Lambda}$ seřazeny podle absolutních hodnot λ_i . \mathbf{Q} a $\mathbf{\Lambda}'$ vznikne odstraněním komponent jejichž podíl na celkové informaci je menší než τ . Parametr τ je volitelný, implicitní hodnota $\tau = 10^{-5}$.

Se vzniklými \mathbf{Q} a $\mathbf{\Lambda}'$ se dále pracuje shodně jako v části 2.2.1.

Jedná se o vlastní návrh úpravy výchozího předpokládání. Výhodou tohoto přístupu je, že v bodě 2 nemusí být odstraněno takové množství momentových funkcí jako při použití výchozí verze předpokládání. To je dáno tím, že se pracuje s absolutními hodnotami vlastních čísel, takže větší záporná vlastní čísla se dostanou do kladného spektra a mohou být nad hodnotou τ . Na druhou stranu v případě velmi nepřesných odhadů je přistoupeno k radikálnějšímu oříznutí momentů (bod 1). Díky tomu je zajištěna stabilní konvergence numerického řešení. Mezi nevýhody patří někdy obtížné hledání vhodného parametru τ .

Na obrázku 3.2 jsou znázorněny vlastní čísla výsledné Hessovy matice $\mathbf{H}(\hat{\boldsymbol{\lambda}})$ numerického řešení. Porovnány jsou jejich hodnoty v případě použití výchozího předpokládání a toho s pomocí PCA. Ukazuje se, že při použití PCA dochází k většímu oříznutí vlastních čísel a snižuje se číslo podmíněnosti matice $\mathbf{H}(\hat{\boldsymbol{\lambda}})$. Vliv tohoto přístupu na rekonstruované hustoty pravděpodobnosti je zkoumán v části 6.6.



Obrázek 3.2: Porovnání vlastních čísel Hessiany matice $\mathbf{H}(\hat{\lambda})$ s výchozím předpodmíněním a předpodmíněním využívajícím PCA, σ - úroveň chyby momentů. Pro každé σ a metodu předpodmínění jsou uvedeny čísla podmíněnosti κ_2 matice $\mathbf{H}(\hat{\lambda})$.

4 Srovnávací rozdělení pravděpodobnosti

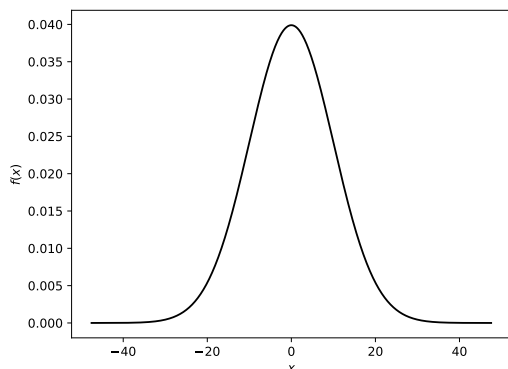
Představené algoritmy jsou testovány na sadě vybraných rozdělení pravděpodobnosti, která byla zvolena podle J. Farmer [16]. Kromě spojitých rozdělení je provedeno testování také na jednom rozdělení, které obsahuje několik bodů nespojitosti. K příkladům z [16] je přidáno lognormální rozdělení, které se často vyskytuje v praktických úlohách. Následuje stručný popis jednotlivých rozdělení včetně zobrazení jejich tvaru hustoty pravděpodobnosti. Pro všechna provedená testování jsou rozdělení pravděpodobnosti omezena na interval $[x_{0.001}, x_{0.999}]$.

4.1 Normální rozdělení

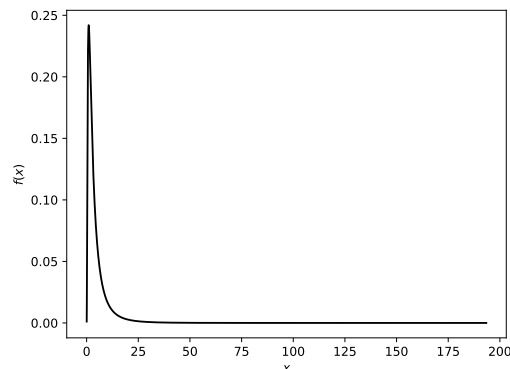
Prvním příkladem je normální rozdělení

$$f(x) = \mathcal{N}(x|\mu = 0, \sigma = 10), x \in (-\infty, \infty).$$

Toto rozdělení (obrázek 4.1) má velmi jednoduchý tvar, symetrický kolem střední hodnoty. Výhodou je, že většina hodnot leží na poměrně malém intervalu. Lze jej z hlediska MEM aproximovat pomocí jedné exponenciály a pro určení PDF postačí malý počet momentů R . Jedná se tak o nejjednodušší případ z použitých rozdělení.



Obrázek 4.1: Normální rozdělení



Obrázek 4.2: Lognormální rozdělení

4.2 Lognormální rozdělení

Lognormální rozdělení (obrázek 4.2) je často používané spojité rozdělení, které je vhodné pro náhodné veličiny, jež mohou nabývat pouze kladných hodnot. Lze ho chápat jako logaritmus, který je normálně rozdělený. Platí, že pokud X má lognormální rozdělení, pak $Y = \ln(X)$ má normální rozdělení. Hustota pravděpodobnosti je zde stanovena takto

$$f(x) = \mathcal{LN}(x|\mu = 4.5, \sigma = 5.9), x \in (0, \infty).$$

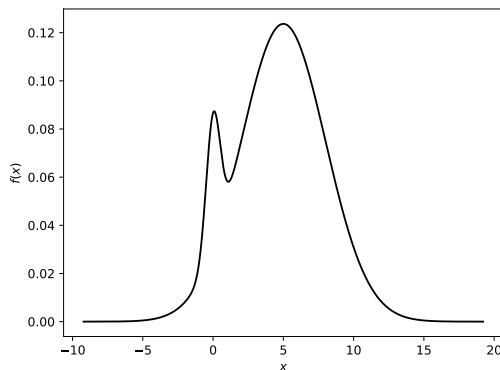
V praxi se lognormální rozdělení používá například pro fyzikální veličiny, které nemohou být záporné, vyskytuje se také často v analýze spolehlivosti systému.

4.3 Rozdělení two-gaussians

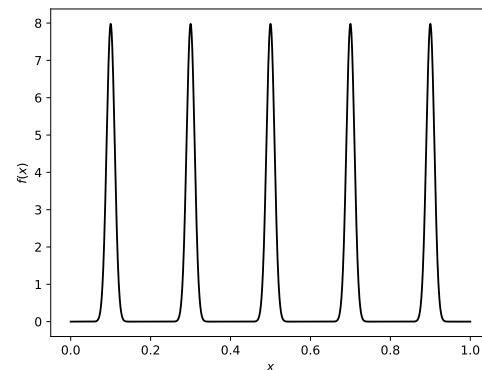
Toto rozdělení obsahuje součet dvou normálních rozdělení s odlišnými středními hodnotami a směrodatnými odchylkami

$$f(x) = \frac{7}{10}\mathcal{N}(x|\mu_1 = 5, \sigma_1 = 3) + \frac{3}{10}\mathcal{N}(x|\mu_2 = 0, \sigma_2 = 0.5), x \in (-\infty, \infty).$$

Průběh funkce hustoty pravděpodobnosti (obrázek 4.3) nemůže být z hlediska MEM vyjádřen jednoduše pomocí jedné exponenciály. Pro nalezení přesného tvaru je zapotřebí větší množství momentů. Pro malé počty momentů je velmi těžké rozlišit mezi zvlněním způsobeným nedostatkem momentů (chybou aproximace) a přítomností dalšího normálního rozdělení, jehož vrchol vybočuje z jedné strany dominantního rozdělení.



Obrázek 4.3: Rozdělení two-gaussians



Obrázek 4.4: Rozdělení five-fingers

4.4 Rozdělení five-fingers

Jedná se o součet pěti normálních rozdělení a rovnoměrného rozdělení

$$f(x) = w \sum_{k=1}^5 \mathcal{N}\left(x \middle| \mu_k = \frac{2k-1}{10}, \sigma_k = \frac{1}{100}\right) + (1-w).$$

Zde je snahou otestovat aproximaci pro velmi špičatá rozdělení, která jsou blízko sebe na krátkém intervalu $[0, 1]$ (obrázek 4.4). Vzhledem k tomu, že jsou rozdělení velmi úzká, tak lze tento interval považovat za úplný a neoříznutý. Podobně jako v předchozím případě i zde by mělo být potřeba k přesnému odhadu použití mnoha momentů. Obtížné bude také, pro použitý algoritmus, přesné zachycení velmi úzkých a špičatých vrcholů a rovněž strmých přechodů mezi jednotlivými vrcholy.

4.5 Cauchyho rozdělení

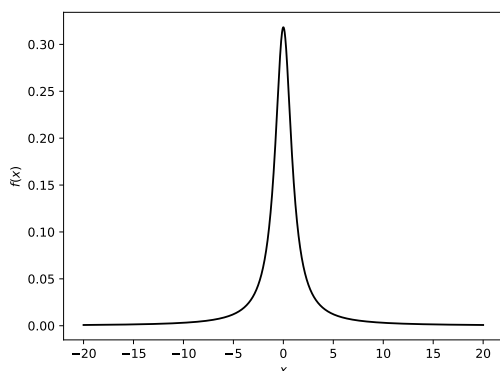
Cauchyho rozdělení (obrázek 4.5) je podílem dvou normálně rozdělených náhodných veličin. Kde rozdělení ve jmenovateli má střední hodnotu rovnu 0.

$$f(x) = \frac{1}{\pi(x^2 + 1)}, x \in (-\infty, \infty).$$

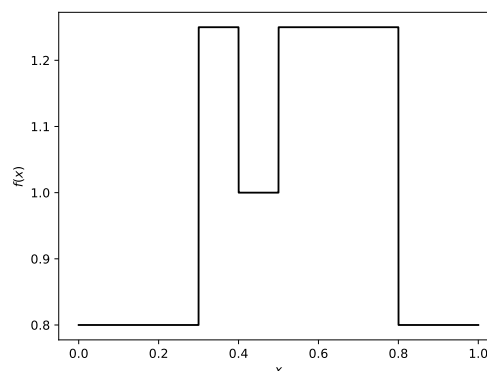
Zvláštností tohoto rozdělení je, že nemá definovanou střední hodnotu ani směrodatnou odchylku resp. rozptyl. Neexistují zde konečné momenty žádného stupně a neexistuje ani jeho momentová vytvořující funkce.

Z důvodu existence řešení a stability MEM je oříznutý interval, na kterém se hledá výsledná hustota (viz [16, s. 20]):

$$[x_{0.25} - 7(x_{0.75} - x_{0.25}), x_{0.75} + 7(x_{0.75} - x_{0.25})].$$



Obrázek 4.5: Cauchyho rozdělení



Obrázek 4.6: Nespojité rozdělení

4.6 Nespojité rozdělení

Posledním testovacím případem (obrázek 4.6) je rozdělení, které má celkem 8 bodů nespojitosti a je definované pro $x \in [0, 1]$

$$f(x) = \begin{cases} 4/5, & x < 0.3 \text{ nebo } x > 0.8 \\ 1, & 0.4 < x < 0.5 \\ 5/4, & \text{jinde.} \end{cases} \quad (4.1)$$

Toto rozdělení není možné zachytit přesně pomocí MEM. Polynomy, které jsou použité pro výpočet momentů mají nosič přes celý interval, takže nedokáží zohlednit body nespojitosti. K alespoň přibližnému zachycení tvaru bude potřeba velký počet momentů, pravděpodobně největší ze všech uvedených rozdělení pravděpodobnosti.

5 Analýza KL divergence MEM

Správnost výsledků výchozího algoritmu je ověřena na zavedených srovnávacích rozděleních pravděpodobnosti (viz kapitola 4). Použito je zde výchozí předpokládání a numerická metoda optimalizace popsaná v části 2.2.2. Cílem je kromě vizuální kontroly určených hustot pravděpodobnosti, také analýza KL divergence (viz část 2.1.2).

Jak bylo uvedeno ve větě 2.1.1, tak KL divergenci je možné rozdělit na chybu aproximace a chybu odhadu. Obě tyto chyby jsou zde rozebrány a numerické výsledky jsou porovnány s teoretickými odhady.

Prvním případem je popsání chyby aproximace. Provedena je MEM s přesnými hodnotami momentů $\boldsymbol{\mu}$. Zkoumán je vliv jejich počtu R na přesnost řešení. Následuje popis chyby odhadu. Zde je pozornost zaměřena na posouzení vlivu chyby momentů na KL divergenci. Metoda maximální entropie je provedena pro různé chyby odhadu momentů $\hat{\boldsymbol{\mu}}$ při konstantním R .

5.1 Způsob výpočtu odhadů

Momenty mohou být odhadnuty pomocí metod Monte Carlo. Pro testování není ovšem tento způsob ideální. Výpočty mohou trvat poměrně dlouho a navíc může být obtížné obdržet výsledky s malými chybami. MC a MLMC odhady jsou s praktickými úlohami použitelné při celkové směrodatné odchylce zhruba do 0.01 až 0.001. Nižší chyby lze dosáhnout jen v případě MLMC s velmi dobrým poklesem rozptylu napříč úrovněmi (viz [41, s. 9646]) nebo při modelování triviálních úloh. Ovšem pro celkovou analýzu metody maximální entropie je vhodné zkoumat i vliv menší chyby odhadů.

Z toho důvodu nejsou v této práci, až na uvedené výjimky, použity pro odhady momentů nebo kovarianční matice metody Monte Carlo. Chyby jsou uměle modelovány přidáním Gaussovského šumu k numericky co nejpřesněji spočítaným hodnotám. Intenzita šumu je řízena směrodatnou odchylkou σ . Odhady momentů mají pak tuto podobu:

$$\hat{\boldsymbol{\mu}} = \boldsymbol{\mu} + \mathcal{N}(0, \sigma^2),$$

kde $\boldsymbol{\mu}$ jsou numericky spočítané přesné momenty, $\mathcal{N}(0, \sigma^2)$ je normální rozdělení se střední hodnotou 0 a rozptylem σ^2 . Implementace Gaussovského šumu používá generátor pseudonáhodných čísel, který má nastavený pevný počáteční stav 1234. Tím je zajištěna reprodukovatelnost dosažených výsledků.

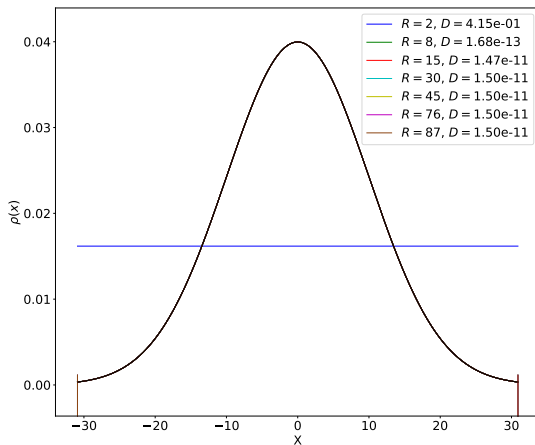
Cílem je, aby rekonstrukce rozdělení pravděpodobnosti poskytovala dobré výsledky pro $\sigma \geq 0.001$, tak aby bylo možné v reálných úlohách používat MLMC. Metoda optimalizace i způsob práce s odhady momentů je stejný napříč tímto textem, pokud to není popsáno jinak.

5.2 Chyba aproximace

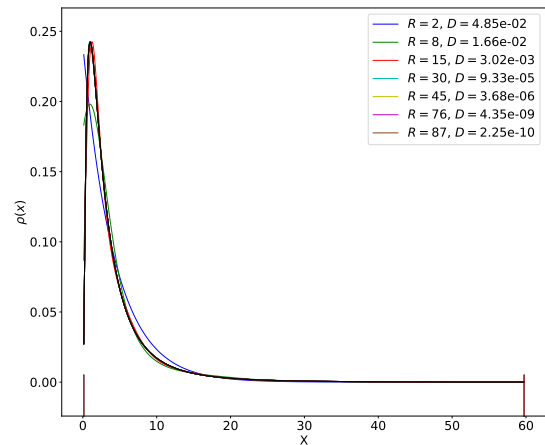
V této části je popsána chyba aproximace provedených rekonstrukcí funkce hustoty pravděpodobnosti. Použity jsou různé počty R přesných momentů μ . Vypočtená chyba aproximace KL divergence $D(\rho||\rho_R)$ je porovnána s teoretickým odhadem.

Na obrázku 5.1 jsou pro jednotlivá srovnávací rozdělení a vybrané hodnoty R vykresleny hustoty pravděpodobnosti ρ , které jsou výsledkem MEM (viz popis v části 2.2). Již z vizuálního posouzení plyne, že s počtem momentů se zlepšuje přiblížení výsledného tvaru PDF ke tvaru referenční PDF. V případě normálního rozdělení (obrázek 5.1a) se tak stane už pro malý počet momentů ($R = 8$). Pro rozdělení two-gaussians (obrázek 5.1c) dochází k poměrně dobrému přiblížení až zhruba od 30 momentů výše. Podobná situace je u lognormálního rozdělení (obrázek 5.1b), kde je pro malé počty momentů ($R < 15$) obtížné zachytit vrchol resp. strmé klesání z vrcholu k počátku sledovaného intervalu.

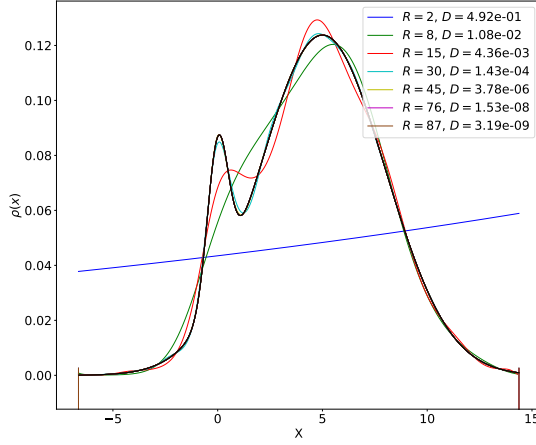
V případě Cauchyho rozdělení (obrázek 5.1e) lze hovořit o dobré aproximace až od více jak 50 momentů. U nespojitého rozdělení dochází k velmi pozvolnému zlepšování aproximace. Ani pro 100 momentů není v tomto případě výsledek příliš dobrý. Zvláštním případem je rozdělení five-fingers (obrázek 5.1d), zde dochází k obstojné aproximace okolo 30 momentů. Jedná se ovšem o jediné z pozorovaných rozdělení, u kterého došlo k oříznutí počtu momentů (podrobněji sekce 2.2.1), konkrétně se to stalo pro případ $R = 100$.



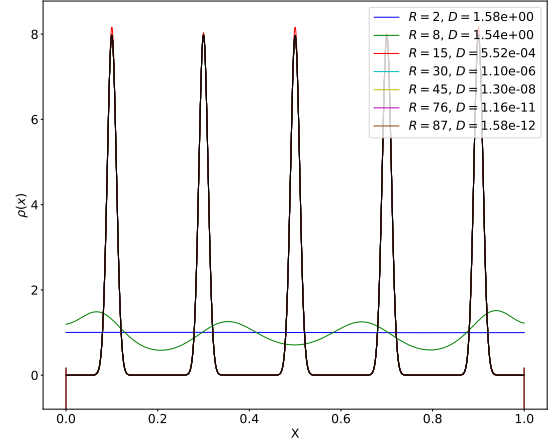
(a) Normální rozdělení



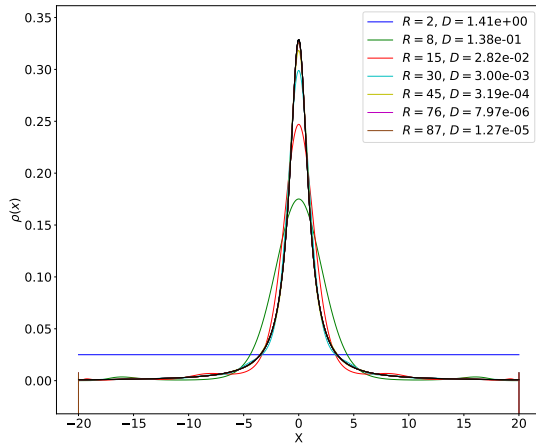
(b) Lognormální rozdělení



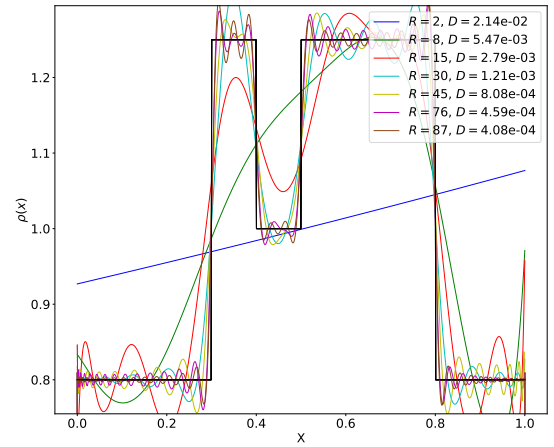
(c) Rozdělení two-gaussians



(d) Rozdělení five-fingers



(e) Cauchyho rozdělení



(f) Nespojité rozdělení

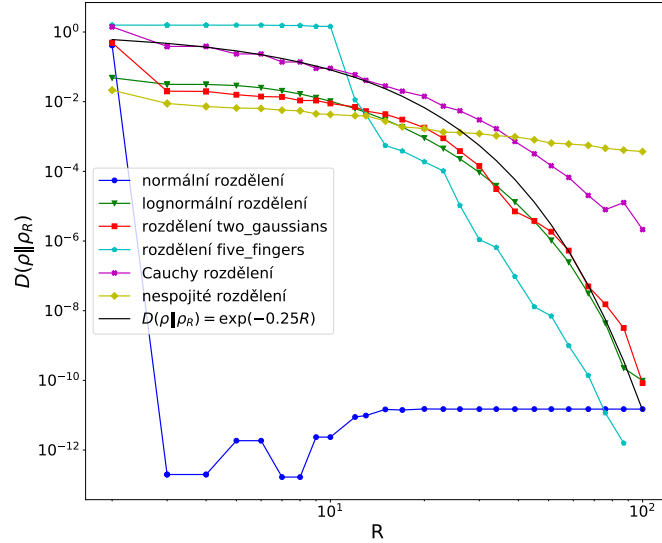
Obrázek 5.1: Rekonstrukce rozdělení pravděpodobnosti s přesnými momenty, R - počet momentů, D - chyba aproximace $D(\rho||\rho_R)$.

Dále bylo přistoupeno k analýze chyby aproximace KL divergence. Pro ni existují teoretické odhady, platí (viz [4, s 1363-1364]):

$$\|\ln \rho - \ln \rho_R\|_2 = \mathcal{O}(1/R)^k,$$

kde $\ln \rho \in W_2^k$. V případě, že je $\ln \rho$ analytické, tak podle [31, s. 22] je $D(\rho||\rho_R) \lesssim \exp(-bR)$, b je konstanta. Z teoretických předpokladů vyplývá, že chyba aproximace $D(\rho||\rho_R)$ závisí na počtu momentů R .

Obrázek 5.2 znázorňuje, pro jednotlivá rozdělení, vývoj chyby aproximace v závislosti na R .



Obrázek 5.2: Průběh KL divergence s rostoucím počtem momentů R

Nejprve jsou popsány křivky pro rozdělení, která lze považovat za analytická, pro ně je přidána křivka ilustrující teoretický odhad $D(\rho \parallel \rho_R) = \exp(-0.25R)$.

V případě normálního rozdělení klesne $D(\rho \parallel \rho_R)$ velmi rychle na úroveň strojové přesnosti, na které se drží i s narůstajícím R . U ostatních rozdělení dochází k poklesu později. Křivky pro lognormální rozdělení, Cauchyho rozdělení a rozdělení two-gaussians mají velmi podobný průběh, což koresponduje s vizuálním posouzením správnosti aproximace MEM.

Zpočátku konstantní je chyba aproximace u rozdělení five-fingers, kde dochází k poklesu až při použití více než deseti momentů. Jiná je situace pro nespojitě rozdělení, které nemá spojitě ani první derivace. A tak by podle teoretických předpokladů nemělo docházet k poklesu $D(\rho \parallel \rho_R)$. Při numerickém řešení nicméně dochází k mírnému lineárnímu poklesu.

Pro $R = 100$ vychází $D(\rho \parallel \rho_R)$ řádově podobně pro normální a lognormální rozdělení i pro rozdělení five-fingers a two-gaussians. Pro Cauchyho rozdělení je o zhruba 4 řády horší, ovšem podle tvaru křivky a provedených experimentů lze předpokládat poměrně strmý pokles se zvyšujícím se R . A tedy pro obdobné výsledky nebude potřeba výrazné navýšení R . Oproti tomu u nespojitě rozdělení, by nemělo docházet k výraznějšímu poklesu ani pro podstatně větší hodnoty R , z experimentů vyplývá, že přinejmenším pro $R < 250$ to platí.

Pro všechny sledované hustoty pravděpodobnosti ρ_R platí trend, že se zvětšujícím se R klesá $D(\rho \parallel \rho_R)$. Zvyšování R s sebou přináší navýšení výpočetní ceny.

5.3 Chyba odhadu

V této části je popsána složka KL divergence, která vyjadřuje chybu odhadu $D(\rho_R \parallel \hat{\rho}_R)$, pro její výpočet byly použity zašuměné momenty $\hat{\mu}$ (viz část 5.1). Kromě zobrazených rekonstruovaných hustot se zde nachází také porovnání numerických

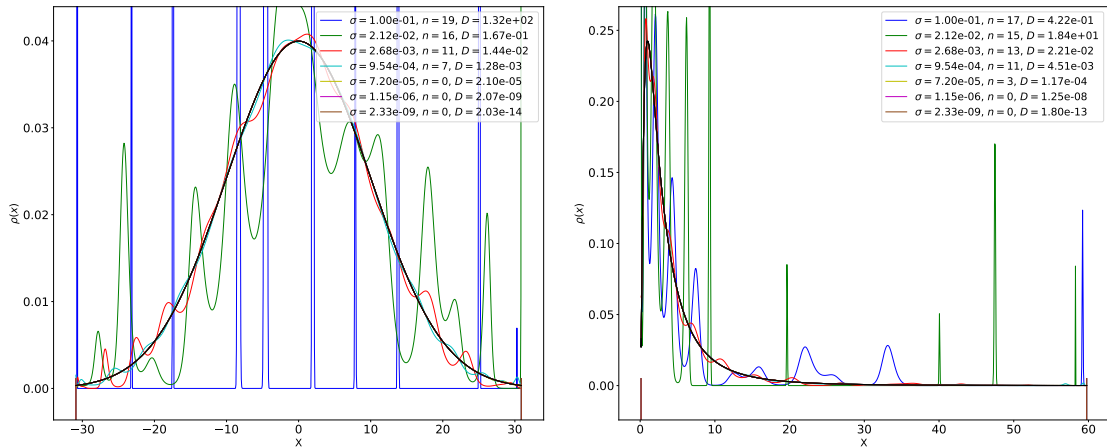
výsledků chyby odhadu s teoretickými předpoklady. Počet momentů byl stanoven na $R = 35$. Tento počet byl zvolen vzhledem k výsledkům chyby aproximace v předchozí části 5.2. Jedná se o kompromis mezi přesností řešení a výpočetní cenou řešení.

Obrázek 5.3 obsahuje pro každé testovací rozdělení porovnání tvaru referenční hustoty a hustoty získané pomocí MEM. Uvedeny jsou případy s úrovní šumu $\sigma = \|\mu - \hat{\mu}\|^2$ v rozpětí přibližně od 10^{-1} do 10^{-9} . Pro každý z těchto šumů je zaznamenána výsledná chyba odhadu $D = D(\rho_{35} \|\hat{\rho}_{35})$ a počet momentů, které byly odstraněny (popis odstranění momentů viz část 2.2.1).

Pro všechna rozdělení vychází velmi špatná aproximace pro $\sigma = 0.1$, dochází zde k největšímu počtu odstraněných momentových funkcí z množiny ϕ . V případě lognormálního nebo Cauchyho rozdělení toto platí i pro $\sigma = 0.0212$. To jsou hodnoty, pro které je Hessova matice špatně podmíněná.

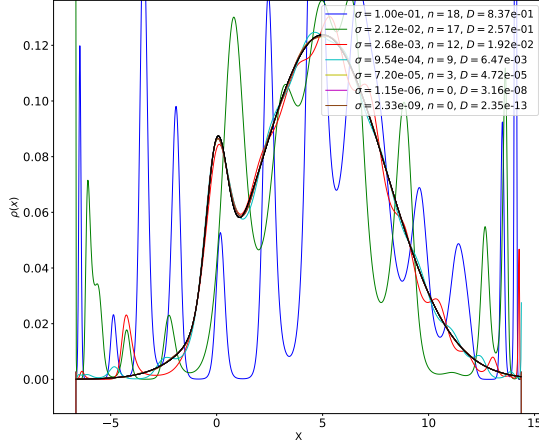
V případě normálního rozdělení nedochází k odstraňování momentů až při malé úrovni šumu ($\sigma \approx 10^{-5}$), podobné to je i u ostatních rozdělení. S výjimkou rozdělení five-fingers, zde nastává odstranění momentů i při malém zašumění dat ($\sigma \approx 10^{-9}$). V případě, že dojde k odstranění některých momentů, tak se zvětšuje chyba aproximace.

Ukazuje se, že chyba odhadu se projevuje zejména ve zvlnění tvaru PDF. To se začíná objevovat, se zvyšujícím se σ , a to především na částech PDF, kde je nejméně informace o rozdělení.

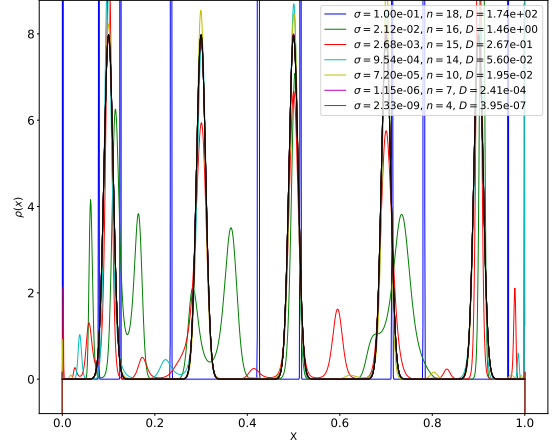


(a) Normální rozdělení

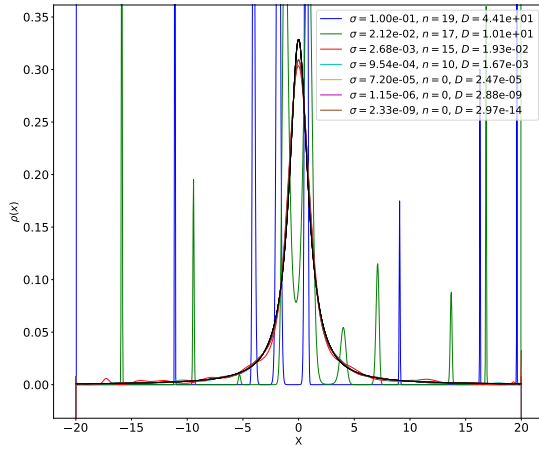
(b) Lognormální rozdělení



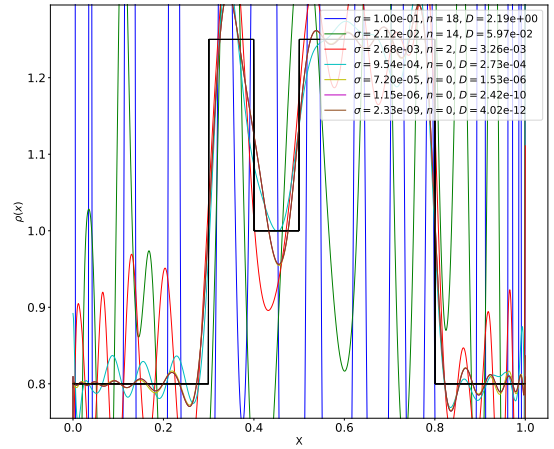
(c) Rozdělení two-gaussians



(d) Rozdělení five-fingers



(e) Cauchyho rozdělení



(f) Nespojité rozdělení

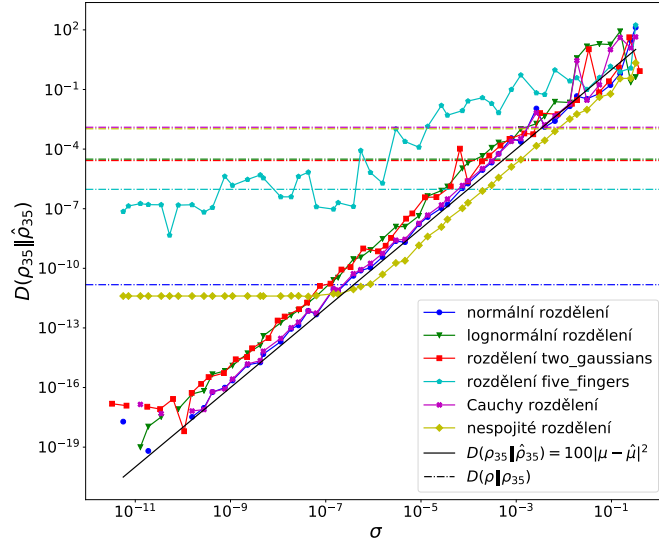
Obrázek 5.3: Rozdělení pravděpodobnosti pro zašuměné momenty, σ - velikost chyby momentů, n - počet oříznutých momentů, D - chyba odhadu pro 35 momentů $D(\rho_{35} \parallel \hat{\rho}_{35})$.

Komplexněji je průběh $D(\rho_{35} \parallel \hat{\rho}_{35})$ pro různé hodnoty σ zobrazen na obrázku 5.4. Přidána je chyba aproximace $D(\rho \parallel \rho_{35})$ a lineární křivka znázorňující teoretický odhad (viz [31, s. 23])

$$D(\rho_R \parallel \hat{\rho}_R) \leq C_R \|\mu - \hat{\mu}\|^2,$$

kde C_R je konstanta, zde $C_R = 10^2$.

Z obrázku je zřejmý velmi podobný průběh pro normální, lognormální i Cauchyho rozdělení, stejně jako pro rozdělení two-gaussians. Jiné je to u nespojitého rozdělení, kde se zastavuje pokles přibližně na úrovni $D(\rho_{35} \parallel \hat{\rho}_{35}) \approx 10^{-11}$. Lze tak konstatovat, že v tomto případě nemá smysl stanovovat $\sigma \lesssim 10^{-7}$. V případě rozdělení five-fingers je trend také klesající ovšem s výraznějšími výkyvy $D(\rho_{35} \parallel \hat{\rho}_{35})$, které jsou způsobeny ořezáváním momentů.

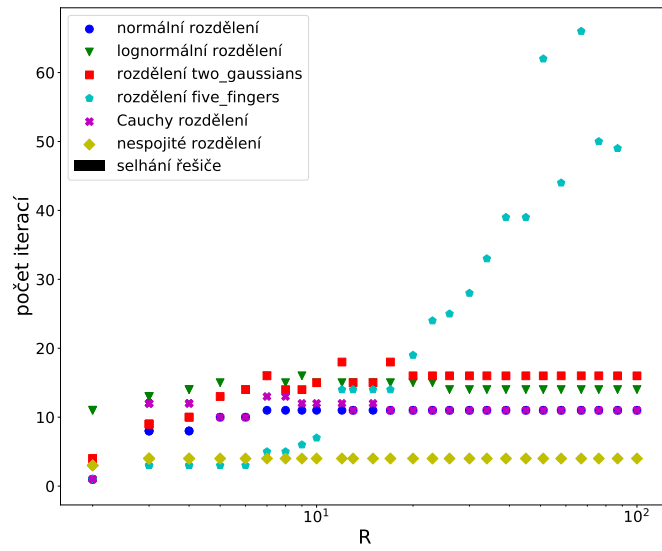


Obrázek 5.4: Průběh chyby odhadu KL divergence s rostoucím σ , horizontální čarou je zobrazena chyba aproximace $D(\rho \|\rho_{35})$.

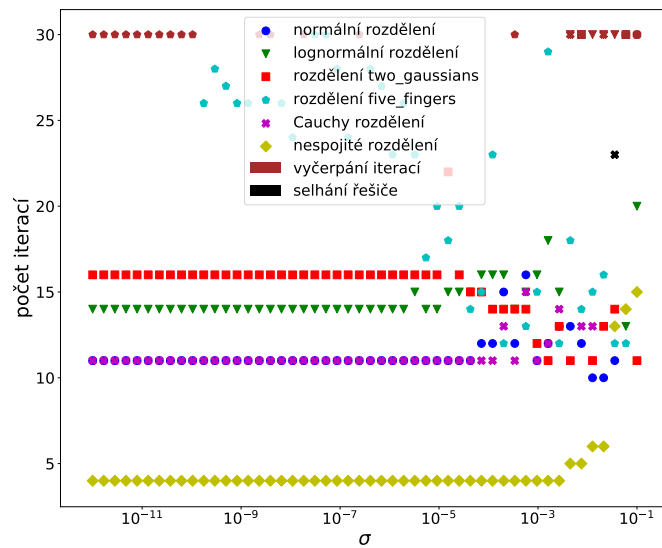
Jak ukazuje obrázek 5.4, tak pro všechna pozorovaná rozdělení zde dochází k tomu, že na určitém intervalu hodnot σ je chyba odhadu $D(\rho_R \|\hat{\rho}_R)$ větší než chyba aproximace $D(\rho \|\rho_R)$. Ta je v obrázku znázorněna vodorovnou čarou. Pro tyto hodnoty σ by mělo být možné stanovit vhodnou regularizaci a pokusit se co nejvíce potlačit chybu odhadu.

Následuje popis schopnosti konvergence numerického řešení. Uvedeny jsou počty iterací algoritmu MEM pro různá R a rovněž pro odlišná σ . Tolerance numerického řešiče zůstává nastavena na 10^{-5} (viz část 2.2.2). V případě výpočtu s použitím přesných momentů (obrázek 5.5) byl zvolen maximální počet iterací 70. Na obrázku 5.5 si lze všimnout toho, že pro rozdělení five-fingers dochází k nárůstu počtu iterací s rostoucím R . Pro ostatní rozdělení je počet iterací přibližně konstantní pro $R > 10$.

Pro případ výpočtu s nepřesnými momenty (obrázek 5.6) byl stanoven maximální počet iterací na 30. A to z toho důvodu, že na základě předchozího případu by měl být tento počet pro většinu rozdělení dostačující. Ukazuje se, že zde dochází často k selhání minimalizačního algoritmu pro největší hodnoty σ , kde je velikost rezidua řádově 0.1 a neklesne tak pod zvolenou toleranci 10^{-5} . Nelze se tomu vyhnout ani navýšením maximálního počtu iterací. Při experimentálním navýšení na 70 iterací tento problém stále přetrvává, projevuje se zde špatná podmíněnost výsledné Hessovy matice, viz část 2.2.1. Pro $\sigma \lesssim 0.001$ již numerický řešič konverguje s výjimkou rozdělení five-fingers, kde algoritmus vyčerpá všech 30 iterací i pro malé hodnoty σ . Velikost reziduí je zde řádově 10^{-4} , se zvýšením počtu iterací (až na 70) lze snížit velikost reziduí pod stanovenou toleranci.



Obrázek 5.5: Vývoj iterací algoritmu pro různý počet momentových funkcí R



Obrázek 5.6: Vývoj iterací algoritmu pro různé chyby momentů σ , hnědou barvou jsou zobrazeny příklady, ve kterých bylo dosaženo maximálního počtu iterací algoritmu, černou barvou je uveden neúspěch řešiče z jiného důvodu.

6 Regularizace

Pojem regularizace zahrnuje celou řadu metod. Lze ho chápat jako techniku vedoucí ke zobecnění původní úlohy. V této kapitole je objasněn význam regularizace a představeny jsou její základní typy. Dále je uveden přehled konkrétních regularizací pro problém rekonstrukce funkce hustoty pravděpodobnosti. Následuje zavedení vybrané regularizace do již představené metody maximální entropie, viz 2.1.4. Efekt regularizace je posouzen na srovnávací sadě rozdělení pravděpodobnosti (viz kapitola 4).

Regularizace je poměrně stará myšlenka, která se objevila již ve 40. letech 20. století (viz A. Tikhonov [53]). Prvotním účelem bylo umožnit řešení tzv. špatně postulovaných (ill-posed) inverzních problémů. Aby byl problém dobře postulovaný (well-posed) tak musí splňovat tři podmínky (Hadamardovu definici):

- existuje řešení problému
- řešení je jednoznačné
- řešení spojitě závisí na vstupních datech a parametrech

Myšlenku řešení inverzních problémů lze vyjádřit obecně jako určení parametrů funkce z provedených pozorování. Příkladem může být právě určení PDF z vypočtených momentů. S ohledem na to, že tyto inverzní problémy často trpí chybou v datech nebo numerickou nestabilitou kvůli diskretizaci, tak mohou být špatně postulované (špatně postulovaný problém může být dobře podmíněný, viz [46, kapitola 2]).

Postupným vývojem se začala uplatňovat regularizace také v oblasti statistického zpracování dat a strojového učení. Zde slouží převážně k zhlazování funkcí a zabránění takzvanému přeučení (overfitting) modelu. Úkolem je zde najít rovnováhu mezi komplexností modelu a přiblížením k testovacím datům.

Lze se setkat se dvěma základními typy regularizací. Jsou jimi L_2 a L_1 regularizace. V případě L_2 regularizace, tedy použití L_2 normy pro regularizační člen, není možné zcela potlačit některé parametry modelu. To znamená, že po regularizaci budou buď zachovány všechny původní atributy, žádný nebude tzv. vynulován, nebo budou potlačeny všechny koeficienty.

L_1 regularizace, která odpovídá použití L_1 normy pro regularizační člen, má také za cíl zmenšit koeficienty, ale navíc poskytuje tzv. řídké řešení. Některé koeficienty mohou být úplně potlačeny. Tato norma má nevýhodu v tom, že není diferencovatelná na celém intervalu. To lze vyřešit například použitím Huber normy nebo implementací vhodného algoritmu optimalizace (viz [48]).

Dále jsou podrobně popsány dvě velmi časté metody regularizace, a to Tikhonovská regularizace a TSVD. V literatuře se lze setkat i s dalšími, které zde nejsou dále zmíněny. Patří mezi ně iterativní regularizační metody, podrobně se jim věnují H. Engl [14] nebo B. Kaltenbacher [34]. Hybridní regularizace, jež typicky kombinují více regularizačních členů, některé aplikace jsou uvedeny v [7, 56, 60]. V oblasti statistiky představili hojně používané metody regularizace P. Bickel a M. Li [5].

6.1 Tikhonovská regularizace

Princip regularizace poprvé uceleně popsal A. Tikhonov [53]. Prvotní motivací bylo zajistit, aby byl řešený problém tzv. dobře postulovaný (well-posed). Jedná se o základní regularizaci, která spočívá v přidání nezáporného konvexního omezení k původnímu funkcionalu. Používá se dnes široce v případech řešení lineárních i nelineárních inverzních problémů (viz [30]).

Postup je vysvětlen na špatně postulovaném lineárním diskretním problému:

$$\mathbf{y} = \mathbf{A} \mathbf{x} + \epsilon, \quad (6.1)$$

kde \mathbf{y} je vektor pozorování, \mathbf{A} je regresní matice (někdy označována jako matice plánu) o hodnotě t , \mathbf{x} je vektor neznámých a ϵ vyjadřuje chybu \mathbf{y} .

Regularizovaný minimalizační problém (v tomto případě lineární regrese) vypadá následovně:

$$\min_x = \{ \|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 + \alpha \|\mathbf{x}\|^2 \}, \alpha \in \mathbf{R}_+$$

Zde první člen $\|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2$ je tzv. cenová funkce, která říká jak dobře řešení \mathbf{x} aproximuje zašuměná pozorovaná data \mathbf{y} . Mohou tu nastat dvě nepříznivé situace. Pokud je tento člen příliš velký, tak to znamená, že \mathbf{x} není dobré řešení. Naopak pokud je příliš malý a tedy pro \mathbf{x} je problém vyřešen velmi přesně, pak dochází k přeučení. To znamená, že se podařilo aproximovat \mathbf{y} včetně šumu. To je výsledek, který také není dobrý, a to především s ohledem na univerzálnost řešení vůči jiné úrovni šumu.

Člen $\|\mathbf{x}\|^2$ definuje regularitu řešení. Největší váha je přikládána největším hodnotám v \mathbf{x} . Koeficient α je tzv. regularizační parametr, který řídí velikost regularizace. Čím větší je α , tím větší vliv má regularizace na výsledek a naopak. Stanovení vhodného α je klíčové pro získání optimálního řešení. V části 6.5 jsou představeny možnosti určení α .

Variety Tikhonovské regularizace a navazující Ivanovovu regularizaci popisuje Ch. Classon [11].

6.2 TSVD

Další široce používanou regularizací je tzv. truncated SVD metoda (viz [27][43, kapitola 4]). Jedná se o „oříznutí“ singulárního rozkladu. Někdy dochází k tomu, že ty největší chyby jsou způsobeny komponentami, které odpovídají nejmenším

singulárním hodnotám z SVD rozkladu. Cílem metody TSVD je odstranění těchto komponent.

Postup této metody je vysvětlen, stejně jako v předchozím případě, na lineárním problému uvedeném v rovnici 6.1. Nejprve je proveden singulární rozklad matice \mathbf{A}

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^T,$$

kde \mathbf{U} a \mathbf{V} jsou ortonormální matice. Pro i -tý sloupcový vektor \mathbf{u}_i matice \mathbf{U} a i -tý sloupcový vektor \mathbf{v}_i matice \mathbf{V} platí, že $\mathbf{u}_i^T \mathbf{u}_j = \delta_{ij}$ a $\mathbf{v}_i^T \mathbf{v}_j = \delta_{ij}$. $\mathbf{\Sigma}$ je diagonální matice singulárních hodnot matice \mathbf{A} . Předpokládá se, že prvky σ z matice $\mathbf{\Sigma}$ jsou kladné a seřazené od největšího po nejmenší: $\sigma_1 \geq \sigma_2 \geq \sigma_t \geq 0$.

Řešení rovnice 6.1 pomocí metody nejmenších čtverců lze zapsat takto:

$$\mathbf{x}_{ls} = \sum_{i=1}^t \frac{\mathbf{u}_i^T \mathbf{y}}{\sigma_i} \mathbf{v}_i.$$

Při použití TSVD dochází k ponechání prvních k komponent. Zbylé, které odpovídají nejmenším hodnotám σ , jsou odstraněny

$$\mathbf{x}_k = \sum_{i=1}^k \frac{\mathbf{u}_i^T \mathbf{y}}{\sigma_i} \mathbf{v}_i,$$

kde k určuje tzv. práh, tedy index mezní hodnoty pro odstraňování komponent.

Výhodou TSVD je jednoduchost a také fakt, že jakmile je jednou vypočítán singulární rozklad, tak je možné velmi jednoduše volit různou intenzitu regularizace pomocí nastavování parametru k . Mezi nevýhody tohoto přístupu patří výpočetní náročnost SVD, která může být zejména pro rozsáhlé úlohy velmi limitující.

Jedná se o analogický postup k předpokládání MEM, kde se využívá odhadu výsledné Hessovy matice, podrobněji je to popsáno v sekci 2.2.1. Místo rozkladu na singulární hodnoty je tam použit rozklad na vlastní čísla a vektory.

6.3 Regularizace aproximace PDF

Nedlouho po prvních publikacích, které popisovaly princip regularizace, se objevila tato idea v řešeních rekonstrukce hustoty pravděpodobnosti. Již v roce 1971 zavedl Good [23], pro neparametrický maximálně věrohodný odhad (MLE) hustoty pravděpodobnosti ρ , regularizaci

$$J(\rho) = \int_{\Omega} (\sqrt{\rho(x)})^2 dx, \quad (6.2)$$

snahou bylo minimalizovat Fisherovu informaci funkce ρ v bodě x , podrobnosti o Fisherově informaci viz [42]. Přidaný člen v podobě L_2 regularizace je konvexní, a proto v principu neztěžuje efektivní vyřešení MLE. Jednalo se o poměrně jednoduchý odhad rozdělení pravděpodobnosti v diskrétních bodech na základě N pozorování.

Cílem regularizace zde bylo zaručit stabilitu řešení a zabránit tzv. Diracově katastrofě, tzn. výsledný odhad hustoty je průměr Diracových pulsů v jednotlivých N bodech.

Ještě v témže roce přišli Good a Gaskins [24] s vylepšením této regularizace, které spočívalo v použití druhé derivace $\sqrt{\rho}$. To již nabízí přímočařejší interpretaci v podobě penalizace hrubosti rozdělení. Použití odmocniny zde zaručuje, že PDF nemůže nabývat záporných hodnot.

Později prezentoval B. Silverman [49] myšlenku penalizace logaritmu funkce hustoty pravděpodobnosti

$$J(\rho) = \int_{\Omega} (\ln \rho(x)')^2 dx. \quad (6.3)$$

Silverman také poznamenává, že logaritmus PDF je přirozený pro maximálně věrohodné odhady nebo neparametrickou diskriminační analýzu.

Zatímco v regularizaci 6.2 je použita první derivace, která slouží ke změření sklonu funkce v bodě. Druhá derivace, která se objevuje v regularizaci 6.3, odpovídá tomu, o kolik se změní sklon v tomto bodě. Tedy druhou derivaci lze v tomto kontextu chápat jako ukazatel hrubosti dané funkce. Tam kde se hustota výrazně vlní bude druhá derivace velká v absolutní hodnotě. Naopak v místech kde je funkce hladká bude blízko nule. Regularizace 6.3 tak popisuje množství celkové změny funkce $\ln \rho(x)'$ na oblasti Ω . V případě, že je $\ln \rho(x)$ hladká funkce, pak $\ln \rho(x)'$ bude téměř konstanta a $J(\rho)$ bude malé. Na druhou stranu pokud $\ln \rho(x)$ osciluje, pak bude také $\ln \rho(x)'$ kolísat a hodnota $J(\rho)$ bude velká. Toto chování regularizace 6.3 je důvodem, proč je vhodné její použití pro zhlazování funkcí hustoty pravděpodobnosti.

Další variantou je penalizace třetí derivace logaritmu hustoty:

$$J(\rho) = \int_{\Omega} (\ln \rho(x)''')^2 dx \quad (6.4)$$

Tato regularizace je zejména výhodná pro normální rozdělení. Je to proto, že čím větší je 6.4, tím lépe $\ln \rho$ aproximuje kvadratickou funkci a v důsledku se tak bude ρ blížit k normálnímu rozdělení (podrobněji Ramsay a Silverman [45, s. 120]).

Další vývoj směřoval k aplikaci L_1 regularizací. R. Koenker a I. Mizera [36] upravili regularizaci 6.3 do podoby penalizace totální variace

$$J(\rho) = \int_{\Omega} |\ln \rho(x)''| dx. \quad (6.5)$$

Výhodou by mělo být snadnější zachycení strmých záhybů a úzkých vrcholů hustot (viz [37]). Nevýhodou je, že L_1 norma nemá definovanou derivaci na celém intervalu Ω .

6.4 Metoda maximální entropie s regularizací

Prvotní snahou bylo do metody maximální entropie implementovat regularizaci spočívající v penalizaci totální variace tak, jak ji představili Koenker a Mizera.

Z důvodu zachování struktury výchozího předpokládání, které využívá odhad kovarianční matice, bylo od tohoto přístupu upuštěno. Dále byly otestovány některé varianty penalizace logaritmu hustoty pravděpodobnosti. Nakonec byla nezávisle na Silvermanovy jako vhodná regularizace určena penalizace druhé derivace logaritmu hustoty pravděpodobnosti.

Do cenového funkcionálu (uvedeného v části 2.1.4) pro výpočet hustoty pravděpodobnosti pomocí metody maximální entropie je přidán regularizační člen $J(\rho)$.

$$\ln \rho = \boldsymbol{\lambda} \cdot \boldsymbol{\phi}$$

$$J(\rho) = (\ln(\rho))''^2 = (\boldsymbol{\lambda} \cdot \boldsymbol{\phi}'')^2.$$

Hledaná hustota ρ nyní závisí také na přidané regularizaci

$$\rho = e^{\boldsymbol{\lambda} \cdot \boldsymbol{\phi}} + \alpha J(\rho).$$

Funkcionál s regularizačním členem:

$$F(\boldsymbol{\lambda}; \alpha) = \int_{\Omega} e^{\boldsymbol{\lambda} \cdot \boldsymbol{\phi}} + \frac{\alpha}{2} (\boldsymbol{\lambda} \cdot \boldsymbol{\phi}'')^2 dx - \boldsymbol{\lambda} \cdot \boldsymbol{\mu}.$$

Podmínka optimality 2.8 má nyní tuto podobu

$$\mathbf{G}(\boldsymbol{\lambda}; \alpha) = \int_{\Omega} \boldsymbol{\phi} e^{\boldsymbol{\lambda} \cdot \boldsymbol{\phi}} + \alpha \boldsymbol{\lambda} (\boldsymbol{\lambda} \cdot \boldsymbol{\phi}'') dx - \boldsymbol{\mu}.$$

Nakonec Hessova matice je následující:

$$\mathbf{H}(\boldsymbol{\lambda}; \alpha) = \int_{\Omega} \boldsymbol{\phi} \otimes \boldsymbol{\phi} e^{\boldsymbol{\lambda} \cdot \boldsymbol{\phi}} + \alpha \int_{\Omega} \boldsymbol{\phi}'' \otimes \boldsymbol{\phi}''.$$

Stejně jako v části 2.2.1 i zde je aplikováno předpokládání. Postup výpočtu je totožný. Předpokládá se, že výsledná Hessova matice $\mathbf{H}(\boldsymbol{\lambda}; \alpha)$ je blízká necentrální kovarianční matici s přičtenou regularizací. Její odhad je označen $\langle \mathbf{H} \rangle$ a je zatížen chybou σ . S vědomím toho, že matice regularizace $\mathbf{R}_H(\alpha) = \alpha \int_{\Omega} \boldsymbol{\phi}'' \otimes \boldsymbol{\phi}''$ má první dva sloupce a řádky nulové, je pomocí \mathbf{M} provedeno centrování $\langle \mathbf{H} \rangle$ následovně:

$$\mathbf{M} \langle \mathbf{H} \rangle \mathbf{M}^T = \hat{\mathbf{C}} + \mathbf{M} \mathbf{R}_H(\alpha) \mathbf{M}^T = \hat{\mathbf{C}} + \alpha \tilde{\mathbf{R}}_H,$$

kde $\hat{\mathbf{C}}$ odpovídá centrované kovarianční matici. Při použití výchozího předpokládání je proveden rozklad na vlastní čísla a jsou odstraněna ta vlastní čísla, která jsou menší než σ , rovněž jsou oříznuty odpovídající vlastní vektory, pak $\langle \mathbf{H} \rangle \approx \mathbf{M} \mathbf{Q} \boldsymbol{\Lambda}' \mathbf{Q}^T \mathbf{M}^T$. Po uskutečnění RQ rozkladu je zredukována množina momentových funkcí $\boldsymbol{\phi}$:

$$\tilde{\boldsymbol{\phi}} = \mathbf{L} \boldsymbol{\phi}, \quad \mathbf{L} = \mathbf{R} \boldsymbol{\Lambda}'^{-\frac{1}{2}} \mathbf{Q}^T \mathbf{M}.$$

6.5 Volba regularizačního parametru

Pokud má být regularizace použita efektivně, tak je nutné vhodně určit regularizační parametr α . Jeho vlastností by mělo být, že roste s chybou dat a naopak pokud jsou data přesná, pak by měl být až nulový.

Neexistuje obecný postup, který by vždy přinášel ideální volbu α . V literatuře [3, 17, 43] se objevují různé přístupy, které jsou často použitelné pro odlišné typy regularizací. Pro následující popis těch nejběžnějších z nich je použit opět lineární model popsany v části 6.1.

Mezi nejrozšířenější přístupy patří tzv. diskrepanční princip (viz [43, část 5.4.1]), zde se předpokládá předchozí znalost odhadu velikosti chyby modelu ϵ . Optimální α je takové, pro které rozdíl aproximace a přesného řešení je roven chybě dat σ , nebo v diskrétním případě je menší než σ

$$\|\mathbf{Ax}_\alpha - \mathbf{y}\|_2 = \sigma,$$

kde $\epsilon \leq \sigma$.

Jestliže není známá chyba dat dopředu, tak je možné použít zobecněnou křížovou validaci, její podrobný popis následuje v části 6.5.1. Další používanou metodou je kritérium L-křivky (L-curve criterion) (viz [26]). L-křivka popisuje vztah mezi normou regularizovaného řešení $\|\mathbf{x}_\alpha\|_2$ a normou chyby řešení $\|\mathbf{Ax}_\alpha - \mathbf{y}\|_2$. Optimální parametr α je takový, který odpovídá „rohu“ L-křivky. V té chvíli je docíleno maximální možné rovnováhy mezi přeučením a nedoučením modelu.

Z pohledu statistického učení se regularizace provádí pomocí rozdělení dat na trénovací, testovací a případně i validační sadu. To je proveditelné v situacích, ve kterých je k dispozici dostatečně velké množství dat. Model je poté proveden (naučení) s trénovacími daty a na testovacích datech jsou ověřeny výsledky. Takto lze postupně vyzkoušet různé hodnoty α a například pomocí minimalizace střední kvadratické chyby (MSE) vybrat tu nejvhodnější z nich.

V praxi často není dostupný dostatek dat tak, aby bylo možné použít dostatečně rozsáhlou testovací sadu. Z toho důvodu byly zavedeny techniky, které vyjmou některá data z procesu trénování a výsledný model je na nich pak otestován. Tyto metody se obecně označují jako křížová validace.

6.5.1 Křížová validace

Křížová validace (cross-validation) (viz [22] nebo [3, část 4.7]) a její varianty jsou jednou z vůbec nejpoužívanějších metod nejen pro odhad regularizačních parametrů. Princip spočívá v náhodném rozdělení dostupných dat na trénovací část a na validační část, se kterou je otestován natrénovaný model.

Na trénovacích datech lze teoreticky vytvořit model s libovolně malou chybou. Ovšem to zpravidla znamená velkou chybu při použití testovacích dat a tedy vznik již zmíněného efektu přeučení. Chyba modelu s validačními daty poskytuje odhad testovací chyby. Obvykle se jako ukazatel velikosti chyby používá střední kvadratická chyba (MSE).

6.5.2 LOOCV

Jednou z variant křížové validace je LOOCV (leave-one-out cross validation), viz [3, část 4.7]. Na rozdíl od klasické křížové validace zde nejsou data rozdělena na dvě velké části, ale ze všech dat je vybrán jeden vzorek (x_1, y_1) pro validaci modelu a ostatní vzorky $[(x_2, y_2), \dots, (x_n, y_n)]$ jsou použity pro trénování. $\text{MSE}_1 = (y_1 - \hat{y}_1)^2$ je rovno přibližně nestrannému odhadu testovací chyby, který má ovšem velký rozptyl, proto se postup opakuje i pro všechny další dvojice trénovacích a validačních dat

$$CV_n = \frac{1}{n} \sum_{i=1}^n \text{MSE}_i.$$

Určení CV_n se provede pro všechny regularizační parametry α ze stanovené množiny. Jako optimální parametr α_{opt} je vybrán ten, pro který je CV_n nejmenší.

6.5.3 LOOCV s metodou maximální entropie

Z principu LOOCV vychází algoritmus 1, který slouží k nalezení α_{opt} pro regularizovanou metodou maximální entropie (viz sekce 6.4).

Nejprve jsou pro momentové funkce ϕ určeny zašuměné momenty $\hat{\mu}$, zde je klíčová volba přidaného šumu, který by měl být podobně velký jako rozptyl momentů. Jako analogie k validačním vzorkům z LOOCV je provedeno N perturbací přesných momentů mírně větší úrovní šumu, než tomu je u $\hat{\mu}$, vzniklé momenty jsou označeny $\hat{\mu}_c^i$, kde $i = 1, \dots, N$. Dále je stanoven vektor K regularizačních parametrů $\alpha = [\alpha_1, \dots, \alpha_K]$. Pro každé α_k , $\hat{\mu}$ a regularizaci J je pomocí MEM stanovena hustota pravděpodobnosti $\hat{\rho}$. S ní jsou následně vypočteny momenty $\hat{\mu}_f$, které jsou porovnány na základě MSE s $\{\hat{\mu}_c^i\}_{i=1}^N$. $\text{MSE}_i = \|\hat{\mu}_f - \hat{\mu}_c^i\|_2^2$.

$$CV = \frac{1}{N} \sum_{i=1}^N \text{MSE}_i$$

$$\alpha_{opt} = \underset{\alpha_k}{\text{argmin}}(CV)$$

Algoritmus 1: Určení regularizačního parametru

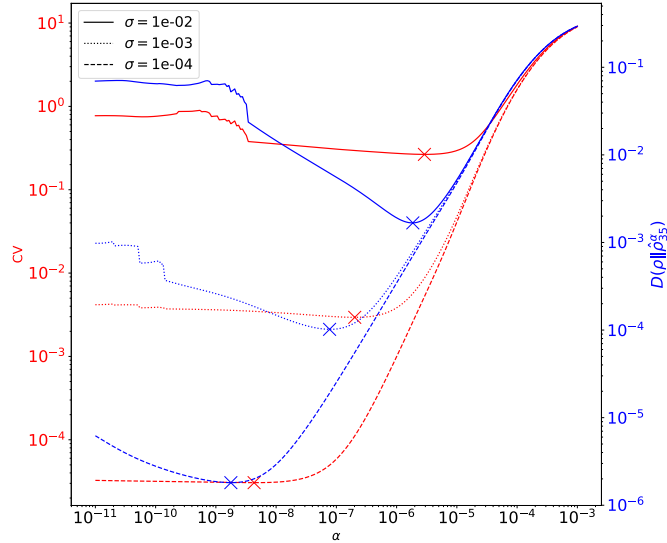
```
1 Function FindOptimalAlpha( $\alpha, \hat{\mu}, \hat{\mu}_c, \phi, J$ ):
2    $N = 100$ 
3    $CV = \text{MAX}$ 
4    $\alpha_{opt} = 0$ 
5   for  $\alpha_k$  in  $\alpha$  do
6      $\hat{\rho} = \text{MEM}(\tilde{\phi}, \hat{\mu}, \alpha_k, J)$ 
7      $\hat{\mu}_f = \int_{\Omega} \phi \hat{\rho}$ 
8      $\text{MSE}_{\alpha_k} = 0$ 
9     for  $i$  in  $N$  do
10       $\text{MSE}_{\alpha_k} = \text{MSE}_{\alpha_k} + \|\hat{\mu}_f - \hat{\mu}_c^i\|^2$ 
11    end
12    if  $\text{MSE}_{\alpha_k}/N < CV$  then
13       $\alpha_{opt} = \alpha_k$ 
14    end
15  end
16  return  $\alpha_{opt}$ 
```

Vzhledem k tomu, že je znám vztah mezi KL divergencí a L_2 normou chyby momentů (viz část 5.3), tak s poklesem této chyby dochází také k poklesu příslušné KL divergence. V popsaném algoritmu se pracuje s Gaussovským šumem namísto přímého použití metody Monte Carlo. Nicméně tento algoritmus 1 je snadno upravitelný pro aplikaci s MC nebo MLMC. Pro odhad $\hat{\mu}_c^i$ je v takovém případě použita metoda bootstrap. Pakliže je znám rozptyl momentů dopředu, pak uvedený postup odpovídá diskrepančnímu principu.

Na obrázku 6.1 jsou znázorněny křivky průběhu CV na jejichž základě je zvoleno α_{opt} . Uveden je příklad pro normální rozdělení a výchozí předpokládání. Zobrazeny jsou případy pro tři různé úrovně chyby momentů σ . Pro každou z hodnot α je vynesena nejen hodnota CV , ale také KL divergence $D(\rho \|\hat{\rho}_{35}^\alpha)$ mezi referenční hustotou pravděpodobnosti a aproximací MEM s použitou regularizací.

Ukazuje se, že je možné touto metodou určit optimální regularizační parametr, který je blízko tomu, pro nějž vychází nejlepší KL divergence. V případě, že je cílem použít velmi malou regularizaci (řádově $\alpha \lesssim 10^{-10}$) při malém σ , tak může docházet k chybnému nalezení α_{opt} . Protože jak se ukazuje z obrázku 6.1, pro tyto hodnoty je průběh CV téměř konstantní a je tak, i vzhledem k velikosti K , obtížné najít minimum.

Uvedený algoritmus byl vytvořen nad rámec zadání diplomové práce, jedná se o prvotní verzi. Pro spolehlivé nalezení α_{opt} bude vhodné tento algoritmus do budoucna zdokonalit.



Obrázek 6.1: Průběh CV a KL divergence při hledání α_{opt} . Znázorněn je červeně průběh CV a modře průběh KL divergence $D(\rho \|\hat{\rho}_{35}^\alpha)$, křížky označují nalezená minima jednotlivých křivek. Typem čar jsou odlišeny různé úrovně chyby momentů σ .

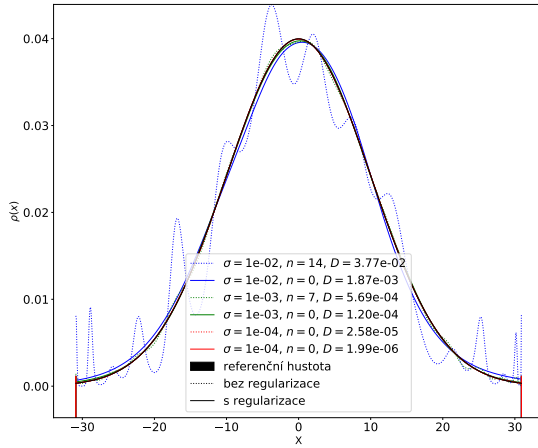
6.6 Testy na srovnávacích rozděleních

Uvedená regularizovaná metoda maximální entropie, včetně postupu pro určení regularizačního parametru, je otestována na srovnávacích rozděleních (viz kapitola 4). Pro určení α_{opt} bylo zvoleno $\alpha = [\alpha_1 = 10^{-11}, \dots, \alpha_{150} = 10^{-5}]$. V analýze výsledků je proto třeba brát v potaz, že nemuselo dojít k nalezení nejlepšího α_{opt} . Testy jsou provedeny nejprve pro výchozí předpokládání, a pak také s předpokládáním využívajícím PCA. V obou případech je použito 35 momentů ($R = 35$).

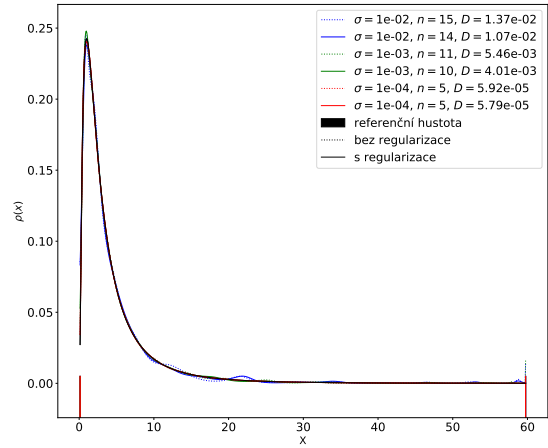
6.6.1 Použití výchozího předpokládání

Nejdříve je provedena rekonstrukce hustot pravděpodobnosti s použitím výchozího předpokládání, ve kterém je redukce momentů řízena chybou odhadu σ (viz část 2.2.1).

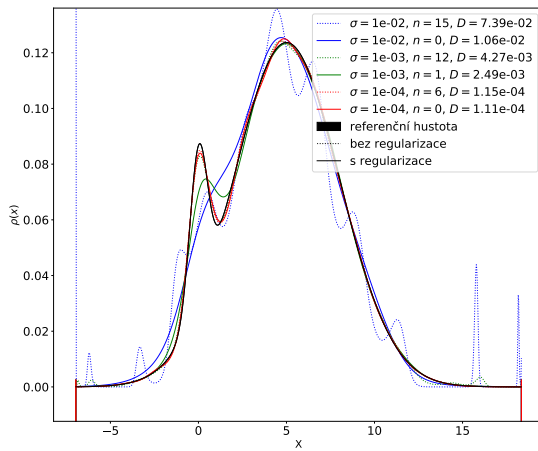
Na obrázku 6.2 jsou zobrazeny rekonstruovaná rozdělení pro MEM bez regularizace a MEM s regularizací. Testovány jsou tři úrovně chyby odhadu momentů $\sigma = [0.01, 0.001, 0.0001]$. Hodnota $\sigma = 0.01$ je přibližně největší chyba, pro kterou bylo možné rekonstruovat rozdělení bez selhání řešiče (viz část 5.3). Pro $\sigma = 0.0001$ již pro většinu rozdělení nedocházelo k příliš velkým chybám aproximace PDF, a proto nejsou menší σ uvažována. Pro každý z těchto případů je uvedena KL divergence $D = D(\rho_{35} \|\hat{\rho}_{35})$ a počet oříznutých momentů n . Neregularizované případy MEM jsou zde zobrazeny pouze pro srovnání a jejich detailnější analýza, byť pro odlišná σ , byla provedena v části 5.3.



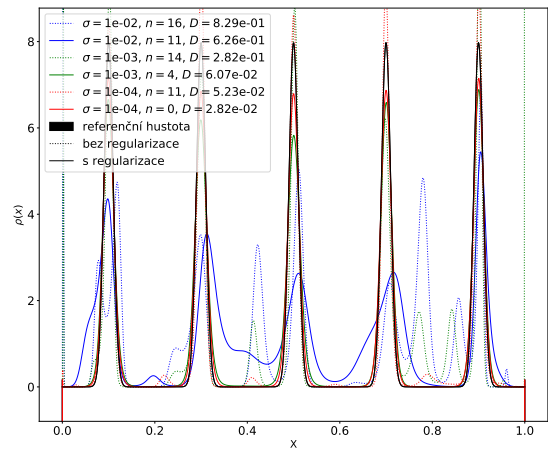
(a) Normální rozdělení



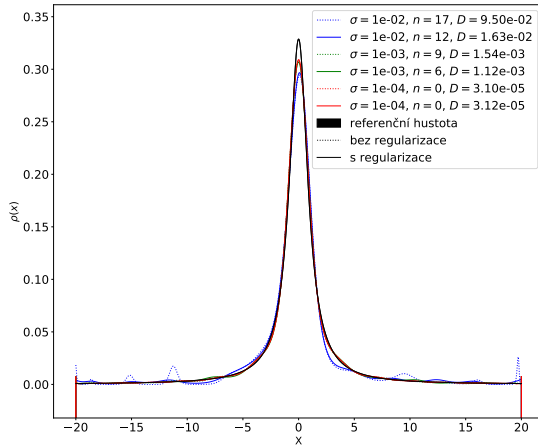
(b) Lognormální rozdělení



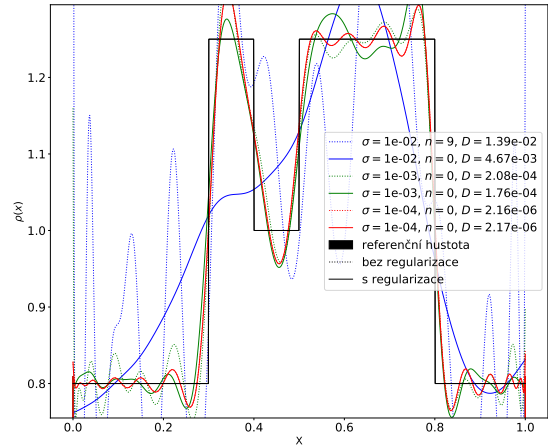
(c) Rozdělení two-gaussians



(d) Rozdělení five-fingers



(e) Cauchyho rozdělení



(f) Nespojité rozdělení

Obrázek 6.2: Porovnání rekonstrukce rozdělení pravděpodobnosti s použitím regularizace a bez ní. Pro každou z chyb σ je uveden tečkovanou čarou výsledek pro MEM bez regularizace a plnou čarou výsledek MEM s regularizací, n je počet oříznutých momentů a D je KL divergence $D(\rho_{35} \parallel \hat{\rho}_{35})$.

V případě normálního rozdělení (obrázek 6.2a) dochází ve všech případech k poklesu KL divergence D při použití MEM s regularizací. Vizuálně se tento efekt nejvíce projevuje při $\sigma = 0.01$, kde je zcela potlačeno zvlnění. Podobného výsledku je dosaženo i pro $\sigma = 0.001$, zde se pro původní MEM objevovalo zvlnění jen na koncích intervalu, to je nyní zhlazeno. Pro nejmenší $\sigma = 0.0001$ již nejsou okem patrné defekty ve vypočtené PDF pomocí MEM. Přesto také zde dochází k poměrně velkému zlepšení KL divergence téměř o jeden řád.

Z obrázku 6.2b lze vyčíst, že pro lognormální rozdělení nedochází k tak velkému snížení D jako pro normální rozdělení. Také se zde nedaří zcela potlačit zvlnění. Je to dáno především kritérii pro volbu regularizačního parametru α_{opt} . Při větším vlivu regularizace lze dosáhnout potlačení zvlnění, ale dojde také k mírnému snížení vrcholu rozdělení a tím je ovlivněna výsledná D . V úlohách, kde by to bylo vhodné, lze vyhladit tzv. dlouhý chvost a tím snížit KL divergenci v této oblasti, ovšem zbytek tvaru hustoty pravděpodobnosti bude výrazně porušen.

V případě rozdělení two-gaussians (obrázek 6.2c) nastává pro $\sigma = 0.01$ situace, ve které sice dochází k dobré eliminaci nežádoucího zvlnění, ale zároveň je velmi potlačen i menší z vrcholů rozdělení. Tomuto efektu se nelze zcela vyhnout, avšak i zde je znatelně sníženo D . Pro zbylá σ se vzhledem k menšímu zvlnění daří lépe zachytávat postranní vrchol tohoto rozdělení a vliv regularizace na KL divergenci se zmenšuje.

Obrázek 6.2d ukazuje výsledky pro rozdělení five-fingers. Zde pro $\sigma = 0.01$ nastává s regularizací jen mírné zlepšení D . Přesto v regularizovaném případě přinejmenším odpovídají vrcholy počtem a umístěním vrcholům referenčního rozdělení. I zde se KL divergence zlepšuje s klesajícím σ .

U Cauchyho rozdělení (obrázek 6.2e) je také s aplikovanou regularizací zmenšeno D . Při $\sigma = 0.01$ je potlačena velká část zvlnění včetně nárůstu PDF na koncích pozorovaného intervalu. Přesto se i v regularizovaném případě malé zvlnění objevuje. Pro $\sigma = 0.0001$ dochází v případě regularizace k mírnému navýšení KL divergence. Zde je to způsobeno hodnotou α_{opt} , která by měla být ještě menší než je mezní hodnota $\alpha_1 = 10^{-11}$.

Na obrázku 6.2f jsou zobrazeny výsledky pro nespojitě rozdělení. KL divergence pro první dvě hodnoty σ klesá. Zvláště pro $\sigma = 0.01$ dochází k úplnému potlačení oscilací. Stejně jako v případě Cauchyho rozdělení i zde je mírné zvýšení D pro $\sigma = 0.0001$, důvod je stejný.

Z prezentovaných výsledků lze konstatovat, že se pomocí regularizace daří ve valné většině případů snižovat KL divergenci. Za pozornost rovněž stojí vliv regularizace na počet oříznutých momentových funkcí. Ve většině případů dochází k poklesu počtu odstraněných momentů. Tím je zlepšeno zachycení tvaru rozdělení.

6.6.2 Použití předpokládání pomocí PCA

Zde je realizována rekonstrukce srovnávacích rozdělení s použitím předpokládání, které pro redukci momentů využívá analýzu hlavních komponent (viz část 3.3). Také na tomto místě jsou porovnány varianty MEM bez regularizace a s regularizací. Značení je stejné jako v předchozí části. Oproti ní je zde kromě použití jiného před-

podmínění také navíc přidána další chyba momentů $\sigma = [0.1, 0.01, 0.001, 0.0001]$. Dále je pozornost věnována především příkladům se $\sigma = 0.1$. Pro většinu rozdělení a ostatní σ jsou výsledky srovnatelné s výchozím předpokmáněním. A tak by bylo jen zopakováno to co bylo již popsáno v části 6.6.1.

V případě normálního rozdělení (obrázek 6.3a) dochází pro $\sigma = 0.1$ k oříznutí 32 momentů, stanovená PDF je vizuálně velmi nepřesná. K výraznému zlepšení dochází s použitím regularizace, kde již nedochází k odstranění žádného momentu a výsledný tvar je ovlivněný především chybou odhadu momentů a nikoliv chybou aproximace. Velice podobně se chovají PDF pro rozdělení two-gaussians (obrázek 6.3c).

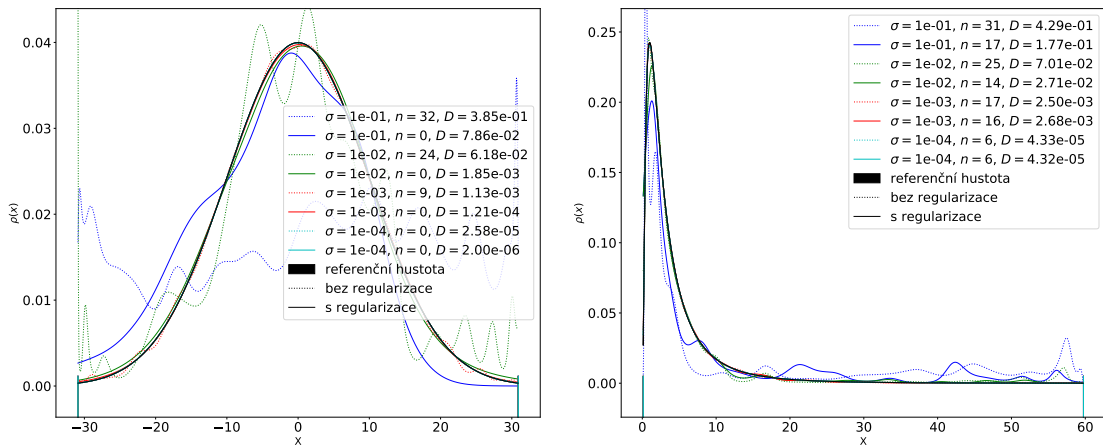
Pro lognormální rozdělení (obrázek 6.3b) je při $\sigma = 0.1$ alespoň rámcově zachycen tvar i bez regularizace. S jejím použitím se aproximace zlepšuje, přesto obsahuje výrazné zvlnění.

U rozdělení five-fingers (obrázek 6.3d) dochází k poměrně špatným aproximacím i pro σ nižší než 0.1. Projevuje se tu již zmíněný vliv odstranění velkého množství momentů. Pro variantu s regularizací jsou již výsledky srovnatelné s výchozím předpokmáněním.

V případě Cauchyho rozdělení (obrázek 6.3e) je pro $\sigma = 0.1$ a použitou regularizací rámcově zachycen referenční tvar. Pro nespojitě rozdělení (obrázek 6.3f) jsou výsledky pro $\sigma = 0.1$ srovnatelné jak s regularizací, tak bez ní.

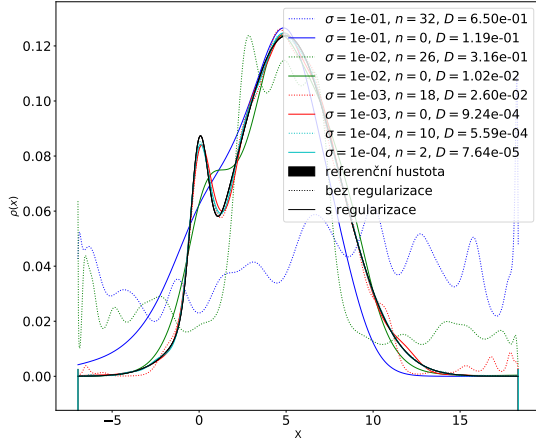
Zvolené předpokmáněním nemá vliv na efekt použité regularizace (v porovnání s výchozím předpokmáněním). Pro $\sigma = 0.1$ a MEM bez regularizace se ve většině případů nedaří zachytit tvar hustoty pravděpodobnosti, výjimkou je zde lognormální rozdělení, které má dlouhý chvost. Pro MEM s regularizací dochází k výrazně lepšímu zachycení tvaru rozdělení.

Smysl tohoto předpokmáněním je v tom, že umožňuje pro velká σ stanovení malé váhy regularizace, aniž by došlo k selhání numerického řešiče. Pro $\sigma \leq 0.01$ jsou výsledky bez regularizace srovnatelné s použitím výchozího předpokmáněním, výjimku zde tvoří rozdělení five-fingers.

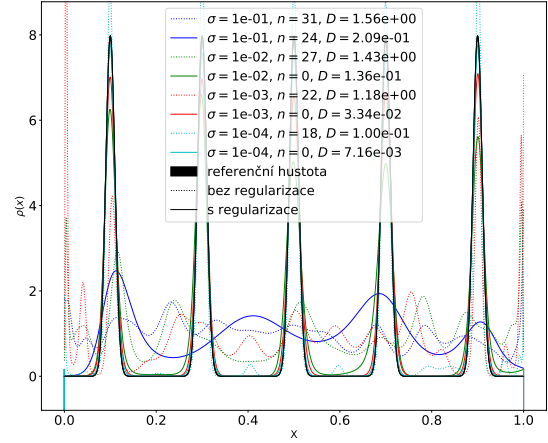


(a) Normální rozdělení

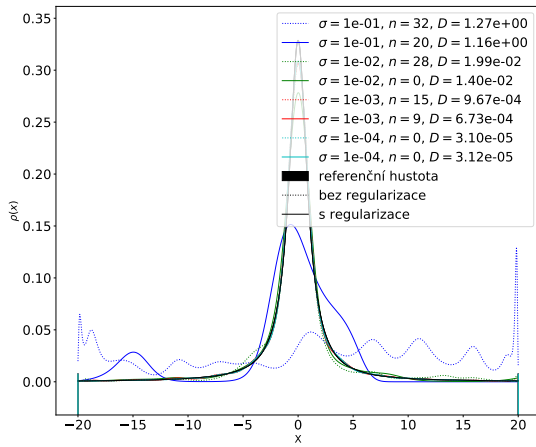
(b) Lognormální rozdělení



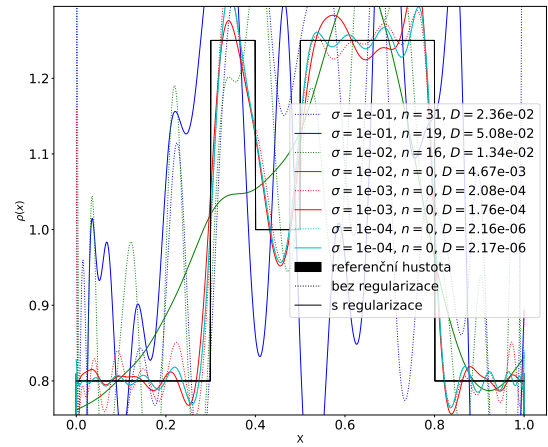
(c) Rozdělení two-gaussians



(d) Rozdělení five-fingers



(e) Cauchyho rozdělení



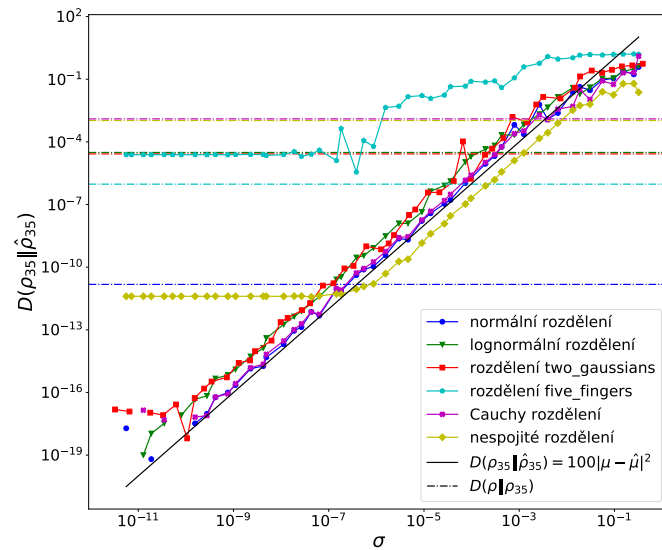
(f) Nespojité rozdělení

Obrázek 6.3: Rekonstrukce rozdělení pravděpodobnosti s regularizací a bez ní, použito je předpokládání pomocí PCA. Pro každou z chyb $\sigma = [10^{-1}, \dots, 10^{-4}]$ je uveden tečkovanou čarou výsledek pro MEM bez regularizace a plnou čarou výsledek MEM s regularizací, n je počet oříznutých momentů a D je KL divergence $D(\rho_{35} \|\hat{\rho}_{35})$.

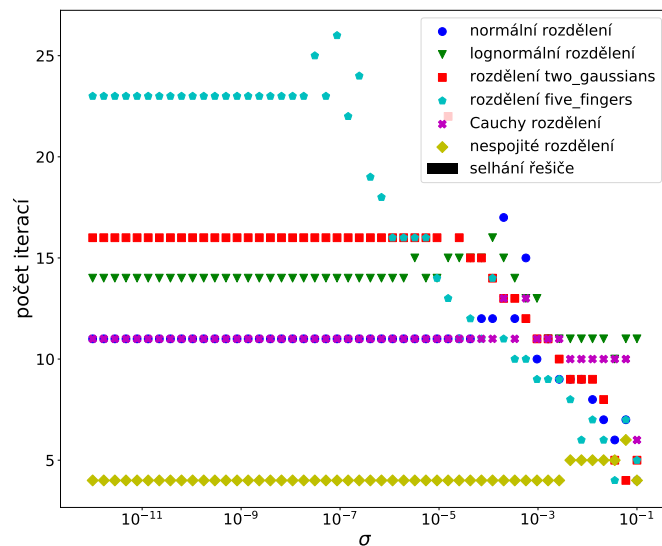
Dále je pro toto předpokládání ukázán na obrázku 6.4 vývoj KL divergence pro různé hodnoty chyby momentů σ , a také vývoj iterací numerického řešiče MEM. V případě KL divergence dochází k poklesu $D(\rho_{35} \|\hat{\rho}_{35})$, podobně jako na obrázku 5.4 v kapitole 5. Pro rozdělení five-fingers se zde pokles zastavuje. Pro malé hodnoty $\sigma \lesssim 10^{-8}$ je již $D(\rho_{35} \|\hat{\rho}_{35})$ v podstatě konstantní. To je způsobeno oříznutím 7 momentových funkcí, což je více než v případě výchozího předpokládání, tam docházelo k postupnému poklesu oříznutých momentů. Toto oříznutí bylo zapříčiněno příliš malými hlavními komponentami, v tomto případě by bylo vhodné zvolit τ menší než je výchozí hodnota 10^{-5} (viz část 3.3).

Obrázek 6.5 ukazuje počty potřebných iterací ke konvergenci numerického řešiče. S narůstajícím σ dochází ve většině případů ke snížení počtu iterací. Je to dáno velkým počtem oříznutých momentových funkcí. Pro tato velká σ převládá v předsta-

veném algoritmu první možnost tzn. ponechání prvních k komponent, které popisují přibližně 100% rozptylu původních momentových funkcí. Ve všech případech dochází ke konvergenci numerického řešení.



Obrázek 6.4: KL divergence pro různé chyby momentů σ a předpodmínění pomocí PCA. Vodorovnou čarou je zobrazena chyba aproximace $D(\rho \|\rho_{35})$, přímka $D(\rho_{35} \|\hat{\rho}_{35}) = 100|\mu - \hat{\mu}|^2$ znázorňuje teoretický předpoklad chyby odhadu.



Obrázek 6.5: Iterace algoritmu pro různé chyby momentů σ a předpodmínění pomocí PCA

7 Rekonstrukce rozdělení pomocí spline funkcí

V této části je uvedena další možnost rekonstrukce rozdělení pravděpodobnosti. Jedná se o interpolaci distribuční funkce pomocí spline křivek. Zatímco přímým výsledkem metody maximální entropie je funkce hustoty pravděpodobnosti. Tak tato metoda stojí na odhadu střední hodnoty indikátorové funkce, pomocí níž je určena přímo distribuční funkce. Následně je vypočítána funkce hustoty pravděpodobnosti.

Náhodná veličina X je v tomto případě již modelována pomocí metod Monte Carlo. Z důvodu zachování přiměřeného rozsahu této práce zde není obecný popis principů metody Monte Carlo a je přistoupeno přímo k její konkrétní implementaci pro stanovení distribuce. Podrobný matematický popis této metody lze nalézt v [19], případně stručnější vysvětlení poskytuje přecházející bakalářská práce [50].

Stanovení distribuční funkce pomocí spline křivek představil ve svém článku M. B. Giles [20], který náhodnou veličinu X modeluje pomocí víceúrovňové metody Monte Carlo. M. B. Giles aplikuje tento přístup na problém řešení stochastických diferenciálních rovnic. Na jeho práci navazují D. Lu a kolektiv [41], kteří přišli s úpravou MLMC algoritmu a aplikovali jej na simulace fyzikálních dějů v ropných ložiscích. Pro výpočet distribuční funkce používají místo spline křivek Lagrangeovy polynomy. Podobným problémem se zabývali také S. Taverniers a D. Tartakovsky [51], kteří rovněž vychází z práce M. B. Gilese a porovnávají jeho přístup s jádrovými odhady hustot (KDE). Tito autoři se zabývali aplikací v úlohách podzemního proudění.

D. Wilson a R. Baker [57] přišli s vlastním algoritmem výpočtu distribuční funkce. I zde se jedná o kombinaci MLMC a odhadů hodnot indikátorové funkce. Autoři se zároveň zabývali vyřešením stejného problému pomocí metody maximální entropie. Aplikací je zde simulace chemického procesu. V závěru autoři konstatují, že pro jejich konkrétní úlohu jsou schopni dosahovat lepších výsledků použitím indikátorové funkce.

Kromě již zmíněných Lagrangeových polynomů, lze místo interpolace pomocí spline křivek použít také jiné přístupy. Jako jsou například Newtonova interpolace nebo interpolace pomocí Taylorova polynomu a další [8]. Velmi používané jsou kubické spline křivky, které pro velkou část praktických úloh dokáží splnit požadované nároky.

Obsahem této kapitoly jsou definice spline funkce a její reprezentace pomocí tzv. B-spline báze funkcí. Následuje samotný popis algoritmu pro určení distribuční funkce. Nejprve je zmíněna varianta pro klasickou metodu Monte Carlo a následně

je tento přístup rozšířen na případ MLMC. V poslední části je popsána realizace výpočtu a použité externí knihovny programovacího jazyka Python. Nakonec jsou porovnány výsledné PDF stanovené metodou maximální entropie a spline interpolací.

7.1 Spline interpolace

Interpolace je proces, jehož cílem je nalézt funkci, která ve stanovených tzv. uzlových bodech nabývá předepsaných hodnot. V některých úlohách není vhodné, aby hledaná funkce procházela přímo předepsanými hodnotami v uzlových bodech. Může se jednat o situace, kdy předepsané hodnoty nejsou stanoveny příliš přesně (jsou například výsledkem měření), pak je vhodnější místo interpolace použít aproximaci. V této kapitole jsou předmětem zájmu především interpolace.

K interpolaci se velmi často používají polynomy. Ty mají celou řadu výhodných vlastností. Patří mezi ně snadné vyhodnocování, diferencovatelnost, integrovatelnost a další [8].

V okamžiku kdy je k dispozici rozsáhlá množina dat, tak začne být polynomiální interpolace problematická. Polynom je vysokého stupně a objevují se oscilace. Východiskem může být použití funkce, která bude po částech polynomem nízkého stupně a jednotlivé části na sebe budou dostatečně hladce navazovat. Jedná se o tzv. spline funkce. Kromě úloh interpolace se využívají také například v počítačové grafice, numerické integraci apod.

Pro další popis se předpokládá, že uzlový vektor $\mathbf{T} = (t_0, t_1, \dots, t_m)$ má své prvky seřazeny neklesajícím způsobem.

Definice 3. B-spline bázové funkce $N_{i,k}(t)$, stupně k jsou na uzlovém vektoru \mathbf{T} definovány rekurzivním předpisem:

$$N_{i,1}(t) = \begin{cases} 1 & t_i \leq t < t_{i+1} \\ 0 & \text{jinde,} \end{cases}$$

pro $k = 1$, a

$$N_{i,k}(t) = \frac{t - t_i}{t_{i+k-1} - t_i} N_{i,k-1}(t) + \frac{t_{i+k} - t}{t_{i+k} - t_{i+1}} N_{i+1,k-1}(t),$$

pro $k > 1$.

Definice 4. Spline funkce stupně k s uzlovým vektorem \mathbf{T} je libovolná lineární kombinace B-spline bázových funkcí stupně k s uzlovým vektorem \mathbf{T} .

Definice 5. Kubickou spline funkcí se nazývá funkce s , která má na daném intervalu $[a, b]$ dvě spojité derivace a na každém jeho podintervalu $[t_{i-1}, t_i]$, $i = 1, \dots, m$, je polynomem stupně 3. Funkce s se označuje jako kubický interpolační spline, jestliže $s(t_i) = y_i$, pro dané y_i .

7.2 Interpolace distribuční funkce

V této části je uveden algoritmus pro interpolaci distribuční funkce $F(X)$ na intervalu $[a, b] \subset \mathbf{R}$. Popsány jsou varianty s klasickou metodou Monte Carlo a také s víceúrovňovou metodou Monte Carlo.

Nejprve jsou připomenuty některé vztahy, které mají důležitý význam pro dále uváděný postup. Pro libovolný bod $q \in \mathbf{R}$ platí, že:

$$F(q) = \int_{-\infty}^q \rho(X) dX,$$

kde $\rho(X)$ je odpovídající funkce hustoty pravděpodobnosti. Využívá se toho, že indikátorová funkce určuje rozdělení pravděpodobnosti náhodné veličiny. Hodnotu distribuční funkce F v bodě q je možné určit jako odhad střední hodnoty indikátorové funkce:

$$F(q) = \mathbb{E}[\mathbb{1}_{(-\infty, q]}(X)] = \int_{-\infty}^{\infty} \mathbb{1}_{(-\infty, q]}(s) \rho(s) ds = \int_{-\infty}^q \rho(s) ds \quad (7.1)$$

a platí

$$\mathbb{1}_{(-\infty, q]}(s) = \begin{cases} 1 & \text{pro } s \in (-\infty, q] \\ 0 & \text{pro } s \in (q, \infty). \end{cases} \quad (7.2)$$

Pro výpočet samotné distribuční funkce F jsou nejprve ze spojitého intervalu $[a, b]$ vybrány ekvidistantní interpolační body $S_h = \{a = q_1 < q_2 < \dots < q_S = b\}$. Výsledná interpolace F_h distribuční funkce F je dána takto:

$$F_h(s) = \sum_{n=1}^S F(q_n) \varphi_n(s) = \sum_{n=1}^S \mathbb{E}[\mathbb{1}_{(-\infty, q_n]}(X) \varphi_n(s)],$$

přičemž $\varphi_n(s)$ jsou v tomto případě B-spline bázové funkce (viz předchozí sekce 7.1).

7.2.1 Aplikace metody Monte Carlo

Pro aplikaci spline interpolace s metodami Monte Carlo je uvažována nová náhodná veličina $Y = P(X)$, kde P je funkce původní náhodné veličiny X . V případě klasické metody Monte Carlo je Y aproximována Y_1 . Nestranný odhad indikátorové funkce v bodě q_n je následující:

$$\hat{\mathbb{1}}_{(-\infty, q_n]}(Y_1) = \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{(-\infty, q_n]}(Y_1^i),$$

kde N je počet nezávislých vzorků z Y_1 .

Interpolace distribuční funkce má následující podobu:

$$\hat{F}_h(s) = \sum_{n=1}^S \hat{\mathbb{1}}_{(-\infty, q_n]}(Y_1) \varphi_n(s). \quad (7.3)$$

Z analýzy chyby $\hat{F}_h(s)$ (viz [41, rovnice 10]) vyplývá, že lze určení distribuční funkce zpřesnit zvětšením N nebo zlepšením aproximace Y pomocí Y_1 . Ať už se zvolí jeden nebo druhý způsob zpřesnění, tak dojde ke zvětšení výpočetní ceny. Proto je vhodným řešením použití MLMC.

7.2.2 Aplikace MLMC

Víceúrovňová metoda Monte Carlo je v principu založena na aproximaci Y pomocí L různě přesných aproximací $\{Y_l, l = 1, \dots, L\}$. V tomto případě Y_1 představuje nejhrubší aproximaci a naopak Y_L je nejjemnější aproximace Y .

Odhad indikátorové funkce v bodě q_n pomocí MLMC:

$$\hat{\mathbb{1}}_{(-\infty, q_n]}(Y_L) = \frac{1}{N_1} \sum_{i=1}^{N_1} \mathbb{1}_{(-\infty, q_n]}(Y_1^i) + \sum_{l=2}^L \left\{ \frac{1}{N_l} \sum_{i=1}^{N_l} [\mathbb{1}_{(-\infty, q_n]}(Y_l^i) - \mathbb{1}_{(-\infty, q_n]}(Y_{l-1}^i)] \right\},$$

kde L je počet úrovní, N_l je počet vzorků na úrovni l . Vztah pro výslednou distribuční funkci 7.3 se nemění, pouze místo Y_1 vstupuje do indikátorové funkce Y_L .

Pro efektivní použití MLMC je třeba dostatečně rychlý pokles rozptylu $\mathbb{V} [\mathbb{1}_{(-\infty, q_n]}(Y_l^i) - \mathbb{1}_{(-\infty, q_n]}(Y_{l-1}^i)]$ napříč úrovněmi. Od té nejnižší $l = 1$, kde je rozptyl největší až po nejvyšší $l = L$, kde je naopak rozptyl nejmenší. Čím rychlejší je tento pokles rozptylu tím MLMC přináší větší úsporu výpočetní ceny v porovnání s MC (viz [41, s. 9648]). V tomto směru se zde projevuje výrazný nedostatek spojený s použitím indikátorové funkce $\mathbb{1}_{(-\infty, q_n]}(Y)$. Z důvodu její nespojitosti dochází k pomalému poklesu rozptylu.

Proto M. B. Giles [21, s. 1] zavádí funkci $g((Y - q_n)/\delta)$, která je spojitou aproximací původní indikátorové funkce. Zde se δ nazývá zjemňovací parametr. Pokud $\delta \rightarrow 0$, pak zde není žádné zjemnění a $g((Y - q_n)/\delta) \rightarrow \mathbb{1}_{(-\infty, q_n]}(Y)$, viz [41, část 4.2]. Čím větší tento parametr je, tím jemněji se mění g v okolí bodu q_n . Funkce g je zvolena tak, aby splňovala následující vlastnosti (viz [20, s. 269]):

1. Výpočetní cena $g(s) \leq C, \forall s \in \mathbf{R}$,
2. g je Lipschitzovsky spojitá
3. $g(s) = \begin{cases} 1 & s < -1 \\ 0 & s > 1 \end{cases}$
4. $\int_{-1}^1 s^j (\mathbb{1}_{(-\infty, 0]}(s) - g(s)) ds = 0$, pro $j = 0, \dots, r - 1$

Pro $r = 3$ je g následující (viz [21, remark 1])

$$g(s) = \begin{cases} 1 & s < -1 \\ 0 & s > 1 \\ \frac{1}{2} + \frac{5s^3 - 9s}{8} & \text{jinde.} \end{cases}$$

Analýzou chyby aproximace $\mathbb{1}_{(-\infty, q_n]}$ pomocí $g((\cdot - q_n)/\delta)$ se zabývá M. B. Giles [20, lemma 2.2], který rovněž podrobuje zkoumání celkovou chybu MLMC odhadu, viz [20, sekce 2.2].

Odhad střední hodnoty $g((Y - q_n)/\delta)$ pomocí MLMC:

$$\hat{g}\left(\frac{Y_L - q_n}{\delta}\right) = \frac{1}{N_1} \sum_{i=1}^{N_1} g\left(\frac{Y_1^i - q_n}{\delta}\right) + \sum_{l=2}^L \left\{ \frac{1}{N_l} \sum_{i=1}^{N_l} \left[g\left(\frac{Y_l^i - q_n}{\delta}\right) - g\left(\frac{Y_{l-1}^i - q_n}{\delta}\right) \right] \right\}.$$

Výpočet distribuční funkce je analogický s případem použití indikátorové funkce 7.3

$$\hat{F}_h(s) = \sum_{n=1}^S \hat{g}\left(\frac{Y_L - q_n}{\delta}\right) \varphi_n(s).$$

Jak již bylo zmíněno, tak volba parametru δ je klíčová pro přesnost aproximace. V rámci této práce bylo zvoleno $\delta = \epsilon^{\frac{1}{r+1}}$ (viz [20, rovnice 2.17]), kde ϵ je zvolená chyba interpolace.

Někteří další autoři, kteří se zabývají touto metodou, často přicházejí s vlastním postupem pro určení hodnoty δ . V článku [41, s. 9650] je výsledné δ určeno na základě provedení prvních N_{init} vzorků v úvodních krocích realizace MLMC algoritmu. Obdobný postup volí i v publikaci [57, s. 9]. Nevýhodou těchto přístupů je úzké provázání provedení MLMC se samotným výpočtem distribuční funkce. To znamená, že tyto metody jsou přímo navrženy s účelem stanovení distribuční funkce, a tak mnohdy postrádají univerzálnost. Ovšem ukazuje se (viz [41, s. 9568, obrázek 9]), že tyto novější přístupy určení δ lépe odpovídají teoretickému odhadu chyby a přináší menší výpočetní cenu MLMC. Obdobná je situace pro určení vhodného počtu interpolačních bodů.

7.3 Porovnání spline interpolace s MEM

V této části je popsán postup výpočtů metody maximální entropie a spline interpolace s použitím metod Monte Carlo. Nejprve je uvedena testovací úloha, která zde byla simulována. Následuje představení použitých externích prostředků a porovnání výsledných PDF získaných pomocí MEM a spline interpolací.

7.3.1 Postup výpočtu

Pro určení odhadů hodnot Y_l pomocí metod Monte Carlo byla použita knihovna `mlmc` [10], která je naprogramována v jazyce Python. Funkce P je modelována umělou simulací:

$$Y = P(X) = X + h\sqrt{10^{-4} + |X|},$$

kde h je simulační krok, který určuje přesnost výsledného odhadu na dané úrovni. V této knihovně lze pro realizaci MLMC nastavit celkový cílový rozptyl \mathbb{V} odhadu Y (více o redukci rozptylu viz [19, s. 4]). V tomto případě je $\mathbb{V} = 10^{-4}$. Jedná se o poměrně malý rozptyl. Ovšem pro některé výpočetně náročné typy simulací

je tento rozptyl maximální reálně spočitatelný. Pro nižší rozptyly je třeba provést mnoho vzorků N_l a může docházet k extrémnímu nárůstu výpočetní ceny.

V případě MEM jsou pomocí knihovny `mlmc` odhadnuty momenty $\hat{\mu}$. A následně je proveden výpočet podle algoritmu popsaného v sekci 2.2. Výpočet je se shodným $\hat{\mu}$ realizován také pro MEM s regularizací (viz část 6.4). Použito je výchozí předpokládání.

Co se týče spline interpolace, tak odhad střední hodnoty indikátorové funkce resp. funkce g je uskutečněn pomocí implementovaného algoritmu, který je uveden v části 7.2. Kód je napsán v programovacím jazyce Python a umožňuje plné začlenění do knihovny `mlmc`.

Parametry B-spline bazových funkcí jsou vypočítány pomocí metody `splrep` z modulu `scipy.interpolate`. Vstupem do `splrep` jsou body definující křivku. V tomto případě se jedná o interpolační body S_h a hodnoty indikátorové funkce resp. funkce g v těchto bodech. Výstupem, který je dále použitý, jsou uzlové body a stupeň spline funkce, preferován je zde kubický spline. Pomocí získaných hodnot a metody `scipy.interpolate.splev` je nakonec vyhodnocena distribuční funkce nebo funkce hustoty pravděpodobnosti.

7.3.2 Porovnání výsledků

Porovnávány jsou tři přístupy: metoda maximální entropie, metoda maximální entropie s regularizací a spline interpolace. V případě MEM s regularizací je eliminována chyba způsobená špatným určením regularizačního parametru. A to tím způsobem, že je optimální regularizační parametr přímo vybrán na základě KL divergence vůči referenčnímu rozdělení. Stejně je postupováno i v případě spline interpolace a volby počtu interpolačních bodů. Zjemňovací parametr δ je zde určen pro $\epsilon = 1 \times 10^{-6}$. Výsledky jsou interpretovány podle KL divergence mezi referenční PDF a provedenou rekonstrukcí PDF.

Obrázek 7.1 obsahuje pro každé srovnávací rozdělení porovnání všech tří přístupů při použití klasické MC a víceúrovňové metody Monte Carlo s pěti úrovněmi. Všechny přístupy jsou v rámci jednotlivé Monte Carlo metody prováděny nad stejnými daty.

V případě normálního rozdělení vychází spline interpolace pro MC (obrázek 7.1a) i MLMC (obrázek 7.1b) lépe než MEM bez regularizace. MEM s regularizací poskytuje v obou případech nejlepší výsledky. Nepatrně horší výsledky pro MLMC v porovnání s MC jsou zapříčiněny tím, že pro MC zde byl pomocí `mlmc` dosažen nižší cílový rozptyl, avšak řádově odpovídající požadovanému $\mathbb{V} = 10^{-4}$.

Pro lognormální rozdělení (MC - obrázek 7.1c a MLMC - obrázek 7.1d) vychází rovněž nejlepší výsledky pro MEM s regularizací. Pomocí spline interpolace se nepodařilo zachytit vrchol hustoty pravděpodobnosti, toho by bylo možné docílit ovšem za cenu nárůstu KL divergence.

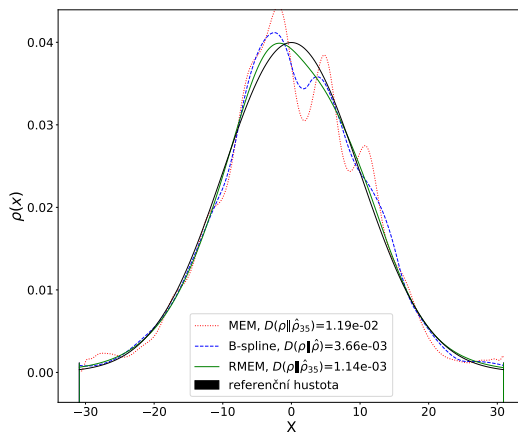
U rozdělení two-gaussians došlo pro MC (obrázek 7.1e) k nejlepšímu přiblížení pomocí spline interpolace. Jedná se pouze o mírný rozdíl v KL divergenci, avšak spline interpolací se podařilo lépe zachytit menší z vrcholů. Toto platí i pro případ MLMC (obrázek 7.1f), kde i přesto vyšla jako nejlepší z hlediska KL divergence

MEM s regularizací.

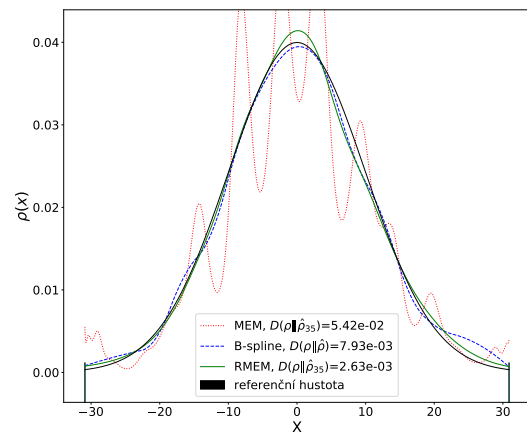
Rekonstrukce Cauchyho rozdělení vychází pro MC (obrázek 7.1i) i MLMC (obrázek 7.1j) srovnatelně. S tím, že opět jsou nejlepší výsledky z hlediska KL divergence pro MEM s regularizací. Nedochozí zde k příliš velkému zlepšení mezi MEM a MEM s regularizací.

Rozdělení five-fingers se daří poměrně dobře určit pomocí MC (obrázek 7.1g). Pro spline interpolace je obtížné zachytit strmé přechody mezi jednotlivými vrcholy, to vede k tomu, že je v těchto oblastech záporná PDF. Při použití MLMC je výsledek spline interpolace velmi špatný, projevuje se zde vliv zjemňovacího parametru δ . Jeho navýšení zde vede ke zlepšení rekonstrukce PDF, ovšem pro některá jiná rozdělení má tato změna δ negativní vliv. MEM s regularizací i zde poskytuje nejlepší výsledky. Velmi podobné problémy nastávají u nespojitého rozdělení (MC - obrázek 7.1k, MLMC - obrázek 7.1l). Avšak objevuje se zde poměrně dobrá aproximace pro MEM s regularizací.

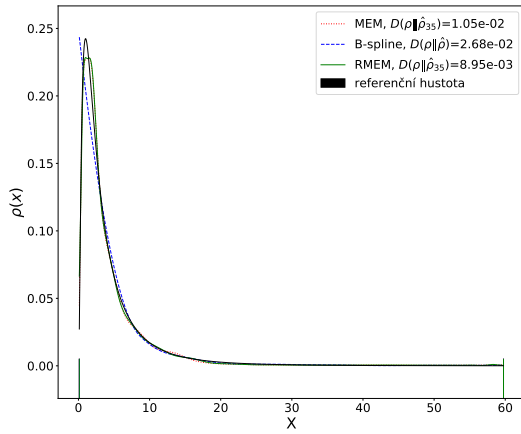
V případě jednoduchých tvarů rozdělení může spline interpolace předčít metodu maximální entropie bez regularizace. V případě složitějších tvarů PDF a použití spline interpolace s MLMC je vhodné lépe určit zjemňovací parametr δ . Při použití MEM s regularizací dochází ve valné většině případů k nejlepšímu přiblížení k referenční hustotě. V některých případech mohou být nepatrně horší výsledky pro MLMC způsobeny méně přesnými odhady z knihovny `mlmc`.



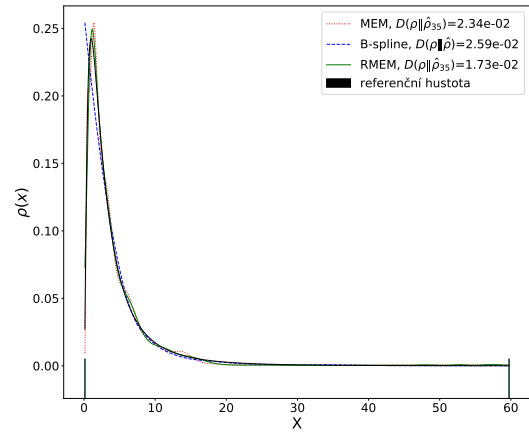
(a) MC, normální rozdělení



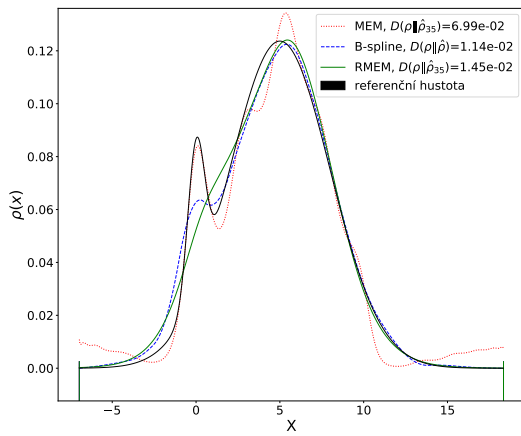
(b) MLMC, normální rozdělení



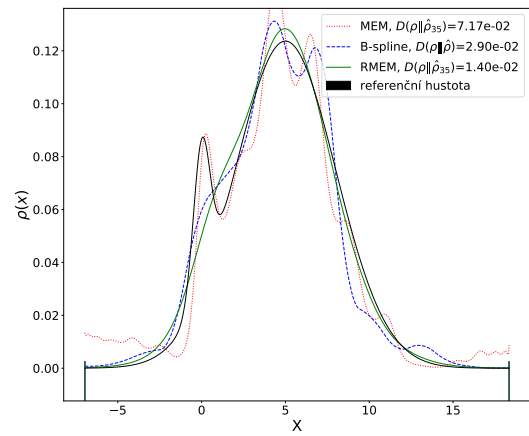
(c) MC, lognormální rozdělení



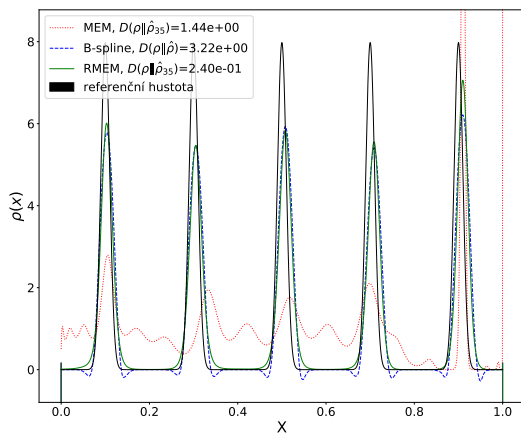
(d) MLMC, lognormální rozdělení



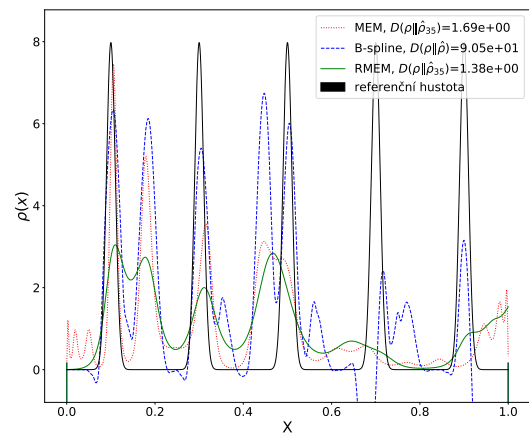
(e) MC, rozdělení two-gaussians



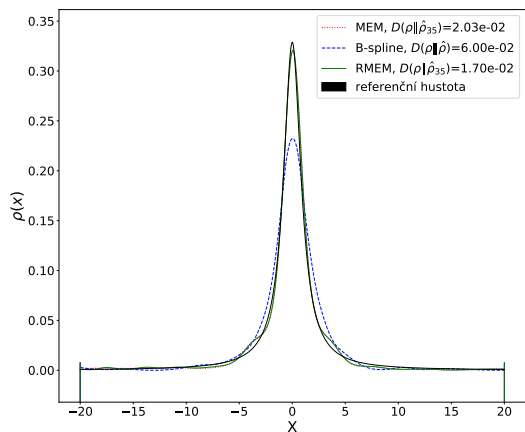
(f) MLMC, rozdělení two-gaussians



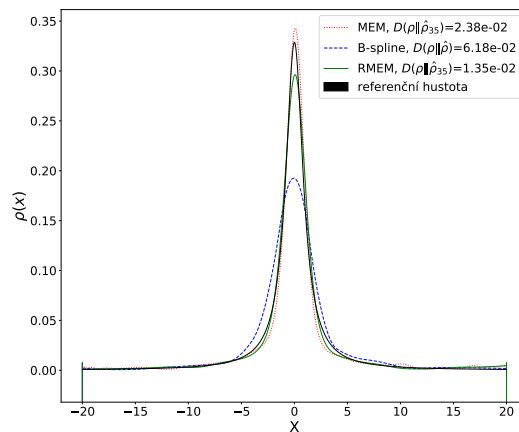
(g) MC, rozdělení five-fingers



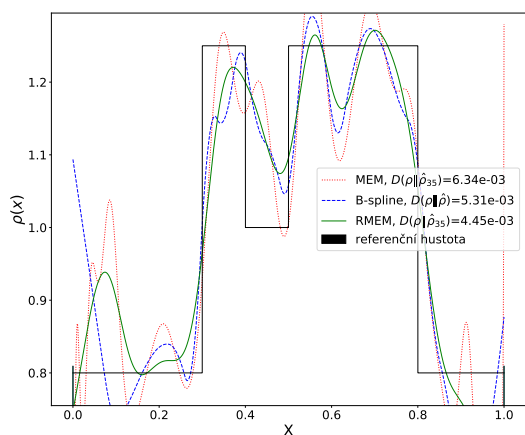
(h) MLMC, rozdělení five-fingers



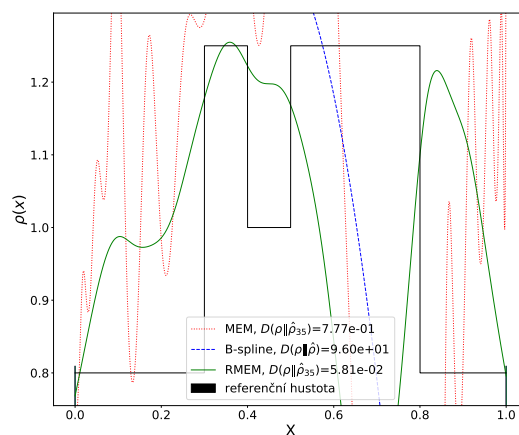
(i) MC, Cauchyho rozdělení



(j) MLMC, Cauchyho rozdělení



(k) MC, nespojité rozdělení



(l) MLMC, nespojité rozdělení

Obrázek 7.1: Porovnání výsledků rekonstrukce rozdělení pravděpodobnosti pomocí metody maximální entropie a spline interpolací. V levém sloupci jsou uvedeny výsledky při použití MC. V pravém sloupci je použita MLMC, jedná se o pětiúrovňovou variantu. Pro každé z rozdělení je tečkovaná čarou zobrazen výsledek MEM, čárkovaně je vykreslen výsledek B-spline interpolace, plnou čarou je vyobrazen výsledek pomocí MEM s regularizací (označeno jako RMEM).

8 Rozšíření MEM na bivarietní rozdělení pravděpodobnosti

V některých případech nestačí modelovat pouze jednu náhodnou veličinu, ale je zapotřebí zachytit vlastnosti více náhodných veličiny, které na sobě mohou záviset. Pak se hovoří o tzv. multivarietních rozděleních pravděpodobnosti. Z hlediska rekonstrukce PDF a použití metody maximální entropie se jedná o obdobný problém jako v případě jednorozměrné náhodné veličiny. Ovšem stává se výpočetně náročné až nemožné rekonstruovat mnohorozměrná rozdělení.

Reálně lze za rozumnou výpočetní cenu aproximovat bivarietní rozdělení pravděpodobnosti. Jedná se o dvourozměrné rozdělení pravděpodobnosti (viz [40]). V tomto případě jsou uvažovány dvě spojité náhodné veličiny, jejichž rozdělení lze popsat funkcí sdružené pravděpodobnosti.

Cílem je najít nezápornou funkci hustoty pravděpodobnosti $\rho(X, Y)$ pro náhodný vektor (X, Y) , kde X a Y jsou náhodné veličiny definované na stejném pravděpodobnostním prostoru. Toto rozdělení lze vyjádřit pomocí spojité sdružené distribuční funkce: $F(x, y) = P(X < x) \wedge P(Y < y)$.

Následuje popis metody maximální entropie rozšířené na bivarietní rozdělení. Dále jsou uvedeny dvě testovací PDF, na kterých je ověřena správnost navrženého algoritmu.

8.1 Metoda maximální entropie pro bivarietní rozdělení

Metodu maximální entropie lze přirozeně rozšířit do více dimenzí. Představen je popis odhadu PDF pro bivarietní rozdělení. Jedná se o analogický zápis s tím, který byl uveden pro univarietní rozdělení v kapitole 2.

Úkolem je najít sdruženou funkci hustoty pravděpodobnosti, která bude mít maximální entropii

$$H(\rho) = - \iint_{\Omega} \rho(x, y) \ln(\rho(x, y)) dx dy \quad \text{za podmínek}$$
$$\iint_{\Omega} \rho(x, y) dx dy = 1,$$
$$\iint_{\Omega} \phi_{r,s}(x, y) \rho(x, y) dx dy = \mu_{r,s}, \quad r = 1, \dots, R \text{ a } s = 1, \dots, S,$$

kde $\phi_{r,s}(x, y) = \phi_r(x)\phi_s(y)$ jsou funkce pro výpočet sdružených momentů a $\mu_{r,s}$ jsou odhady středních hodnot těchto sdružených momentů. I v tomto případě lze vyjádřit ρ ve formě rozdělení exponenciálního typu

$$\rho(x, y) = \exp \left[\sum_{r=1}^R \sum_{s=1}^S \lambda_{r,s} \phi_{r,s}(x, y) \right].$$

Bivarietní PDF lze rovněž porovnávat pomocí KL divergence (viz [54, s. 266]). A tak je možné využít i všech vlastností KL divergence, které jsou popsány v části 2.1.2. Algoritmus řešení MEM se téměř neliší od verze pro univarietní rozdělení (viz část 2.1.4). Hlavním rozdílem je řešení dvojného integrálu místo jednoduchého.

Je zřejmé, že po teoretické stránce je rozšíření metody maximální entropie na variantu pro rekonstrukci bivarietních rozdělení analogické s původní verzí. Z programátorského hlediska bylo třeba upravit stávající kód pro výpočet momentů $\boldsymbol{\mu}$. Také bylo nutné upravit funkce pro výpočet hodnot funkcionálu $F(\boldsymbol{\lambda})$, gradientu $\mathbf{G}(\boldsymbol{\lambda})$ a Hessovy matice $\mathbf{H}(\boldsymbol{\lambda})$, které se používají v rámci numerického řešení MEM.

8.2 Otestování MEM na bivarietních rozděleních

V této části jsou popsány dvě bivarietní rozdělení. Jedná se o variantu normálního rozdělení a rozdělení two-gaussians z kapitoly 4. S těmito rozděleními je následně provedeno otestování správnosti výsledků rozšířené MEM. K porovnání referenčních rozdělení s rekonstruovanými PDF je opět použita KL divergence.

Bivarietní normální rozdělení

Prvním testovacím rozdělení je bivarietní normální rozdělení, které je pro stížení úlohy rotované.

$$\rho(x, y) = \mathcal{N}(x, y | \boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad x, y \in (-\infty, \infty),$$

kde $\boldsymbol{\mu} = [1, 0]$ jsou střední hodnoty marginálních rozdělení, kovarianční matice $\boldsymbol{\Sigma} = \mathbf{R}\mathbf{S}\mathbf{R}^T$ vznikla pomocí matice rotace \mathbf{R} s přeškálováním \mathbf{S}

$$\mathbf{R} = \begin{bmatrix} \cos(\pi/3) & -\sin(\pi/3) \\ \sin(\pi/3) & \cos(\pi/3) \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}.$$

V tomto případě by mělo docházet k dobré aproximaci již pro malý počet momentových funkcí.

Bivarietní rozdělení two-gaussians

Těžším případem pro algoritmus MEM je obměna původního rozdělení two-gaussians (viz kapitola 4). To se v tomto případě skládá ze dvou bivarietních normálních rozdělení.

$$\rho(x, y) = \frac{8}{10} \mathcal{N}(x, y | \boldsymbol{\mu}_1, \boldsymbol{\Sigma}) + \frac{2}{10} \mathcal{N}(x, y | \boldsymbol{\mu}_2, \boldsymbol{\Sigma}), \quad x, y \in (-\infty, \infty),$$

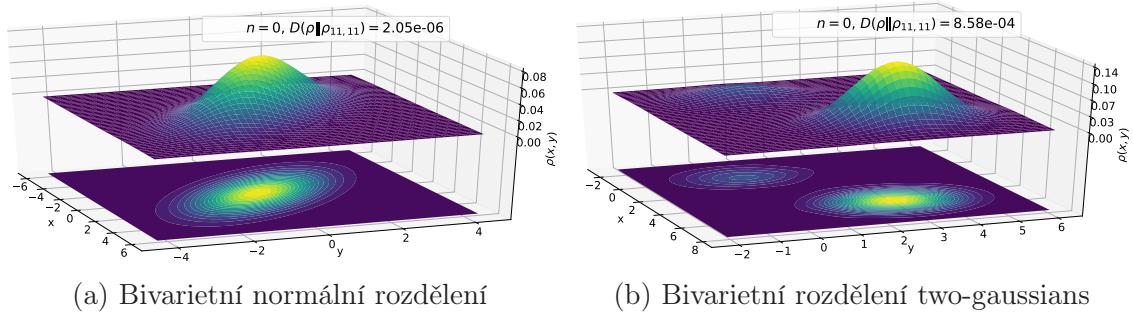
kde $\boldsymbol{\mu}_1 = [5, 3]$, $\boldsymbol{\mu}_2 = [0, 0.5]$ a kovarianční matice $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

Zde se předpokládá, že bude třeba pro dobré přiblížení k referenční PDF velký počet momentových funkcí.

8.2.1 Výsledky testů

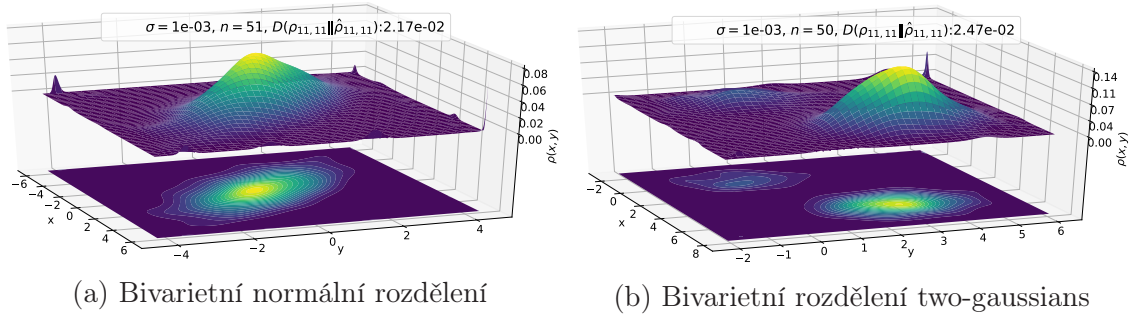
Následuje popis výsledků MEM pro uvedená bivariální rozdělení. Použito bylo předpokmínění pomocí PCA (viz část 3.3). Popsány jsou dvě varianty. První je použití daného počtu přesných momentů $\boldsymbol{\mu}$. V druhém případě jsou uplatněny momenty $\hat{\boldsymbol{\mu}}$ s přičtenou chybou odpovídající Gaussovskému šumu. Podrobně byla tato technika uvedena v části 5.1.

Obrázek 8.1 obsahuje zrekonstruovanou PDF pro počet momentů $R = 11$ a $S = 11$, které jsou určeny přesně. Pro bivariální normální rozdělení (obrázek 8.1a) je již s tímto počtem momentů dosaženo poměrně nízké KL divergence $D(\rho||\rho_{R,S}) \approx 2.1 \times 10^{-6}$. U bivariálního rozdělení two-gaussians (obrázek 8.1b) je KL divergence pro stejně zvolená R a S zhruba o dva řády větší $D(\rho||\rho_{R,S}) \approx 8.6 \times 10^{-4}$. Podobné úrovně $D(\rho||\rho_{R,S})$ jako pro bivariální normální rozdělení lze dosáhnout až pro například $R = 27$, $S = 27$.



Obrázek 8.1: Rekonstruovaná bivariální rozdělení s přesnými momenty, $R = 11$, $S = 11$, n - počet oříznutých momentů, $D(\rho||\rho_{11,11})$ - KL divergence vůči referenční PDF.

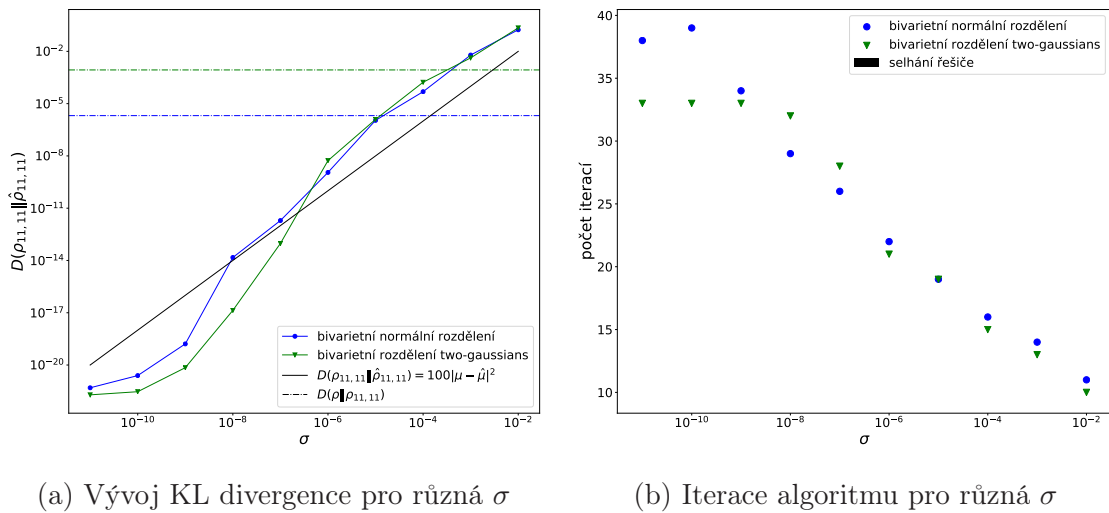
V případě použití zašuměných momentů (obrázek 8.2) se ukazuje, že rovněž u bivariálních rozdělení dochází ke zvlnění tvaru hustoty. Toto zvlnění se začíná s přibývajícím chybou momentů nejdříve projevovat v oblastech rozdělení, kde se vyskytují data s malou pravděpodobností.



Obrázek 8.2: Rekonstruovaná bivarietní rozdělení se zašuměnými momenty, $R = 11$, $S = 11$, n - počet oříznutých momentů, $D(\rho_{11,11} \|\hat{\rho}_{11,11})$ - KL divergence vyjadřující chybu odhadu.

Pro lepší představu o vývoji chyby odhadu KL divergence a numerické stabilitě řešení jsou přidány další dva obrázky. Obrázek 8.3a ukazuje, že i v tomto případě dochází k poklesu $D(\rho_{11,11} \|\hat{\rho}_{11,11})$ se snížením chyby $\hat{\mu}$. Sklon tohoto poklesu přibližně odpovídá teoretickému předpokladu, který je v obrázku vyznačen přímkou $D(\rho_{11,11} \|\hat{\rho}_{11,11}) \leq 100\|\mu - \hat{\mu}\|^2$. Čárkovanou horizontální čarou je uvedena chyba aproximace $D(\rho \|\rho_{11,11})$.

Vývoj iterací numerického řešiče ilustruje obrázek 8.2b. Maximální počet iterací byl stanoven na 40. Ve všech případech nastala konvergence numerického řešení.



Obrázek 8.3: Bivarietní rozdělení - KL divergence a počty iterací numerického řešiče v závislosti na chybě momentů

Vzhledem k přítomnosti zvlnění v rekonstruovaných rozděleních by i zde bylo vhodné zavést obdobnou regularizaci jako v případě univarietních rozdělení. Z časových důvodů k tomu již v rámci této diplomové práce nedošlo. Nicméně byla provedena rešerše používaných přístupů a vytipováno možné řešení, které představili R. Koenker a I. Mizera [37, sekce 2.2]. Jeho princip tkví v potlačování „hrubosti“ povrchu rozdělení.

9 Závěr

Tato diplomová práce se zabývala vybranými neparametrickými metodami rekonstrukce rozdělení pravděpodobnosti. Největší pozornost byla věnována metodě maximální entropie, která využívá tzv. zobecněné momenty. Jejich hodnoty mohou být efektivně odhadovány pomocí víceúrovňové metody Monte Carlo. Pomocí ní lze v praktických úlohách dosahovat odhadů s nejnižší chybou na úrovni $\sigma \approx 0.001$. Z toho důvodu byl hlavní důraz kladen na správnost rekonstrukce rozdělení při těchto úrovních chyby odhadů.

Na vybraných srovnávacích rozděleních pravděpodobnosti byla pomocí KL divergence zanalyzována přesnost výsledků metody maximální entropie. Bylo ověřeno, že KL divergence klesá se zvyšujícím se počtem momentových funkcí. Při jejich konstantním počtu dochází ke snížení KL divergence se zpřesňujícím se odhadem momentů. Z provedených pozorování vyplynulo, že chyba v odhadu momentů vede ke zvlnění tvaru hustoty pravděpodobnosti, které se zvýrazňuje s rostoucím σ .

S cílem potlačit toto zvlnění byla do funkcionálu metody maximální entropie přidána regularizace. Konkrétně se jednalo o penalizaci druhé derivace logaritmu funkce hustoty pravděpodobnosti. Navržen byl rovněž algoritmus pro určení optimálního regularizačního parametru. Na provedených testech se ukázalo, že lze aplikaci zvolené regularizace zmírnit až úplně potlačit přítomnost zvlnění. Nejvýrazněji se tento efekt projevil pro největší zkoumanou chybu momentů $\sigma = 0.01$. Na případu lognormálního rozdělení se ukázalo, že může být do budoucna vhodné použít adaptivní regularizaci, která by byla schopná zhladit pouze část tvaru rozdělení. Rovněž bude účelné dále zdokonalovat algoritmus pro výpočet regularizačního parametru.

Vzhledem k tomu, že numerické řešení selhávalo pro $\sigma > 0.01$. Tak bylo přistoupeno k modifikaci výchozího předpokládání. Snahou zůstalo zajistit pozitivní definitnost kovarianční matice, která je klíčová pro celý výpočet. Původně toho bylo dosaženo prostřednictvím odstranění vlastních čísel této matice, které byly menší než σ . Návrh nového předpokládání využívá analýzu hlavních komponent. Podstatou je zde ponechání pouze tolika hlavních komponent, které popisují přibližně celkový rozptyl obsažený v momentech. Díky tomuto přístupu lze rekonstruovat rozdělení, pro něž byly momenty odhadnuty se $\sigma > 0.01$. Kvalita výsledné hustoty pravděpodobnosti závisí především na chybě momentů a velikosti regularizace. Navržené předpokládání, které především zaručuje konvergenci numerického řešiče, ovlivňuje chybu aproximace.

Kromě metody maximální entropie byla pozornost směřována také na interpolaci distribuční funkce pomocí B-spline bázových funkcí. Realizován byl algoritmus založený na Monte Carlo odhadu střední hodnoty indikátorové funkce včetně jeho

rozšíření na víceúrovňovou metodu Monte Carlo používající hladkou aproximaci indikátorové funkce. Došlo na porovnání MEM bez regularizace, MEM s regularizací a spline interpolace. Z pohledu KL divergence se jeví, na základě provedený experimentů, jako nejlepší použití MEM s regularizací. Spline interpolace poskytuje pro jednoduché tvary rozdělení lepší nebo srovnatelné výsledky, než je tomu při použití MEM bez regularizace. Avšak pro komplexnější tvary hustot pravděpodobnosti s aplikací MLMC dochází vlivem použité volby zjemňovacího parametru ke špatnému určení hustoty pravděpodobnosti. Implementovanou metodu spline interpolace by bylo vhodné do budoucna rozšířit o nejnovější přístupy, které sofistikovaněji určují klíčové parametry algoritmu.

Metoda maximální entropie byla rozšířena o možnost rekonstrukce bivariálních rozdělení pravděpodobnosti. V tomto případě již nedošlo na implementaci regularizace. Přesto byl představen směr, kterým by bylo možné tento úkol do budoucna řešit.

Literatura

- [1] ANDREWS, Larry C. *Special Functions of Mathematics for Engineers*. 2nd ed. 1000 20th Street, Bellingham, WA 98227-0010 USA: SPIE, 1997. DOI: 10.1117/3.270709. ISBN 9780819483713.
- [2] AOKI, Yasunori, Rikard NORDGREN a Andrew C. HOOKER. Preconditioning of Nonlinear Mixed Effects Models for Stabilisation of Variance-Covariance Matrix Computations. *The AAPS Journal*. 2016, **18**(2), 505-518. DOI: 10.1208/s12248-016-9866-5. ISSN 1550-7416. Dostupné také z: <http://link.springer.com/10.1208/s12248-016-9866-5>
- [3] ASTER, Richard C., Brian BORCHERS a Clifford H. THURBER. *Parameter estimation and inverse problems*. 3rd ed. Cambridge, MA: Elsevier, 2019. ISBN 978-0-12-804651-7.
- [4] BARRON, Andrew R. a Chyong-Hwa SHEU. Approximation of Density Functions by Sequences of Exponential Families. *The Annals of Statistics*. 1991, **19**(3), 1347-1369. Dostupné také z: <https://projecteuclid.org/euclid.aos/1176348252>
- [5] BICKEL, Peter J., Bo LI, Alexandre B. TSYBAKOV, et al. Regularization in statistics. *Test*. 2006, **15**(2), 271-344. DOI: 10.1007/BF02607055. ISSN 1133-0686. Dostupné také z: <http://link.springer.com/10.1007/BF02607055>
- [6] BIERIG, Claudio a Alexey CHERNOV. Approximation of probability density functions by the Multilevel Monte Carlo Maximum Entropy method. *Journal of Computational Physics*. 2016, **314**, 661-681. DOI: 10.1016/j.jcp.2016.03.027. ISSN 00219991. Dostupné také z: <https://linkinghub.elsevier.com/retrieve/pii/S0021999116001790>
- [7] BÖCKMANN, Christine. Hybrid regularization method for the ill-posed inversion of multiwavelength lidar data in the retrieval of aerosol size distributions. *Applied Optics*. 2001, **40**(9). DOI: 10.1364/AO.40.001329. ISSN 0003-6935. Dostupné také z: <https://www.osapublishing.org/abstract.cfm?URI=ao-40-9-1329>
- [8] BOOR, Carl de. *A practical guide to splines*. Rev. ed. New York: Springer-Verlag, 2001. Applied mathematical sciences. ISBN 0-387-95366-3.

- [9] BUEHLMANN, Peter a S. A. van de GEER. *Statistics for high-dimensional data: methods, theory and applications*. New York: Springer, c2011. Springer series in statistics. ISBN 3642201911.
- [10] BŘEZINA, Jan, Martin ŠPETLÍK a Klára ŠTEKLOVÁ. MLMC [software]. Duben 2019. [cit 8.5.2020]. Dostupné z: <https://pypi.org/project/mlmc/>
- [11] CLASON, Christian, Barbara KALTENBACHER a Elena RESMERITA. Regularization of Ill-Posed Problems with Non-negative Solutions. BAUSCHKE, Heinz H., Regina S. BURACHIK a D. Russell LUKE, ed. *Splitting Algorithms, Modern Operator Theory, and Applications*. Cham: Springer International Publishing, 2019, 2019-11-07, s. 113-135. DOI: 10.1007/978-3-030-25939-6_5. ISBN 978-3-030-25938-9. Dostupné také z: http://link.springer.com/10.1007/978-3-030-25939-6_5
- [12] CONN, Andrew R., Nicholas I. M. GOULD a Philippe L. TOINT. *Trust Region Methods*. Philadelphia PA: Society for Industrial and Applied Mathematics, 2000. ISBN 978-0-89871-460-9.
- [13] DUDÍK, Miroslav. *Maximum entropy density estimation and modeling geographic distributions of species*. 2017. Disertace. Princeton University, Department of computer science.
- [14] ENGL, Heinz W., Martin HANKE a Andreas NEUBAUER. *Regularization of Inverse Problems*. Dordrecht: Springer Netherlands, 1996. DOI: 10.1007/978-94-009-1740-8. ISBN 978-0-7923-6140-4.
- [15] EVANS, D. J. The Use of Pre-conditioning in Iterative Methods for Solving Linear Equations with Symmetric Positive Definite Matrices. *IMA Journal of Applied Mathematics*. 1968, **4**(3), 295-314. DOI: 10.1093/imamat/4.3.295. ISSN 0272-4960. Dostupné také z: <https://academic.oup.com/imamat/article-lookup/doi/10.1093/imamat/4.3.295>
- [16] FARMER, Jenny, Donald JACOBS a James P. BRODY. High throughput nonparametric probability density estimation. *PLOS ONE*. 2018, **13**(5). DOI: 10.1371/journal.pone.0196937. ISSN 1932-6203. Dostupné také z: <https://dx.plos.org/10.1371/journal.pone.0196937>
- [17] FARQUHARSON, Colin G. a Douglas W. OLDENBURG. A comparison of automatic techniques for estimating the regularization parameter in non-linear inverse problems. *Geophysical Journal International*. 2004, **156**(3), 411-425. DOI: 10.1111/j.1365-246X.2004.02190.x. ISSN 0956540X. Dostupné také z: <https://academic.oup.com/gji/article-lookup/doi/10.1111/j.1365-246X.2004.02190.x>
- [18] FREUND, Rudolf J., William J. WILSON a Donna L. MOHR. *Statistical Methods*. 3rd ed. Boston: Elsevier, 2010. DOI: 10.1016/C2009-0-20216-9. ISBN 9780123749703.

- [19] GILES, Michael B. Multilevel Monte Carlo methods. *Acta Numerica*. 2015, **24**, 259-328. DOI: 10.1017/S096249291500001X. ISSN 0962-4929. Dostupné také z: https://www.cambridge.org/core/product/identifier/S096249291500001X/type/journal_article
- [20] GILES, Michael B., Tigran NAGAPETYAN a Klaus RITTER. Multilevel Monte Carlo Approximation of Distribution Functions and Densities. *SIAM/ASA Journal on Uncertainty Quantification*. 2015, **3**(1), 267-295. DOI: 10.1137/140960086. ISSN 2166-2525. Dostupné také z: <http://epubs.siam.org/doi/10.1137/140960086>
- [21] GILES, Michael B., Tigran NAGAPETYAN a Klaus RITTER. *Adaptive Multilevel Monte Carlo Approximation of Distribution Functions* [online]. 2017 [cit. 2020-05-05]. DOI: 10.1137/140960086. Dostupné z: <https://arxiv.org/pdf/1706.06869.pdf>
- [22] GOLUB, Gene H., Michael HEATH a Grace WAHBA. Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter. *Technometrics*. 1979, **21**(2), 215-223. DOI: 10.1080/00401706.1979.10489751. ISSN 0040-1706. Dostupné také z: <http://www.tandfonline.com/doi/abs/10.1080/00401706.1979.10489751>
- [23] GOOD, I. J. Non-parametric Roughness Penalty for Probability Densities. *Nature Physical Science*. 1971, **229**(1), 29-30. DOI: 10.1038/physci229029a0. ISSN 0300-8746. Dostupné také z: <http://www.nature.com/articles/physci229029a0>
- [24] GOOD, I. J. a R. A. GASKINS. Nonparametric roughness penalties for probability densities. *Biometrika*. 1971, **58**(2), 255-277. DOI: 10.1093/biomet/58.2.255. ISSN 0006-3444. Dostupné také z: <https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/58.2.255>
- [25] GOU, Wenhui. *Estimating Value-at-Risk using Multilevel Monte Carlo Maximum Entropy method*. Oxford, 2018. Diplomová práce. University of Oxford.
- [26] HANSEN, P. C. The L-Curve and its Use in the Numerical Treatment of Inverse Problems. *Computational Inverse Problems in Electrocardiology*, ed. P. Johnston, *Advances in Computational Bioengineering*. WIT Press, 2000, s. 119-142.
- [27] HANSEN, P. C. The truncated SVD as a method for regularization. *BIT*. 1987, **27**(4), 534-553. DOI: 10.1007/BF01937276. ISSN 0006-3835. Dostupné také z: <http://link.springer.com/10.1007/BF01937276>
- [28] HENDERSON, Shane G. a Barry L. NELSON. Chapter 1 Stochastic Computer Simulation. *Simulation*. Elsevier, 2006, 2006, s. 1-18. Handbooks in Operations Research and Management Science. DOI: 10.1016/S0927-0507(06)13001-7. ISBN 9780444514288. Dostupné také z: <https://linkinghub.elsevier.com/retrieve/pii/S0927050706130017>

- [29] HOLTON, Glyn A. *Value-at-Risk: Theory and Practice* [online]. 2nd ed. 2014 [cit. 2020-05-29]. Dostupné z: <https://www.value-at-risk.net/>
- [30] HOU, Zong-Yi a Qi-Nian JIN. Tikhonov regularization for nonlinear ill-posed problems. *Nonlinear Analysis: Theory, Methods & Applications*. 1997, **28**(11), 1799-1809. DOI: 10.1016/S0362-546X(95)00235-N. ISSN 0362546X. Dostupné také z: <https://linkinghub.elsevier.com/retrieve/pii/S0362546X9500235N>
- [31] CHERNOV, Alexey a Claudio BIERIG. Estimation of probability density functions by the Maximum Entropy Method. In: *GAMM Activity Group on Uncertainty Quantification (AGUQ)* [online]. 2018-03-18 [cit. 2020-04-02]. Dostupné z: <http://www.mathematik.tu-dortmund.de/lsix/events/UQ18/talks/Chernov-Dortmund-2018.pdf>
- [32] JAYNES, E. T. Information Theory and Statistical Mechanics. *Physical Review*. 1957, **106**(4), 620-630. DOI: 10.1103/PhysRev.106.620. ISSN 0031-899X. Dostupné také z: <https://link.aps.org/doi/10.1103/PhysRev.106.620>
- [33] JOLLIFFE, I. T. *Principal Component Analysis*. New York: Springer-Verlag, 2002. Springer Series in Statistics. DOI: 10.1007/b98835. ISBN 0-387-95442-2. Dostupné také z: <https://link.springer.com/book/10.1007/b98835>
- [34] KALTENBACHER, Barbara, Andreas NEUBAUER a Otmar SCHERZER. *Iterative Regularization Methods for Nonlinear Ill-Posed Problems*. Berlin, New York: Walter de Gruyter, 2008. Radon Series on Computational and Applied Mathematics. DOI: 10.1515/9783110208276. ISBN 9783110208276.
- [35] KILMER, Misha E. a Dianne P. O'LEARY. Choosing Regularization Parameters in Iterative Methods for Ill-Posed Problems. *SIAM Journal on Matrix Analysis and Applications*. 2001, **22**(4), 1204-1221. DOI: 10.1137/S0895479899345960. ISSN 0895-4798. Dostupné také z: <http://epubs.siam.org/doi/10.1137/S0895479899345960>
- [36] KOENKER, Roger a Ivan MIZERA. DENSITY ESTIMATION BY TOTAL VARIATION REGULARIZATION. NAIR, Vijay. *Advances in Statistical Modeling and Inference*. WORLD SCIENTIFIC, 2007, 2007-03-21, s. 613-633. DOI: 10.1142/9789812708298_0030. ISBN 978-981-270-369-9. Dostupné také z: http://www.worldscientific.com/doi/abs/10.1142/9789812708298_0030
- [37] KOENKER, Roger a Ivan MIZERA. Penalized triograms: total variation regularization for bivariate smoothing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2004, **66**(1), 145-163. DOI: 10.1111/j.1467-9868.2004.00437.x. ISSN 1369-7412. Dostupné také z: <http://doi.wiley.com/10.1111/j.1467-9868.2004.00437.x>
- [38] KULLBACK, S. a R. A. LEIBLER. On Information and Sufficiency. *The Annals of Mathematical Statistics*. 1951, **22**(1), 79-86. DOI: 10.1214/aoms/1177729694. ISSN 0003-4851. Dostupné také z: <http://projecteuclid.org/euclid.aoms/1177729694>

- [39] LEHMANN, E. L. a George CASELLA. *Theory of Point Estimation*. New York: Springer-Verlag, 1998. Springer Texts in Statistics. DOI: 10.1007/b98854. ISBN 0-387-98502-6.
- [40] LONG, Dou a Roman KRZYSZTOFOWICZ. A Family of Bivariate Densities Constructed from Marginals. *Journal of the American Statistical Association*. 1995, **90**(430). DOI: 10.2307/2291086. ISSN 01621459. Dostupné také z: <https://www.jstor.org/stable/2291086?origin=crossref>
- [41] LU, Dan, Guannan ZHANG, Clayton WEBSTER a Charlotte BARBIER. An improved multilevel Monte Carlo method for estimating probability distribution functions in stochastic oil reservoir simulations. *Water Resources Research*. 2016, **52**(12), 9642-9660. DOI: 10.1002/2016WR019475. ISSN 00431397. Dostupné také z: <http://doi.wiley.com/10.1002/2016WR019475>
- [42] LY, Alexander, Maarten MARSMAN, Josine VERHAGEN, Raoul P.P.P. GRASMAN a Eric-Jan WAGENMAKERS. A Tutorial on Fisher information. *Journal of Mathematical Psychology*. 2017, **80**, 40-55. DOI: 10.1016/j.jmp.2017.05.006. ISSN 00222496. Dostupné také z: <https://linkinghub.elsevier.com/retrieve/pii/S0022249617301396>
- [43] MUELLER, Jennifer L. a Samuli SILTANEN. *Linear and Nonlinear Inverse Problems with Practical Applications*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2012. DOI: 10.1137/1.9781611972344. ISBN 978-1-61197-233-7.
- [44] NOCEDAL, Jorge a Stephen J. WRIGHT. *Numerical optimization*. 2nd ed. New York: Springer, c2006. ISBN 978-0-387-30303-1.
- [45] RAMSAY, J. O. a B. W. SILVERMAN. *Functional data analysis*. 2nd ed. New York: Springer, c2005. ISBN 038740080X.
- [46] RUSSEN, S. C. Numerical analysis in modern scientific computing: an introduction (2nd edn), by Peter Deuffhard and Andreas Hohmann. Pp. 337. £42. 2003. ISBN 0 387 95410 4 (Springer-Verlag). *The Mathematical Gazette*. 2004, **88**(512), 414-414. DOI: 10.1017/S0025557200176016. ISSN 0025-5572. Dostupné také z: https://www.cambridge.org/core/product/identifier/S0025557200176016/type/journal_article
- [47] SHANNON, C. E. A Mathematical Theory of Communication. *Bell System Technical Journal*. 1948, **27**(3), 379-423. DOI: 10.1002/j.1538-7305.1948.tb01338.x. ISSN 00058580. Dostupné také z: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6773024>
- [48] SCHMIDT, Mark, Glenn FUNG a Rómer ROSALES. Fast Optimization Methods for L1 Regularization: A Comparative Study and Two New Approaches. KOK, Joost N., Jacek KORONACKI, Raomon Lopez de MANTARAS,

- Stan MATWIN, Dunja MLADENIČ a Andrzej SKOWRON, ed. *Machine Learning: ECML 2007*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, s. 286-297. Lecture Notes in Computer Science. DOI: 10.1007/978-3-540-74958-5_28. ISBN 978-3-540-74957-8. Dostupné také z: http://link.springer.com/10.1007/978-3-540-74958-5_28
- [49] SILVERMAN, B. W. On the estimation of a probability density function by the maximum penalized likelihood method. *The Annals of Statistics*. 1982, **10**(3), 795-810. Dostupné také z: <https://projecteuclid.org/euclid.aos/1176345872>
- [50] ŠPETLÍK, Martin. *Aplikace víceúrovňové metody Monte-Carlo v hydrogeologii*. Technická univerzita v Liberci, 2018. Bakalářská práce. Technická univerzita v Liberci. Vedoucí práce Jan Březina.
- [51] TAVERNIERS, Soren a Daniel M. TARTAKOVSKY. Estimation of distributions via multilevel Monte Carlo with stratified sampling. *ArXiv* [online]. 2019, **2019**(abs/1906.00126) [cit. 2020-05-29]. Dostupné z: <https://arxiv.org/pdf/1906.00126.pdf>
- [52] THOMPSON, James R. a Richard A. TAPIA. *Nonparametric Function Estimation, Modeling, and Simulation*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1990. DOI: 10.1137/1.9781611971712. ISBN 978-0-89871-261-2.
- [53] TIKHONOV, A. N. On the stability of inverse problems. *Doklady Akademii Nauk SSSR*. 1943, **39**(5), 195-198.
- [54] VERDOOLAEGE, Geert, Yves ROSSEEL, Michiel LAMBRECHTS a Paul SCHEUNDERS. Wavelet-based colour texture retrieval using the kullback-leibler divergence between bivariate generalized Gaussian models. In: *2009 16th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2009, 2009, s. 265-268. DOI: 10.1109/ICIP.2009.5413405. ISBN 978-1-4244-5653-6. Dostupné také z: <http://ieeexplore.ieee.org/document/5413405/>
- [55] WAINWRIGHT, Martin J. a Michael I. JORDAN. Graphical Models, Exponential Families, and Variational Inference. *Foundations and Trends® in Machine Learning*. 2007, **1**(1—2), 1-305. DOI: 10.1561/22000000001. ISSN 1935-8237. Dostupné také z: <http://www.nowpublishers.com/article/Details/MAL-001>
- [56] WANG, Yanfei, Yan CUI a Changchun YANG. Hybrid regularization methods for seismic reflectivity inversion. *GEM - International Journal on Geomatics*. 2011, **2**(1), 87-112. DOI: 10.1007/s13137-011-0014-1. ISSN 1869-2672. Dostupné také z: <http://link.springer.com/10.1007/s13137-011-0014-1>
- [57] WILSON, D. a R. E. BAKER. Multi-level methods and approximating distribution functions. *AIP Advances*. 2016, **6**(7). DOI: 10.1063/1.4960118. ISSN 2158-3226. Dostupné také z: <http://aip.scitation.org/doi/10.1063/1.4960118>

- [58] WU, Ximing. Exponential Series Estimator of multivariate densities. *Journal of Econometrics*. 2010, **156**(2), 354-366. DOI: 10.1016/j.jeconom.2009.11.005. ISSN 03044076. Dostupné také z: <https://linkinghub.elsevier.com/retrieve/pii/S0304407609002942>
- [59] YAMAN, Fatih, Valery G. YAKHNO a Roland POTTHAST. A Survey on Inverse Problems for Applied Sciences. *Mathematical Problems in Engineering*. 2013, **2013**, 1-19. DOI: 10.1155/2013/976837. ISSN 1024-123X. Dostupné také z: <http://www.hindawi.com/journals/mpe/2013/976837/>
- [60] YE, Hailiang, Feilong CAO a Dianhui WANG. A hybrid regularization approach for random vector functional-link networks. *Expert Systems with Applications*. 2020, **140**. DOI: 10.1016/j.eswa.2019.112912. ISSN 09574174. Dostupné také z: <https://linkinghub.elsevier.com/retrieve/pii/S095741741930630X>