

VÝPRAVY DO HLUBIN DAT

Daniel Hlubinka

Klíčová slova: Hloubka dat, kvantilové kontury, směrové kvantily, zobecněná poloprostorová hloubka

Abstrakt: V článku je představena hloubka dat jako zobecnění pojmu uspořádání pro účely mnohorozměrné statistické analýzy. Pro náhodné vektory jsou zopakovány hlavní typy hloubek spolu s jejich vlastnostmi, výhodami i nevýhodami. V druhé části článku jsou diskutovány dvě alternativní definice hloubky, které umožňují oproti běžným hloubkám zvýraznit lokální vlastnosti rozdělení.

Abstract: The data depth is introduced as a generalization of ranking to be used in the multivariate statistical analysis. The most common definitions of depth are recalled together with their properties and both advantages and disadvantages. In the second part of the paper two alternative definitions of depth are discussed. These definitions allow to emphasize, in contrast to most usual depths, the local behaviour of the underlying distribution.

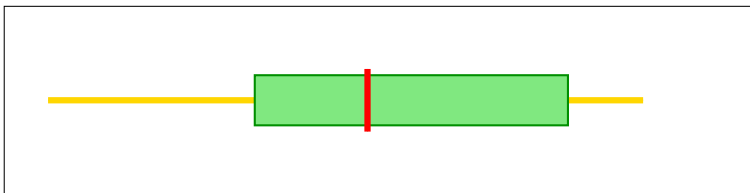
1 Kvantil a problém uspořádání

Ve statistické analýze se občas objeví chvíle, kdy potřebujeme hovořit o oblasti spolehlivosti, predikční oblasti, kritické oblasti. V jednorozměrném případě si obvykle můžeme pomoci kvantily, ve většině případů nejde o složitý problém. Nemusíme ale dlouho vzpomínat, abychom si uvědomili, že kvantil *tak jak jej známe* je pojem úzce vázaný na jednorozměrnou náhodnou veličinu. Většina z nás bude velmi váhat s definicí kvantilu pro dvou- a více-rozměrná rozdělení. Jednou z výjimek je mnohorozměrné normální rozdělení, kde existují přirozeně definované elipsoidy (definované pomocí střední hodnoty a varianční matice), které plně uspokojí naše požadavky na uvedené oblasti spolehlivosti. Jak se ale obejít bez předpokladu normality? V tomto článku budeme hledat nějaký vhodný neparametrický přístup k pojmu kvantil ve více rozměrech. Napřed si ale připomeňme, co víme o jednorozměrných kvantilech a mohlo by při našem hledání být užitečné.

Pro jednorozměrnou náhodnou veličinu X je kvantil definován různými způsoby. Nejčastěji tak, že hodnoty náhodné veličiny kvantil q_α rozděluje na menší a větší; přitom platí zároveň

$$P[X \leq q_\alpha] \geq \alpha, \quad P[X \geq q_\alpha] \geq 1 - \alpha,$$

a hodnota $\alpha \in [0, 1]$ je předem pevně zvolená. Kvantil v \mathbb{R}^p musí být p -rozměrný vektor. Zde je jádro problémů pro náhodné vektory—v prostoru \mathbb{R}^p není přirozené lineární uspořádání. My ale musíme být schopni porovnat



Obrázek 1: Inspirace krabicovým diagramem.

každé dvě hodnoty náhodného vektoru, každé dva vektory, jinak není možné definovat oblast spolehlivosti analogicky jednorozměrnému případu.

Pro jednorozměrnou náhodnou veličinu existuje nástroj, který nám může být inspirací i pro náhodné vektory. Podívejme se na známý krabicový diagram, jednoduchý obrázek patřící do *obyčejné* popisné statistiky. V něm pomocí jednorozměrných kvantilů zobrazíme podstatnou informaci o rozdělení. Zaujme nás *prostřední* bod, *medián*, *vnitřní krabička*, ohraničená dolním a horním kvantilem, a *tykadla—vousy*, ukazující k nejvyšší a nejnižší hodnotě¹. Náš zrak je zde veden nejenom zleva doprava, od malých hodnot k větším, ale spíš zprostřed ke krajům, od středových hodnot ke vnějším. Každé dvě hodnoty tak mohou být porovnány s ohledem na *vzdálenost* od středu, přičemž pojem *vzdálenosti* je pevně svázán s rozdělením; v jednorozměrném případě je přímo definován pomocí kvantilů.

Zde se potkávají dva koncepty. Na jedné straně je to *míra odlehlosti od středu*². Na druhé straně to je naopak určitá *obklopenost* ostatními hodnotami náhodné veličiny. Pro medián je polovina realizací náhodné veličiny větších a polovina menších, je tedy *uprostřed*, nebo *hluboko v rozdělení*. Pro dolní kvartil *jen čtvrtina* realizací náhodné veličiny je menších, je tedy z této strany mnohem méně obklopen rozdělením než medián, je *méně hluboko*.

Takto se budeme dívat i na mnohorozměrné náhodné vektory. Každá hodnota může být s jinou porovnána s ohledem na *odlehlost* od středu. Budeme proto posuzovat, jak *hluboko* se určitý bod nachází v daném rozdělení. Neboli zkoumáme, jak je bod ve výběrovém prostoru *obklopen* ostatními body s ohledem na pravděpodobnostní rozdělení. Tomuto konceptu budeme říkat *hloubka*, *hloubka dat*³. Funkce, které bodům ve výběrovém prostoru přiřazují jejich hloubku, nazveme *hloubky*, případně *hloubkové funkce*. Takových funkcí je jistě celá řada. V příspěvku nejprve pohlédneme na hloubku trochu z nadhledu, ujasníme si čím jsou hloubkové funkce inspirovány, jak se dají třídit do skupin, jaké jsou jejich žádoucí a nežádoucí vlastnosti. Také si představíme několik známých i méně známých hloubkových funkcí i s jejich základními výhodami a nevýhodami. Ve druhé části článku se budeme věnovat dvěma konceptům hloubky, které v současnosti podrobněji zkoumáme.

¹Obvykle se usekávají v určité vzdálenosti od krabičky, ale to každý dobře zná

²V angličtině se většinou používá výraz *outlyingness*

³Anglický termín je *depth*, *data depth*. O hloubce *dat* hovoříme proto, že původně tento koncept vyšel z popisné statistiky a byl aplikován na náhodné výběry.

2 Kvantil v říši více rozměrů

Již jsme uvedli, že pro dvou- a vícerozměrné náhodné vektory místo uspořádání hodnot podle velikosti uijeme lineární kvaziuspořádání podle *odlehlosti*⁴. Formálně postupujeme takto. Označme \mathcal{P} soubor všech pravděpodobnostních rozdělení na prostoru \mathbb{R}^p .

Definice 1. *Hlubkovou funkcí rozumíme zobrazení $H : \mathcal{P} \times \mathbb{R}^p \rightarrow \mathbb{R}^+$.*

Bod $x \in \mathbb{R}^p$ je pro rozdělení $P \in \mathcal{P}$ hlubší než bod $y \in \mathbb{R}^p$, pokud $H(P, x) \geq H(P, y)$. Bod $d \in \mathbb{R}^p$ je pro rozdělení $P \in \mathcal{P}$ nejhlubší, pokud $H(P, d) \geq H(P, x)$ pro všechna $x \in \mathbb{R}^p$. Označme $d_{P,H} = H(P, d)$ hloubku nejhlubšího bodu.

Nebude-li třeba zdůrazňovat rozdělení P budeme psát krátce $H(x)$ místo $H(P, x)$ a d_H místo $d_{P,H}$, případně napíšeme ještě stručněji pouze d .

Dva různé body mohou mít stejnou hloubku, proto jde o *kvaziuspořádání* a každé dva body jsou vzájemně porovnatelné s ohledem na svou hloubku, proto se jedná o *lineární kvaziuspořádání*.

Nejhlubší bod budeme chápat jako mnohorozměrnou obdobu mediánu. Nemusí být dán jednoznačně, stejně jako nemusí být jednoznačný medián v jednorozměrném případě. Díky hloubce lze definovat *centrální* oblasti pro dané rozdělení a *kvantilovou konturu*.

Definice 2. *Pro dané rozdělení P a hlubkovou funkci H definujeme a hlubokou oblast*

$$R_{P,H}(a) := \{x : H(P, x) \geq a\}, 0 \leq a \leq d_{P,H}.$$

Hranice oblasti $R_{P,H}$ se nazývá a -hlubkovou konturou a budeme ji značit $c_{P,H}(a)$. V případě, kdy nemůže dojít k záměně, značíme a hlubokou oblast zkráceně $R(a)$, a hlubkovou konturu obdobně $c(a)$.

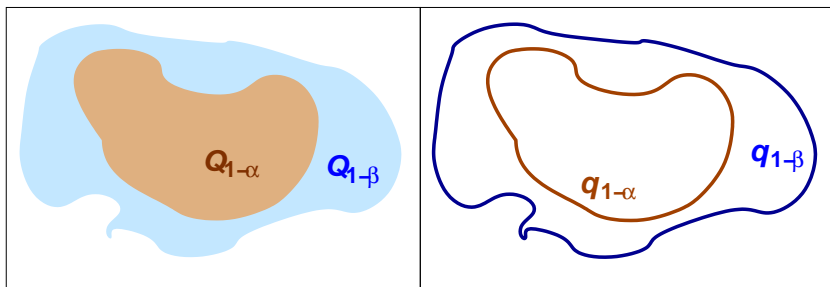
Jak je vidět z definice, a hluboké oblasti, $0 \leq a \leq d_{P,H}$ jsou *úrovňové množiny* hlubkové funkce H . Problematika úrovňových množin je velmi aktuální tematikou současné matematiky a i my se k ní v tomto článku stručně vrátíme o něco později.

Poznámka 1. *Pro jednorozměrnou náhodnou veličinu X a její rozdělení můžeme hloubku bodu x definovat*

$$H(x) = 1 - 2|F(x) - 1/2|, \quad (1)$$

kde $F(x)$ je distribuční funkce X . Hloubka mediánu \tilde{X} je zřejmě 1 a hloubka klesá v obou směrech od mediánu. Body se stejnou hodnotou distribuční funkce mají stejnou hloubku a a hlubokou oblastí je množina

⁴Též můžeme říkat uspořádání ve smyslu vnitřní/vnější, nebo centrální/odlehlý, ...



Obrázek 2: Centrální oblast (vlevo) a kvantilová kontura. $0 < \beta < \alpha < 1$.

$$R(a) = \{x : a/2 \leq F(x) \leq 1 - a/2\}.$$

V případě spojitě a rostoucí distribuční funkce je

$$c(a) = \{F^{-1}(a/2), F^{-1}(1 - a/2)\},$$

a zřejmě

$$P[R(a)] = 1 - a, \quad \text{pro všechna } a \in [0, 1]. \quad (2)$$

Obecně samozřejmě nemusí platit, že pro zvolenou hloubku H je hloubka nejhlubšího bodu rovna 1 pro všechna $P \in \mathcal{P}$. Nemusí ani platit (a pro většinu klasických hloubek opravdu neplatí) vztah (2) $P[R_{P,H}(a)] = 1 - a$. Proto je nezbytné definovat centrální oblasti a kvantilové kontury rozdělení P v hloubce H .

Definice 3. *Definujeme*

$$h(1 - \alpha) = h_{P,H}(1 - \alpha) = \sup\{a : P[R_{P,H}(a)] \geq 1 - \alpha\}.$$

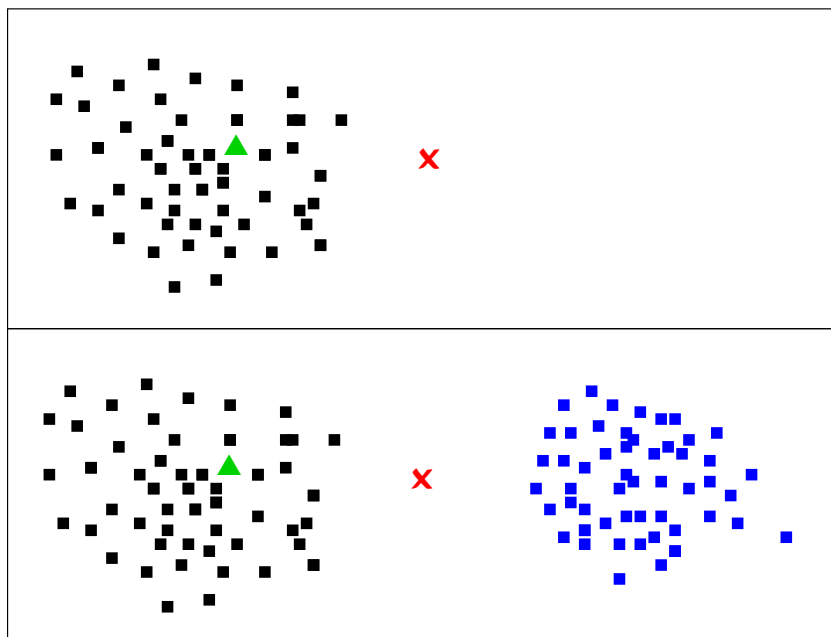
Oblast $Q_{P,H}(1 - \alpha) = R_{P,H}(h_{P,H}(1 - \alpha))$ nazveme $1 - \alpha$ centrální oblastí.

Hranici $q_{P,H}(1 - \alpha)$ oblasti $Q_{P,H}(1 - \alpha)$ nazveme $1 - \alpha$ kvantilovou konturou.

Jak je vidět z definice oblasti $Q(1 - \alpha)$ a hodnoty $h(1 - \alpha)$, bude k jejich určení obvykle nezbytné najít všechny úrovně množiny hloubkové funkce H a jejich pravděpodobnosti. To pochopitelně není snadný úkol. Později si ukážeme, že existuje možnost definovat specifickou hloubku tak, aby hodnota hloubky a pravděpodobnost a hluboké oblasti byly svázané předpisem

$$P[H(P, X) \geq \alpha] = 1 - \alpha \quad \text{pro všechna } \alpha \in [0, 1], \quad (3)$$

čímž se velmi usnadní hledání požadovaných $1 - \alpha$ centrálních oblastí.



Obrázek 3: Je bod ▲ hluboko? Je bod × hluboko?

Klasický jednorozměrný kvantil umožňuje vytvářet centrální množiny ve tvaru intervalů. Někdy jsou upřednostňovány úrovně množiny hustoty, které mohou být tvořeny sjednocením intervalů. Ve vyšších rozměrech se nám však, díky bohatší geometrii, otevírá svět rozmanitějších⁵ tvarů. Čím více rozměrů, tím bohatší tvary centrálních oblastí můžeme uvažovat. Přesto se dnes hlavní proud výzkumu v oblasti hloubky omezuje jen na hloubkové funkce, pro něž jsou centrální oblasti (úrovně množiny) konvexní.

Je dobré si připomenout, že každá redukce informace o rozdělení, v našem případě do několika centrálních oblastí, nám jiným způsobem poukazuje na podstatu náhody v uvažovaném modelu (pokusu). Každá hloubka svým způsobem zvýrazní určité rysy rozdělení, zatímco ostatní vlastnosti rozdělení hodnotu hloubky v podstatě neovlivní. Každá hloubka má také své výpočetní aspekty; jak při výpočtu teoretické hloubky bodu pro dané rozdělení (vesměs musíme používat numerické metody), tak i při výpočtu hloubky bodů v náhodném výběru. Podívejme se na obrázek 3. Na obou obrázcích bychom o bodu ▲ prohlásili, že je hluboko. Bod × je na horním obrázku zřejmě mimo data, co však na dolním?

Uděláme-li si na dolním diagramu v obrázku 3 marginální rozdělení (projekce do os), bude v obou případech bod × blízko mediánu. A dopadneme tak

⁵Mnohdy také jen těžko představitelných.

i v případě projekce na libovolnou přímku. Není proto možné jednoznačně říci, zda \times je či není hluboko. Tento příklad nám ukazuje, že ponětí o tom, který bod *je hluboko* a který *není hluboko* v rozdělení, závisí na tom, co od hloubky očekáváme.

3 Tisíc podob hloubky

Příklad 1. *Hloubka (1) jednorozměrné náhodné veličiny X definovaná pomocí distribuční funkce F_X jako $H(P, x) = 1 - 2|F(x) - 1/2|$ pro $x \in \mathbb{R}$ je funkcí bodu a pravděpodobnostní míry. Na druhou stranu je zřejmé, že pravděpodobnostní míra popisuje vnější prostředí, jinak do definice hloubkové funkce nepřispívá. Nemění totiž samotnou podstatu hloubky. Pro všechna rozdělení je koncept H shodný.*

Motivační příklad 1 nás navádí k tomu, že hloubková funkce by měla vycházet z nějaké obecné myšlenky, přičemž konkrétní pravděpodobnostní rozdělení tento koncept nemění. Jak by takový přístup k hloubkové funkci měl vypadat? Serfling v práci [45] navrhuje řadu *žádoucích vlastností* hloubky, potažmo hloubkové funkce.

3.1 Serflingův průvodce vlastnostmi hloubky

- I. *Afinní invariance hloubky.* Označme A nějakou afinní transformaci \mathbb{R}^p . Má platit $H(P \circ A, A(x)) = H(P, x)$. Hloubka nezávisí na souřadnicovém systému, hloubkové a kvantilové kontury jsou afinně ekvivariantní.
- II. *Maximalita ve středu.* Je-li rozdělení P v nějakém smyslu symetrické kolem $s \in \mathbb{R}^p$, pak $H(P, s)$ je maximální. Neboli s je nejhlubším bodem pro P .
- III. *Přenos symetrie.* Je-li rozdělení P v nějakém smyslu symetrické kolem $s \in \mathbb{R}^p$, pak $H(P, x)$ je odpovídajícím způsobem symetrické kolem s .
- IV. *Pokles podél paprsků.* Je-li s nejhlubší bod pro P , pak funkce $H(P, x)$ je nerostoucí na všech polopřímkách s počátkem v s (ve směru od s).
- V. *Zanedbatelnost v nekonečnu.* Platí $H(P, x) \rightarrow 0$ kdykoliv $\|x\| \rightarrow \infty$.
- VI. *Spojitosť v prostoru.* Zobrazení $x \mapsto H(P, x)$ je spojitě, nebo shora polospojité. Díky této vlastnosti jsou úrovněvé množiny H a tedy i a hluboké oblasti a $1 - \alpha$ centrální oblasti uzavřené.
- VII. *Spojitosť na rozděleních.* Zobrazení $P \mapsto D(P, x)$ je spojitě vzhledem ke slabé konvergenci měř.
- VIII. *Kvazikonkavita.* Úrovněvé množiny (a hluboké oblasti, $1 - \alpha$ centrální oblasti) jsou konkávní.

Uvedeným vlastnostem se budeme v článku průběžně věnovat. V tuto chvíli se pozastavíme jen u požadavků na symetrii. V bodech II a III se hovoří o symetrii *v nějakém smyslu*, jejich splnění je proto závislé na formě uvažované symetrie⁶. V bodě III je dobré specifikovat, jak se symetrie rozdělení P má přenést na (a na jakou) symetrii hloubkové funkce H .

3.2 Račte si vybrat . . .

Na příkladech známých i méně známých hloubek si ukážeme si, že v jednotlivých definicích lze vysledovat několik základních přístupů k mnohorozměrným rozdělením, na kterých jsou hloubkové funkce založené. Kromě definice a některých základních vlastností se pokusíme alespoň trochu pohled⁷ na silné a slabé stránky studované hloubky.

3.2.1 Poloprostorová hloubka Sluší se začít nejklaštější ze všech hloubek (matkou hloubek). Jde o koncept, který představil ve svém článku [47] John Tukey v roce 1975.

Definice 4. *Bud' P rozdělení na \mathbb{R}^p . Definujme poloprostorovou hloubku⁸ bodu x jako*

$$H(P, x) = \inf\{P(D) : D \text{ je uzavřený poloprostor, } x \in D\}. \quad (4)$$

Zřejmě stačí uvažovat poloprostory, pro něž je x na hranici.

Tato definice je docela názorná. Jde o poměrně jednoduché zobecnění kvantilu. V jednorozměrném případě máme pouhé dva směry; hodnoty větší než x a hodnoty menší než x . V mnohorozměrném případě je směrů pochopitelně nespočetně.

V dvourozměrném prostoru lze poloprostorovou hloubku názorně definovat také pomocí *projekcí* na přímku. Každá přímka p procházející bodem x je tímto bodem rozdělena na dvě polopřímky p_1 a p_2 . Označme P_p rozdělení vzniklé projekcí pravděpodobnostní míry P na přímku p . Hloubka bodu x je v tomto případě

$$H(P, x) = \inf\{\min\{P_p(p_1), P_p(p_2)\} : p \text{ je přímka, } x \in p\}^9.$$

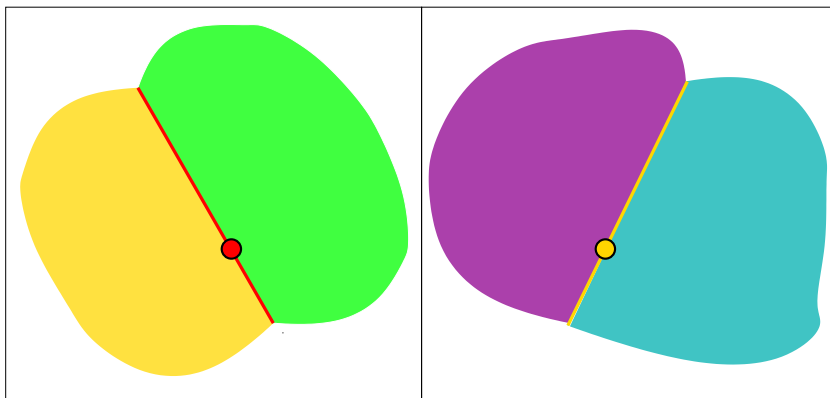
Je zřejmé, že v této definici nijak nezohledňujeme vzdálenosti ostatních bodů od x a svým způsobem ani jejich polohu vůči x . Tím se někdy zbytečně ochuzujeme o důležitou informaci. Najdeme-li totiž poloprostor D takový,

⁶Symetrií lze uvažovat celou řadu; může jít o klasickou středovou symetrii, osovou symetrii, středově-angulární symetrii, eliptickou symetrii, ale třeba i o symetrii vůči otáčení atd.

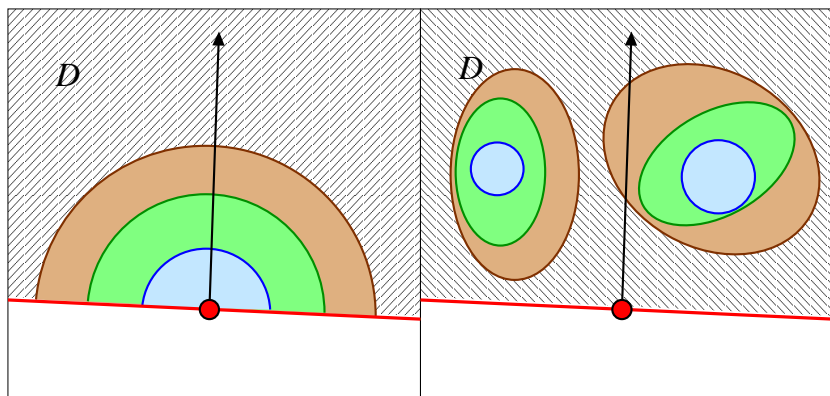
⁷značně subjektivní

⁸Anglicky *halfspace depth*

⁹Čtenář si může sám domyslet, co by nastalo, kdyby místo *všech* přímek obsahujících x uvažoval pro každý bod jen dvě přímky rovnoběžné s osami.



Obrázek 4: Hledání poloprostorové hloubky.



Obrázek 5: Poloprostorová hloubka využívá jen částečně tvar nosiče rozdělení.

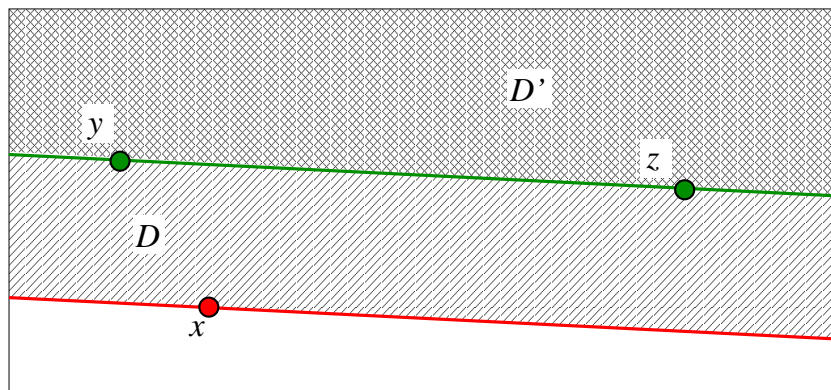
že $x \in D$ a $P(D)$ je minimální, nevíme, jak vypadá rozdělení na D . Je rozdíl, je-li rozdělení na D symetrické kolem normálního vektoru, nebo velmi nesymetrické (viz obrázek 5).

Na druhou stranu, pro náhodný výběr (empirickou míru) platí, že znalost *všech* kvantilových kontur charakterizuje tuto empirickou míru [46].

Zaměříme se na některé základní vlastnosti poloprostorové hloubky. Afinity invariance vyplývá z faktu, že poloprostor po afinní transformaci zůstane poloprostorem. Vzájemná poloha bodů se tak, z *pohledu definice poloprostorové hloubky*, nemění.

Nechť absolutně spojitě rozdělení P je středově symetrické kolem středu s ¹⁰, pak s je nejhlubším bodem. Jakýkoliv uzavřený podprostor na jehož hranici leží s musí mít pravděpodobnost $1/2$, takže hloubka $H(P, s) = 1/2$;

¹⁰Platí tedy $P(s + A) = P(s - A)$ pro každý náhodný jev A .



Obrázek 6: Poloprostorová hloubka klesá směrem od nejhlubšího bodu a je kvazikonkávní.

větší hloubku nemůže mít žádný bod pro libovolné spojitě rozdělení. Pro spojitá rozdělení se souvislým nosičem lze dokázat i jednoznačnost nejhlubšího bodu. V diskrétním případě (speciálně pro empirické míry) se může stát, že existuje konvexní množina nejhlubších bodů.

Pro středově symetrická rozdělení platí symetrie hloubky kolem středu s , neboli

$$H(P, (s + x)) = H(P, (s - x)) \text{ pro všechna } x \in \mathbb{R}^p.$$

Monotonie a kvazikonkavita poloprostorové hloubky jsou také přímým důsledkem používání poloprostorů. Na obrázku 6 je znázorněn bod x a D , poloprostor s nejmenší pravděpodobností, na jehož hranici x leží. Zřejmě

$$\max\{H(P, y), H(P, z)\} \leq P(D') \leq P(D) = H(P, x) \leq 1/2$$

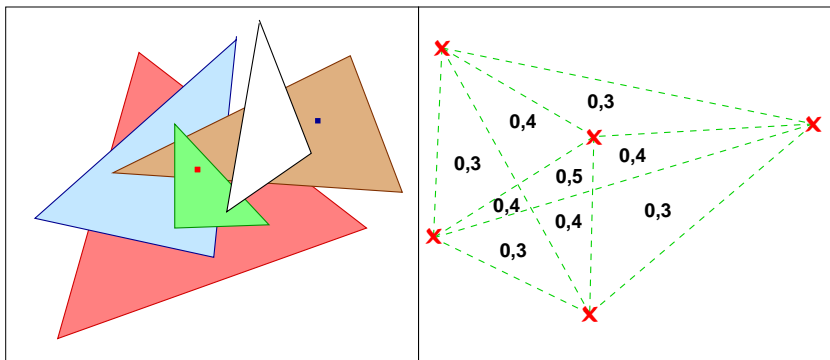
a pro spojitá rozdělení se souvislým nosičem dokonce platí ostré nerovnosti

$$H(P, x) > 0 \Rightarrow \max\{H(P, y), H(P, z)\} < H(P, x).$$

Tím jsme dokázali pokles hloubky podél polopřímek s počátkem v nejhlubším bodě. Podobné úvahy pomohou i k důkazu konvexity centrálních oblastí.

Poloprostorová hloubka splňuje další z požadovaných vlastností, konzistenci. Označíme-li $H(P_n, x)$ hloubku bodu x vzhledem k empirické míře P_n , platí $H(P_n, x) \rightarrow H(P, x)$ skoro jistě. Čtenáře odkazujeme na klasický článek [11], díky kterému se poloprostorová hloubka stala velice populární. V uvedené práci je diskutována také robustnost poloprostorové hloubky a další zajímavé pokročilé vlastnosti.

Velkou nevýhodou poloprostorové hloubky zůstává její výpočetní náročnost. I počítání hloubky byt' jen jediného bodu je s rostoucí dimenzí a s rostoucím počtem pozorování časově velmi náročné. Nejčastěji používané algoritmy vyvinul Rousseeuw se spolupracovníky [42, 43].



Obrázek 7: Náhodné simplexu v \mathbb{R}^2 a simplexová hloubka.

Na druhou stranu zajímavé rozšíření poloprostorové hloubky přináší *regresní hloubka*. Místo hloubky bodů je definována (poloprostorová) hloubka regresní přímky a tím i *nejhlubší regrese* spolu s centrálními oblastmi regresních přímek. Jde o zajímavý alternativní robustní přístup k regresním úlohám. Detailní rozbor regresní hloubky lze nalézt v článku [44].

3.2.2 Simplexová hloubka Tato hloubka se objevuje v článkách Reginy Y. Liu(ové) [29, 30]. Jde o hloubku, kterou můžeme zařadit do třídy hloubek založených na pravděpodobnosti pokrytí bodu *náhodnou oblastí určitého tvaru*. V případě simplexové hloubky je tím předepsaným tvarem *simplex*, čili konvexní obal $p + 1$ bodu $x_1, \dots, x_{p+1} \in \mathbb{R}^p$.

Definice 5. *Bud' P rozdělení na \mathbb{R}^p . Náhodným simplexem rozumíme útvar vzniklý konvexním obalením bodů $V_i, i = 1, \dots, p + 1$, kde V_i jsou nezávislé náhodné vektory s rozdělením P . Definujme simplexovou hloubku¹¹ bodu x jako*

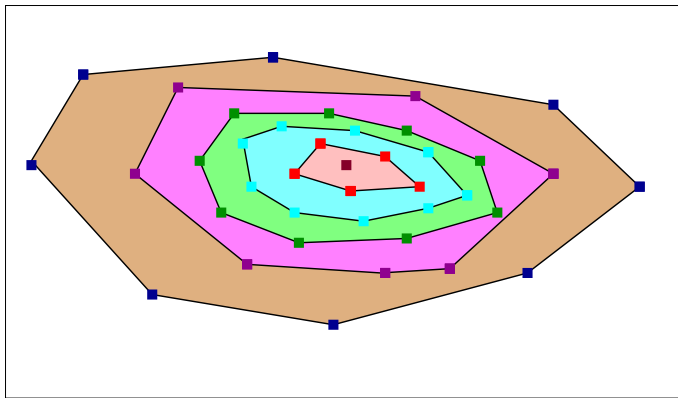
$$H(P, x) = P[x \in S], \text{ kde } S \text{ je náhodný simplex.} \quad (5)$$

Místo simplexu můžeme brát i jiné útvary definované nezávislým výběrem k bodů z rozdělení P . Tak můžeme uvažovat náhodné koule, elipsoidy, mnohoúhelníky.

Náhodné simplexu mají před ostatními tvary zásadní výhodu. Po afinní transformaci zůstává simplex simplexem, což například pro koule a elipsoidy neplatí. Proto hloubka založená na simplexech afinně invariantní *je*, zatímco ostatní obdobné hloubky afinně invariantní *být nemusí* (a typicky nejsou).

Ačkoliv jsou hloubky založené na pravděpodobnosti pokrytí náhodným útvarem, podobně jako poloprostorová hloubka, velmi názorné, jejich vlastností se tak snadno nedokazují. Asymptotické chování simplexové hloubky

¹¹Anglicky *simplicial depth*



Obrázek 8: Loupání slupek.

je shrnuto v článku [13]. Simplexová hloubka je kvazikonkávní, konzistentní, afinně invariantní, spojitá a má spoustu dalších vlastností.

Na druhou stranu je simplexová hloubka výpočetně velmi náročná. V teoretických rozdělení je téměř nemožné ji určit. Pro empirickou míru jde o výpočetně náročný úkol. Možná i proto je simplexová hloubka poněkud ve stínu slavnější a oblíbenější poloprostorové hloubky. Přesto její vývoj neustává, jak lze zjistit například v článku [4].

3.2.3 Konvexní slupky Přísně vzato nejde o hloubku, protože neexistuje definice pro spojitá rozdělení, ale pouze pro výběry ze spojitých rozdělení, tedy pro diskrétní rovnoměrné rozdělení. Jak již název napovídá, *do hloubi dat* se vypravíme skrze odstraňování konvexních obalů výběru¹².

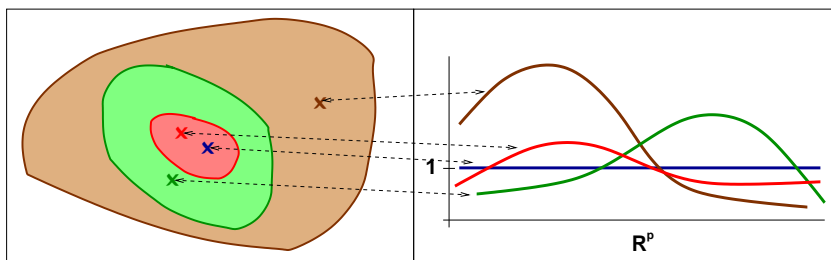
Definice 6. *Bud' P_n empirická míra příslušná náhodnému výběru X_1, \dots, X_n o rozsahu n z rozdělení P . Konvexní obal hloubky $1/n$, označme jej C_1 , je konvexním uzávěrem extrémálních bodů množiny $\mathbf{X} = \{X_1, \dots, X_n\}$. Konvexní obal hloubky k/n , $2 \leq k \leq n$, označme jej C_k , je konvexním uzávěrem extrémálních bodů množiny $\text{int}C_{k-1}$, kde $\text{int}C$ označuje (relativní) vnitřek množiny C . Definujme*

$$H(P_n, x) = k/n \iff x \in C_k \setminus C_{k+1}. \quad (6)$$

Přechod na obecná diskrétní rozdělení, nebo na teoretické spojitá rozdělení není u konvexních slupek jasný. To je ovšem podstatná nevýhoda. Slabou útečou může být, že se výpočetně jedná o výrazně jednodušší verzi hloubky, než jsou simplexová hloubka a poloprostorová hloubka.

Z podstaty definice plyne, že centrální oblasti vymezené konvexními slupkami jsou konvexní. Konvexní slupky jsou také afinně ekvivalentní. O kon-

¹²Odtud se vzal anglický název *convex hull peeling*.



Obrázek 9: Zonoidy a jejich definující funkce.

zistenci se, pochopitelně, hovořit nedá. Podobně problematické je mluvit o symetrii, protože náhodný výběr ze symetrického rozdělení tvoří symetrickou empirickou míru s pravděpodobností nula¹³.

3.2.4 Zonoidy Jde o příklad hloubky, která místo mediánu zobecňuje střední hodnotu. Zonoidy definovali Koshevoy a Mosler v článku [24].

Definice 7. *Bud' $\alpha \in [0, 1]$. Množina*

$$Z_\alpha(P) = \{x \in \mathbb{R}^p : \exists g : \mathbb{R}^p \rightarrow [0, 1/\alpha], E_P g(X) = 1 \text{ a } E_P X g(X) = x\} \quad (7)$$

se nazývá α zonoidem.

Zonoidová hloubka bodu x je definována

$$H(P, x) = \inf\{\alpha : x \in Z_\alpha(P)\} \quad (8)$$

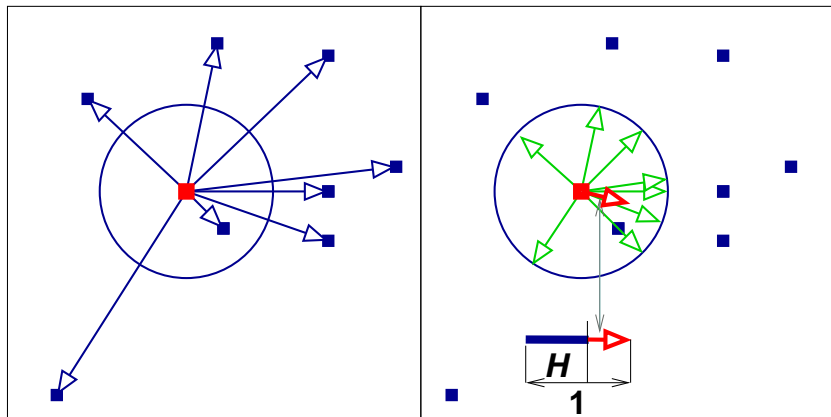
Poprvé se setkáváme s nerobustní definicí hloubky. V definici zonoidu předpokládáme existenci prvního momentu EX . Existuje-li konečná střední hodnota, je vždy nejhlubším bodem. Neexistuje-li konečná střední hodnota, nelze obecně zonoidy definovat.

Pro $\alpha = 1$ je zřejmě jedinou volbou $g \equiv 1$ s.j. a $Z_1 = EX$. Je-li $\alpha_1 < \alpha_2$, pak zjevně (viz obrázek 9) platí $Z_{\alpha_1} \supset Z_{\alpha_2}$. Snadno lze nahlédnout, že zonoidy jsou konvexní. Stačí použít konvexní kombinaci $ag_1(x) + (1-a)g_2(x)$ nějakých dvou funkcí $g_1, g_2 : \mathbb{R}^p \rightarrow [0, 1/\alpha]$. Zonoidová hloubka je tedy kvazikonkávní a navíc, jak lze ověřit přímo z definice, afinně invariantní.

Empirická zonoidová hloubka se dá poměrně rychle spočítat pro konkrétní bod. Je ale mnohem složitější určit vlastní zonoid pro dané α . Algoritmus pro dimenzi $p = 2$ lze najít v [14], jeho výpočetní náročnost pro rozsah výběru n je $O(n^2 \log n)$. Pro vyšší dimenze jde prakticky o nemožný úkol, stejně jako pro teoretická rozdělení.

Další vlastnosti a aplikace zonoidů lze najít v článku [38] či v knize [37].

¹³Je samozřejmě pravda, že symetrická empirická míra vede na symetrickou hloubku—taková situace ale asi v historii lidstva nebyla pozorována

Obrázek 10: Určování směrů a výpočet L_1 hloubky.

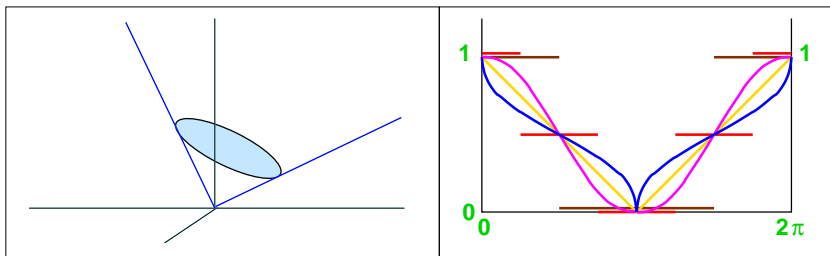
3.2.5 L_1 hloubka Výpočetní náročnost spojená s předchozími definicemi hloubky a s tím spojené obtíže s určováním centrálních oblastí vedly k definicím nových hloubek. Ty často nesplňují úplně všechny klasické požadavky na hloubku (viz část 3.1), ale zato jsou výpočetně mnohem snazší. Jednou z nich je L_1 hloubka uvedená v článku [49]. Podobně jako v případě polo-prostorové hloubky nezávisí L_1 hloubka bodu x na vzdálenostech ostatních bodů. Na rozdíl od polo-prostorové hloubky ale závisí na jejich vzájemném postavení.

Na obrázku 10 vidíme postup pro určení empirické L_1 hloubky. Formálně L_1 hloubku definujeme pomocí délky *průměrného směrového vektoru* k ostatním bodům.

Definice 8. Pro zvolený bod $x \in \mathbb{R}^p$ určíme střední hodnotu $E(X - x) / \|X - x\|$. L_1 hloubkou bodu x nazveme hodnotu

$$H(P, x) = 1 - \|s\|, \text{ kde } s = E \frac{X - x}{\|X - x\|} = \int_{\mathbb{R}^p} \frac{u - x}{\|u - x\|} P(du). \quad (9)$$

Výpočet hloubky je velmi snadný. Není nutné pro každý bod procházet *všechny* možnosti, stačí spočítat střední hodnotu jednoduché transformace náhodného vektoru. Asi nejviditelnější ztrátou za tuto jednoduchost je ztráta afinní invariance hloubky. Po afinní transformaci se může změnit vzájemné postavení bodů; v tomto případě jde o změnu úhlů mezi dvojicemi vektorů. Další, možná ještě horší, vlastností L_1 hloubky je fakt, že ani vně konvexního uzávěru nosiče rozdělení není hloubka nulová. Nulová hloubka znamená, že s pravděpodobností jedna *všechny* hodnoty náhodného vektoru leží *v jediném směru* od uvažovaného bodu. To ale obvykle není možné, takže nulovou hloubku obvykle nemá žádný bod a platí pouze limitní vztah $H(x) \rightarrow 0$,



Obrázek 11: Zobecnění kvantilu na mnohorozměrný M-kvantil. Vlevo ztrátová funkce, vpravo ukázky volby směrových funkcí pro \mathbb{R}^2 .

jestliže $\|x\| \rightarrow \infty$. Velmi nevhodná je tato vlastnost pro náhodné vektory, jejichž nosič není celý prostor.

Pro empirické míry mohou centrální oblasti L_1 hloubky být i nekonvexní. Rozdíl mezi centrální oblastí a jejím konvexním obalem je obvykle zanedbatelný. Proto autoři této metody doporučují použít místo centrálních oblastí jejich *konvexní obaly*¹⁴.

3.2.6 M-kvantily V souvislosti s mnohorozměrnými kvantily se nabízí zobecnění známé definice jednorozměrného kvantilu. Zavedme funkci

$$\varrho(x) = (1 - \alpha)l(x < 0)|x| + \alpha l(x > 0)|x|, x \in \mathbb{R}.$$

Hodnota q_α minimalizující výraz $E\varrho(X - q_\alpha)$ je α kvantilem rozdělení náhodné veličiny X ¹⁵. Připomeňme, že q_α není definováno jednoznačně, ale může jít jak o jediný bod, tak o interval.

Funkci ϱ zobecníme do více rozměrů. Cílem je získat jakýsi „nakloněný polokružel“, obdobu funkce ϱ . Místo indikátorů $l(x < 0)$ a $l(x > 0)$ budeme uvažovat *směrové funkce* $\phi_s(x)$, $s \in \mathbb{R}^p$, $\|s\| = 1$. Funkce $\phi_s(x)$ je funkcí úhlu mezi x a s .

Také vzdálenost l v \mathbb{R}^p můžeme volit různě, nejčastěji se vyskytují \mathcal{L}_1 , \mathcal{L}_2 a \mathcal{L}_∞ vzdálenosti.

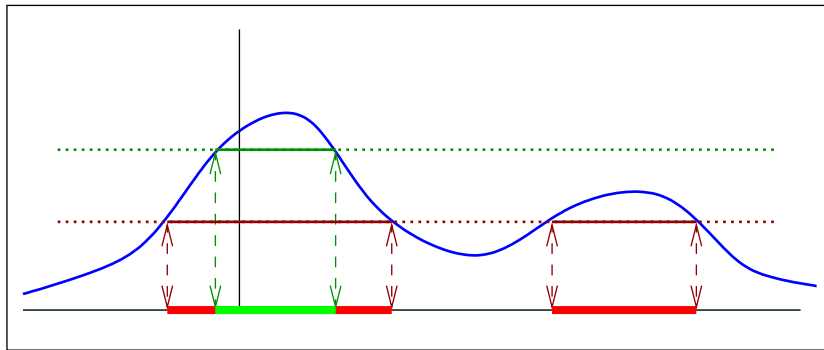
Definujme α M-kvantil pro $\alpha \in [0, 1/2]$ *ve směru* s jako libovolnou hodnotu $q_{\alpha,s}$ minimalizující

$$q_{\alpha,s}(P) = \arg \min_{q \in \mathbb{R}^p} E_P[(1 - 2\alpha)\phi_s(X - q) + \alpha]l(X - q).$$

Volbou $\alpha = 1/2$ získáme mnohorozměrný medián \tilde{x} , se zmenšující se hodnotou α se pro pevné s α M-kvantil vzdaluje od mediánu. Připomeňme, že takto definovaný medián (a potažmo M-kvantily) není afinně ekvivariantní.

¹⁴Jak uvidíme dále, s touto nucenou konvexitou se nedokážeme úplně ztotožnit. Stejně jako s *žádoucí* vlastností hloubky VIII—v tom ale nejsme ojedinelí.

¹⁵Existuje-li první moment náhodné veličiny X



Obrázek 12: Úroňové množiny.

Na M-kvantily je navázána hloubka

$$H(P, x) = \sup\{\alpha : x \notin \{q_{\beta, s}(P); \|s\| = 1, \beta \geq \alpha\}, \alpha \in [0, 1/2]\}. \quad (10)$$

M-kvantily se pravděpodobně poprvé objevují v práci [2]. Poté jsou rozvíjeny v článcích [7, 3] a také u nás v příspěvku [19] a v diplomové práci [17]. Velmi důkladný rozbor teoretických vlastností M-kvantilů a obecnějších M-parametrů mnohorozměrných rozdělení lze najít v článku [23].

Připomeňme, že M-kvantil je funkcí směru s a necentrality α . Praktické pokusy ukazují, že empirické M-kvantily $q_{\alpha, s}$ pro pevné s a pro $\alpha \in [0, 1/2]$ neleží na přímce od středu. V mnoha případech existují dva různé směry s a t a hodnota α takové, že empirické kvantily \tilde{x} , $q_{\alpha, s}$ a $q_{\alpha, t}$ leží na přímce, tedy kvantilová kontura *může být zacyklená*.

3.2.7 Množiny minimálních objemů Poslední z mnoha definic hloubek, které si ukážeme je založena na *úroňových množinách*. Připomeňme, že pro libovolnou funkci $f : \mathbb{R}^p \rightarrow \mathbb{R}$ definujeme úroňové množiny¹⁶ funkce f

$$L_a = L_a(f) = \{x \in \mathbb{R}^p : f(x) \geq a\}, a \in \mathbb{R}.$$

Hlavní výhoda úroňových množin je zřejmá. Buď X p -rozměrný náhodný vektor s absolutně spojitým rozdělením P s hustotou $\psi(x)$. Uvažujme-li úroňovou množinu $L_a(\psi)$ hustoty ψ , pak platí

$$\lambda_p(L_a) = \min\{\lambda_p(B); P(B) = P(L_a)\},$$

kde λ_p je p -rozměrná Lebesgueova míra. Proto hovoříme o *množinách minimálního objemu*¹⁷ v nějaké třídě množin, v tomto případě mezi množinami s danou pravděpodobností.

Hloubka bodu může být definována jednoduše jako hodnota hustoty v daném bodě. Je však vhodnější definovat hloubku tak, aby byla v rozmezí $[0, 1]$.

¹⁶Anglický termín je *level-sets*

¹⁷Minimum volume sets

Definice 9. *Bud' P absolutně spojitě rozdělení pravděpodobností s hustotou ψ . Označme $L_a, a \geq 0$ úrovně množiny hustoty ψ . Definujme hloubku*

$$H(P, x) = 1 - \min\{P(L_a); x \in L_a\} = 1 - P(L_{\psi(x)}).$$

Má-li rozdělení P jeden modus (případně více stejně velkých modů, které dohromady mají pravděpodobnost 0), pak modus je nejhlubším bodem a jeho hloubka je 1. Body mimo nosič hustoty mají zřejmě hloubku 0. Jak je zřejmé i ze schematického obrázku 12, nemusí být centrální množina minimálního objemu ani souvislá, natož konvexní.

Empirická verze úrovně množin je většinou založena na vhodném odhadu hustoty. Od vlastností těchto odhadů se odvíjejí i vlastnosti empirické hloubky, zejména možná afinní invariance a asymptotické vlastnosti. Tomuto problému byla věnována řada prací, například [1, 40, 35].

3.3 Dvě tváře hloubky: třídění podle definice a podle vlastností

Hloubková funkce může být založena na několika různých nápadech, přičemž jsou vždy zdůrazněny jiné globální a jiné lokální vlastnosti rozdělení. Mezi jednotlivými přístupy k hloubce nelze zcela libovolně přecházet, jedna není speciálním případem druhé, těžko by se hledalo zobecnění zastřešující různé přístupy k hloubce. Dá se tedy říci, že při definici hloubky můžeme sledovat pravděpodobnost nějakých útvarů s referenčním bodem x (poloprostorová hloubka), pravděpodobnost, že x je pokryto nějakou množinou (simplexová hloubka), polohu x vůči ostatním bodům (L_1 hloubka, konvexní slupky), řešit optimalizační úlohu (M-kvantily), vázat hloubku přímo na hustotu v daném bodě (úrovně množiny), nebo si vymyslet spoustu dalších konceptů. V každé této třídě lze volit různé varianty základní definice. Místo náhodných simplexů může jít o náhodné elipsoidy, místo poloprostorů o jiné útvary (kužele, pásy, ...), zohlednit vzdálenost, zavést váhy.

Kromě roztřídění hloubek na základě jejich definice lze také posuzovat jejich jednotlivé vlastnosti. Pro stejnou třídu hloubek se chování hloubkové funkce může velmi lišit podle zvolené varianty definice. Typickým příkladem bývá afinní invariance, méně často kvazikonkavita. Také konzistence empirické hloubky se odvíjí od faktu, zda použitá definice splňuje předpoklady některého ze stejnoměrných zákonů velkých čísel (pro asymptotickou normalitu pak nějaké funkcionální centrální limitní věty), které mohou být nalezeny například v [12, 48].

3.4 Jaké vlastnosti má mít ideální hloubka?

Výše jsme uvedli Serflingův seznam vlastností, které by měla splňovat každá hloubka, hloubková funkce.

Začneme bodem I, afinní invariancí. Vzhledem k tomu, že hloubka poukazuje na centrální oblasti a kvantilové kontury, což jsou parametry polohy mnohorozměrného rozdělení, je požadavek afinní invariance přirozený. Dodejme

ještě, že v jednorozměrném případě jsou kvantily ekvivariantní vůči ryze monotónním transformacím (potažmo na kvantilech založená hloubka (1) je invariantní). Obdoba takové vlastnosti není ve více rozměrech obecně možná¹⁸.

Afinní invariance až na výjimky platí pro otočení, překlopení, posunutí. Problém nastává, jestliže zvolená transformace mění různě měřítko jednotlivých os. Tím se totiž mění úhly mezi vektory a hloubka jejíž definice nějakým způsobem úhly mezi vektory zahrnuje, nemůže být vůči této transformaci invariantní. Typickým příkladem je L_1 hloubka, ale i M-kvantily.

Asi největším zklamáním je skutečnost, že přirozený kandidát na mnohorozměrný medián, hodnota minimalizující střední \mathcal{L}_1 odchylku $E\|X - a\|_1$ není afinně ekvivariantní¹⁹. Proto existují různé postupy, jak definovat afinně ekvivariantní medián založený na minimální střední \mathcal{L}_1 odchylce. Častým postupem je standardizace dat, výpočet mediánu a následná zpětná transformace, viz například [5].

Pokud tedy hloubková funkce není afinně invariantní, doporučuje se začít výpočet hloubky vhodnou standardizací dat²⁰ a určením kvantilových kontur pro standardizovaná data. Poté pomocí zpětné transformace určíme hloubku a kvantilové kontury pro původní rozdělení. Možných řešení je spousta, většina je založena na takové transformaci, aby standardizovaná data měla shodný parametr měřítka pro všechna marginální rozdělení a aby byla nekorelovaná. Této problematice se věnují například články [5, 6, 18].

Vlastnosti II a III hovoří o *nějaké* symetrii rozdělení kolem bodu $s \in \mathbb{R}^p$. Bod, kolem kterého je rozdělení symetrické má být nejhlubší. To je určitě rozumný požadavek. Někdy ale jsme před těžkou volbou, zda bod, který není bodem symetrie, ale přesto k němu má v nějakém smyslu blízko, má či nemá být nejhlubším bodem. Tato situace je ilustrována na obrázku 13. Obrázky znázorňují nosiče absolutně spojitého rozdělení, přičemž předpokládáme, že hustota klesá stejně ve všech směrech od vyznačeného modu. V levém sloupci vidíme dvě situace, kde rozdělení není středově symetrické, ale přesto bychom oba vyznačené body mohli považovat za nejhlubší. V pravém sloupci obrázku 13 jsou dvě středově symetrická rozdělení pro která vyznačené body bez váhání označíme za nejhlubší.

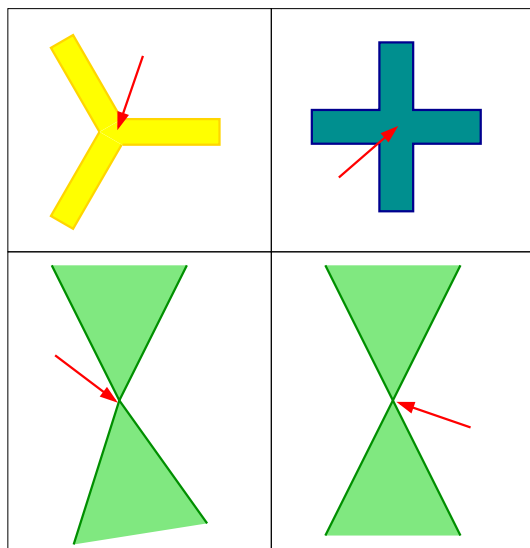
Vlastnost IV, klesání podél paprsků, probereme spolu s konvexitou, se kterou částečně souvisí. Zanedbatelnost hloubky v nekonečnu, vlastnost V, je pro rozdělení náhodného vektoru také přirozeným požadavkem.

Z našeho hlediska je nejvíce diskutabilní vlastnost VIII, kvazikonkavita hloubky. Jedná se o velmi oblíbenou vlastnost, která je jako pozitivní vyzdvihována u nejčastěji používaných hloubek. Je ale zřejmé, že pro úroňové množiny je konvexita centrálních oblastí dosažitelná jen ve speciálních případech. Takže *diktát konvexity* není tak úplný, jak se zdá.

¹⁸Opět se potýkáme s neexistencí lineárního uspořádání a tedy s nemožností definovat rozumně ryzí monotónii zobrazení z \mathbb{R}^p do \mathbb{R} .

¹⁹Afinně ekvivariantní není ani, na první pohled též přirozený, medián po složkách, tedy vektor marginálních mediánů.

²⁰Což bývá vhodné afinní vzájemně jednoznačné zobrazení



Obrázek 13: Nejhlubší bod a střed symetrie.

V práci [10] autoři navrhují uvažovat unimodální hvězdicovité rozdělení a nahradit konvexní konfidenční množiny množinami hvězdicovitými. Připomeňme, že absolutně spojitě unimodální rozdělení s modem v bodě 0 je hvězdicovité kolem nuly, jestliže pro jeho hustotu f platí

$$0 \leq s \leq t < \infty \Rightarrow f(tx) \leq f(sx)$$

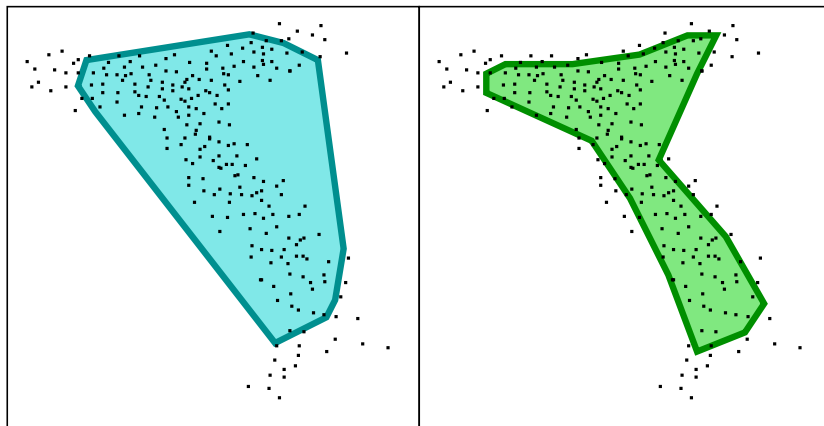
pro všechny body $x \in \mathbb{R}^p$. Množina M je hvězdicovitá vůči bodu x , centru hvězdicovité množiny M , jestliže platí

$$y \in M \Rightarrow \overline{xy} \subset M,$$

kde \overline{xy} značí úsečku s koncovými body x a y . Autoři [10] navrhují najít *nejvhodnější* tvar konfidenční množiny C mezi všemi hvězdicovitými množinami se společným centrem (modem rozdělení) a poté konstruovat centrální oblasti *zvětšováním a zmenšováním* optimální množiny C . Za zmínku stojí, že všechny takto zkonstruované konfidenční množiny mají stejný tvar.

Samozřejmě nechceme zpochybňovat oprávněnost konvexních centrálních oblastí. Jsou případy, kdy se jedná o nezpochybnitelnou volbu. Například pro mnohorozměrné normální rozdělení bychom čekali, že všechny definice hloubky povedou ke známým (a stejným) elipsoidům. Domníváme se ale, že by kvazikonkavita hloubky, tedy konvexita centrálních oblastí, měla být požadována *jen v odůvodněných případech*²¹. Na obrázku 14 je znázorněno,

²¹Jinak jsme sváděni k domněnce, že kvazikonkavita hloubky je pouze z *nouze ctnost*, protože vychází u všech hloubek z *hlavního proudu*



Obrázek 14: Konvexní a hvězdovitá množina.

jak nuceně konvexní tvary centrálních oblastí mohou obsahovat zbytečně mnoho *hluchých* míst. V určitých případech nezbývá než uznat, že mnohorozměrné prostory nabízejí bohatší geometrii než jednorozměrný. Proto se přidáváme k názoru, že je vhodné uvažovat bohatší množinu tvarů, než jsou konvexní množiny.

Opustíme-li tedy požadavek konvexity centrálních oblastí, musíme se zamyslet zda tento požadavek nějak nahradíme. DasGupta a spol. navrhuje používat hvězdovité množiny, tedy zachovat alespoň vlastnost IV, pokud hloubky podél paprsků (navrhují ale použít jeden tvar množiny pro všechny centrální oblasti daného rozdělení, což je hodně omezující požadavek vzniklý snad z výpočetně-optimalizačních důvodů). Na druhé straně stojí úrovnové množiny, které mohou mít zcela libovolný tvar, nemusí být ani souvislé, nemusí být dokonce ani otevřené, ani uzavřené (nesplňují tedy apriori požadavky IV, VI a VIII). Z analogie k jednorozměrným kvantilům se přirozeným požadavkem ukazuje *souvislost* centrálních oblastí, alespoň v případech unimodálních rozdělení²². V každém případě pro obecná rozdělení se nezdá být vhodné omezovat možné tvary centrálních oblastí. Spíše je podstatné vybrat hloubkovou funkci, která z rozdělení *zachová a zvýrazní* ty vlastnosti, které zkoumáme²³.

4 Směrové hloubky

Chceme-li rozšířit možné tvary centrálních oblastí na hvězdovité množiny, musíme se zamyslet nad definicí hloubky, která tuto vlastnost zaručí. Výše jsme uvedli, že v článku [10] se hledá optimální tvar hvězdovité množiny

²²Tato vlastnost je pro unimodální rozdělení úrovnovými množinami pochopitelně splněna.

²³Volba metody je dost subjektivní—to by nás nemělo ve statistice překvapovat.

pro dané rozdělení. Existuje však definice hloubky, která přímo vede na hvězdicovité množiny a navíc α centrální oblasti mohou mít pro různá α velmi odlišné tvary. Jedná se o směrovou hloubku.

4.1 Směrová hloubka na přímkách

Ve svém článku [52] definuje Wei směrové kvantily pomocí rozdělení na přímkách procházejících *středem*.

Definice 10. *Uvažujme absolutně spojitě rozdělení pravděpodobností P na \mathbb{R}^p a jeho hustotu f . Buď S mnohorozměrný medián rozdělení P . Pro libovolnou přímku l tvaru $l = S + tx$, $t \in \mathbb{R}$, kde $x = (x_1, x_2, \dots, x_p)$ je vektor délky 1 a $x_p \geq 0$ (l prochází bodem S , x směřuje na „sever“) definujme podmíněnou hustotu $f_l(t)$ na \mathbb{R} jako hustotu podmíněného rozdělení reálného parametru t za podmínky, že $X = S + tx$ (X leží na přímce l). Pro (jednorozměrné) rozdělení s hustotou f_l najdeme $\alpha/2$ kvantil $q_{\alpha/2}(l)$ a $1-\alpha/2$ kvantil $q_{1-\alpha/2}(l)$. Body $S + q_{\alpha/2}(l)x$ a $S + q_{1-\alpha/2}(l)x$ tvoří α směrové kvantily pro přímku l . Množina všech směrových α kvantilů tvoří α kvantilovou konturu.²⁴*

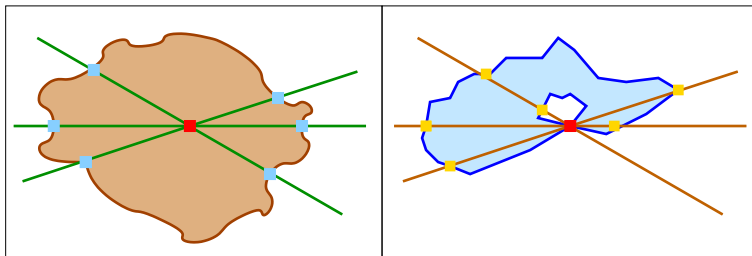
Všimněme si, že směrové kvantily samy o sobě *nedefinují střed*. Ten musí být dodán zvenčí.

Samotná definice směrových kvantilů, tedy pomocí podmíněných rozdělení na přímkách, je afinně invariantní. Přímka zůstane přímkou a pořadí bodů na přímce se buď zachová, nebo převrátí. Pro afinní invarianci je tedy rozhodující afinní invariance navrženého středu. Wei navrhuje použít mnohorozměrný medián, který obecně afinně invariantní není, proto doporučuje nejprve provést standardizace diskutované výše.

Směrové kvantily na přímce mají jednu zásadní nevýhodu. Pro jednotlivé přímky je medián $q_{1/2}(\cdot)$ totožný se středem S jen velmi výjimečně (v případě středově symetrických rozdělení). Může se proto snadno stát, že oba kvantily, $q_{\alpha/2}(l)$ a $q_{1-\alpha/2}(l)$ leží vůči středu S na stejné straně přímky l . Tím dochází k nevhodnému *zacyklení* kvantilových kontur, což je jen velmi obtížně vysvětlitelný fenomén. Zejména postavení středu S je v tomto případě značně podivné.

Další otázkou je, jak počítat směrové kvantily pro empirická rozdělení. V tomto případě skoro všechny přímky neobsahují žádné pozorování. Pro určení směrových kvantilů proto musíme použít nějakou lokální metodu. Nabízí se nahradit kvantily na přímce nějakým odhadem kvantilu na okolí zkoumané přímky. Jinou možností je nejprve odhadnout rozdělení, například pomocí jádrového odhadu hustoty, a poté spočítat směrové kvantily pro odhad hustoty. V každém případě na volbě metody závisí, zda empirické směrové kvantily budou afinně invariantní, případně konzistentní.

²⁴Autor doufá, že zde přesně vystihl význam trochu vágněji napsané definice, kterou lze najít v původním článku.



Obrázek 15: Nezacyklená a zacyklená kvantilová kontura.

Pokud nedochází k zacyklení kvantilových kontur, jsou směrové centrální množiny přirozeně hvězdicovité. V opačném případě pochopitelně nikoliv. V článku [52] autor navrhuje *zacyklenou oblast přidat* do centrální množiny, čímž se odstraní (!) uvedený defekt a centrální množina je hvězdicovitá. Střed S v každém případě zůstává na hranici všech zacyklených kvantilových kontur.

4.2 Polární směrové kvantily—cykly u nás nevedeme

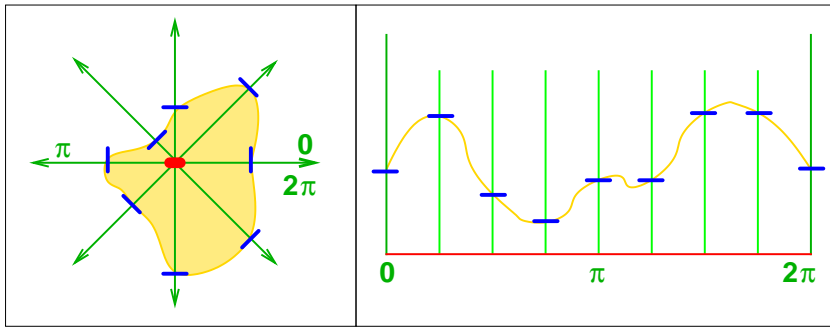
V diplomové práci [17] se objevuje jiná definice směrového kvantilu. Tentokrát se nebudeme dívat na přímky procházející středem S , ale na polopřímky, kterým je S počátkem.

Definice 11. *Uvažujme absolutně spojitě rozdělení pravděpodobností P na \mathbb{R}^p s hustotou f . Bud' S mnohorozměrný medián rozdělení P . Pro libovolnou polopřímku m tvaru $m = S + tx$, $t \geq 0$, kde $\|x\| = 1$ definujme podmíněné rozdělení $f_m(t)$ jako podmíněné rozdělení P za podmínky, že X leží na polopřímce m . Pro (jedorozměrné) rozdělení s hustotou f_m najdeme $1 - \alpha$ kvantil $q_{1-\alpha}(m)$. Body $S + q_{1-\alpha}(m)x$ tvoří $1 - \alpha$ směrové kvantily pro polopřímku m . Množina všech směrových $1 - \alpha$ kvantilů tvoří $1 - \alpha$ kvantilovou konturu.*

Ekvivalentní a názornější je definice *polopřímkového* kvantilu pomocí *polárních souřadnic* vůči středu S .

Uvažujme absolutně spojitě rozdělení pravděpodobností P na \mathbb{R}^p s hustotou f . Bud' S mnohorozměrný medián rozdělení P . Převedme náhodnou veličinu X s rozdělením P na náhodnou veličinu $Z = (D, A)$ vyjadřující hypersférické souřadnice X vůči S . Konkrétně $D \geq 0$ je *vzdálenost* X od S a A určuje *směr*²⁵ X od S . Převod do sférických souřadnic v trojrozměrném Euklidově prostoru lze přehledně nalézt například v [41].

²⁵V hypersférických souřadnicích se jedná o vektor úhlů $\phi_1 \in [0, 2\pi)$, $\phi_i \in [0, \pi)$, $i = 2, \dots, p - 1$.



Obrázek 16: Polopřímkový směrový kvantil a jeho výpočet pomocí polárních (hypersférických) souřadnic.

Označme Q rozdělení náhodné veličiny Z , Q je opět absolutně spojitě, a označme g hustotu Q vůči Lebesgueově míře. Označme dále g_a hustotu podmíněného rozdělení Q_a vzdálenosti D při podmínce $A = a$.

Definice 12. Polárními kvantily rozdělení P vůči středu S rozumíme kvantily podmíněných rozdělení Q_a .

Polární α kvantil ve směru a , označený $q_{1-\alpha}(a)$, je definovaný jako $1 - \alpha$ kvantil jednorozměrné náhodné veličiny D za podmínky $A = a$.

Po převedení hypersférických souřadnic $(q_{1-\alpha}(A), A)$ zpět do kartézských dostaneme stejné kvantilové kontury jako v definici 11. Z praktických důvodů dáváme přednost definici 12, protože názorně ukazuje cestu k výpočtu, viz obrázek 16.

Stejně jako v případě přímkového směrového kvantilu, i polární směrový kvantil je afinně ekvivariantní, je-li afinně ekvivariantní volba středu S . Díky polárním souřadnicím je možné poměrně snadno počítat podmíněné kvantily $q_{1-\alpha}(a)$ ve směru a i pro empirická rozdělení. Tato definice přímo vybízí k použití regresních kvantilů, které v tomto případě musí být ve správném smyslu periodické. Více o tomto přístupu lze nalézt v diplomové práci [26], případně v tomto svazku v příspěvku [27].

Centrální oblasti polárních kvantilů jsou z definice přirozeně hvězdicovité vůči středu S .

4.3 Porovnání obou přístupů

V obou případech se směrovými kvantily spojujeme přirozeně definovanou směrovou hloubku. Pro přímkové směrové kvantily ji lze pro bod x ležící na přímce $l = S + tz$ definovat jako

$$H(P, x) = \sup\{\alpha : x \in [S + q_{\alpha/2}(l)z, S + q_{1-\alpha/2}(l)z]\}. \quad (11)$$

V případě polárních směrových kvantilů je pro bod x ležící ve směru a od středu S hloubkou hodnota

$$H(P, x) = \sup\{\alpha : \|x - S\| \leq q_{1-\alpha}(a)\}. \quad (12)$$

Jinými slovy, bod x má hloubku $H(P, x)$, jestliže na přímce (polopřímce) spojující x a S leží uvnitř každé $1 - h$ kvantilové kontury pro $h < H(P, x)$.

Velkou výhodou směrových kvantilů je fakt, že lze přímo určovat $1 - \alpha$ centrální oblasti. To vyplývá z faktu, že z podmíněných rozdělání ve směrech (na přímkách, nebo v polárních souřadnicích) určuje obyčejné jednorozměrné kvantily, proto oblast vymezená kvantily pro daný směr má pravděpodobnost $1 - \alpha$ a díky větě o *úplné pravděpodobnosti* má příslušná centrální oblast pravděpodobnost $1 - \alpha$. Pro empirickou verzi toto platí samozřejmě *jen* přibližně. Tato vlastnost je velmi výjimečná, u *všech ostatních uvedených* hloubek je nutné postupovat opačně. Pro úrovně množiny hloubkové funkce nejprve určíme pravděpodobnost, za centrální oblast vybereme tu úrovně množinu, která má předepsanou pravděpodobnost.

V obou případech se jedná o výpočetně poměrně jednoduché postupy. Domníváme se, že náš postup pomocí polárních souřadnic je dokonce názornější a přímočarý. Jediné, nač musíme upozornit u polárních kvantilů je nutnost *periodicity* (více v [26, 27]).

Směrové kvantily jsou teoreticky afinně ekvivariantní. V obou případech ale tato vlastnost závisí na volbě středu S (viz výše). Afinní invariance hloubky neplatí pro empirické verze, neboť je nutné nahradit přímkou (polopřímku) oblastmi s nenulovou Lebesgueovou mírou. Zvolíme-li takovou oblast pevně, nemůže být afinně ekvivariantní. Proto lze doporučit vhodnou standardizaci náhodného výběru předtím, než přistoupíme k vlastnímu výpočtu hloubky.

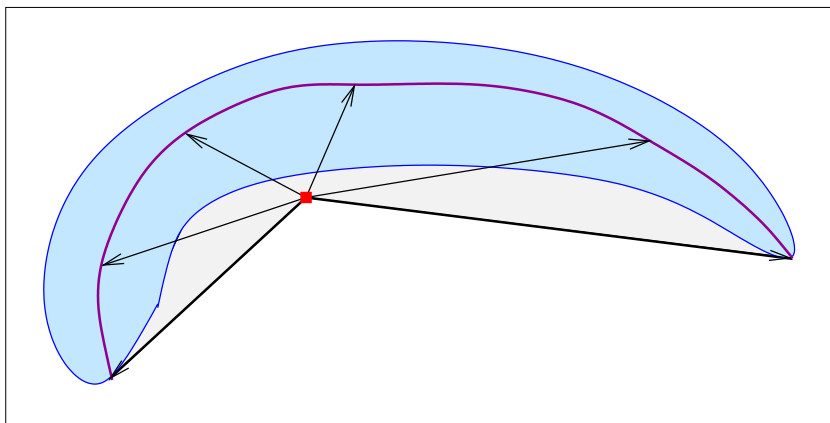
Na polárních směrových kvantilech je dobře vidět, že konzistence hloubky je spjata s konzistencí jednotlivých odhadů. Konzistentní musí být odhad středu, odhad regresních kvantilů a , v případě standardizace, odhad standardizační transformace. V článku [52] je konzistence uvedena jako *zřejmá*.

4.4 Volíme střed

Jak zvolit ten nejvhodnější střed pro směrové kvantily? Wei [52] navrhuje použít afinně ekvivariantní verzi mnohorozměrného mediánu (minimalizujícího střední \mathcal{L}_1 vzdálenost). V tom případě ale potřebujeme předpokládat existenci konečného prvního momentu. Navíc pro nekonvexní nosič rozdělání nemusí mnohorozměrný medián být *uvnitř* nosiče. Ideální kandidát na střed směrových kvantilů by ale *měl být obklopen* oblastí s nenulovou pravděpodobností ve všech směrech. Z podobných důvodů nemusí být vhodným kandidátem ani nejhlubší bod v poloprostorové hloubce, případně v L_1 hloubce.

Zatímco při použití přímkových směrových kvantilů střed mimo nosič rozdělání vede vždy k zacyklení (všech) kvantilových kontur²⁶, pro polární

²⁶K zacyklení může dojít i v případě, kdy střed *je* uvnitř nosiče



Obrázek 17: Polopřímkový směrový kvantil; střed je na hranici všech centrálních oblastí.

směrové kvantily k zacyklení nemůže dojít ani v tomto případě. Přesto přechod k polárním kvantilům není všespásný. Stále zůstává faktem, že polární směrové kvantily jsou vhodné pro *hvězdčovitá* rozdělení. Nemůže sice nastat zacyklení kvantilové kontury, ale střed může stále být na hranici *všech* centrálních oblastí. Na obrázku 17 je proto znázorněno rozdělení *na banánu*, klasický příklad odhalující slabiny mnoha hloubek.

Světle šedá oblast neobsahuje žádné hodnoty náhodné veličiny, přesto je zahrnuta do všech centrálních oblastí. Střed je na hranici všech centrálních oblastí (je obsažen ve všech kvantilových konturách). Problém je zjevně ve volbě středu²⁷. Přitom množina na obrázku 17 hvězdčovitá je vůči mnoha bodům (což sice *neznamená*, že i rozdělení s tímto nosičem je hvězdčovité, ale předpokládejme, že ano), takže by vhodným středem mohl být některý z těchto bodů.

V mnoha situacích by se nejlépe hodil modus rozdělení. Modus se ale nesehná hledá a, což je podstatnější, nemusí být jediný²⁸.

Ve standardních případech symetrických rozdělení je vesměs jedno, kterou hloubku zvolíme pro hledání středu—nejhlubšího bodu. Jsou ale situace, kdy nemůžeme říci, který střed je ten pravý. Snad by stálo za úvahu použít pro hledání vhodného *středu* podobný postup, jaký je v článku [10] navržen pro hledání vhodného *tvaru* hvězdčovité množiny. Vybrat střed (definuje příslušnou hloubku a kvantilové kontury) podle nějakého vhodného kritéria pro (vybrané) centrální oblasti.

²⁷Ale co dělat, když například medián po složkách, nebo poloprostorově nejhlubší bod takto vyjde?

²⁸Modus ovšem může také ležet na hranici nosiče, jako je tomu například pro exponenciální rozdělení

Nezbytnost dosadit zvenčí střed S do definice směrové hloubky může vést ke zvýšení výpočetní náročnosti. Na druhou stranu, ani při použití poloprostorové hloubky není nutné počítat hloubku *všech bodů* k určení nejhlubšího bodu. Mezi rychlé algoritmy výpočtu empirického nejhlubšího bodu patří například postup popsany v článku Jiřího Matouška [36].

Poznámka 2. *Směrová hloubka se hodí zejména na rozdělení hvězdovitěho typu. Oproti kvazikonkávním hloubkám dostaneme širší třídu možných tvarů centrálních množin.*

Přesto v některých případech můžeme potřebovat ještě flexibilnější třídu možných centrálních oblastí. Například pro směsi rozdělení. Nechceme-li přitom používat úrovněvé množiny, musíme nějakým způsobem lokalizovat některou z klasických definic hloubek.

5 Zobecnění poloprostorové hloubky—vážená hloubka

Původně tato hloubka vznikla při hledání vhodného středu pro směrovou hloubku. Ukázalo se ale, že má některé *nečekané*, ale *zajímavé* vlastnosti, viz [20, 50, 51]. Zejména oceňujeme možnost volit mezi zvýrazněním či potlačení lokálních a globálních vlastností zkoumaného rozdělení.

5.1 Definice

Nejprve definujme váhovou funkci.

Definice 13. *Funkce $w_+ : \mathbb{R}^p \rightarrow [0, \infty)$ taková, že $w_+(x) = 0$ pro $x \in \{(x_1, \dots, x_p); x_p < 0\}$ se nazývá váhová funkce hloubky. Označme dále $w_-(x) = w_+(-x)$.*

Z mnoha důvodů se vyplatí používat jen *rozumné* váhové funkce. Například takové, že jejich hodnota závisí na *vzdálenosti* podél osy x_p a na vzdálenosti $x_1^2 + \dots + x_{p-1}^2$ od osy x_p . Tedy

$$w_+(x_1, \dots, x_{p-1}, x_p) = w(x_1^2 + \dots + x_{p-1}^2, x_p), \quad (13)$$

což je požadavek určité *silné symetrie* funkce w_+ okolo osy x_p . Slabším požadavkem je *symetrie* váhové funkce hloubky

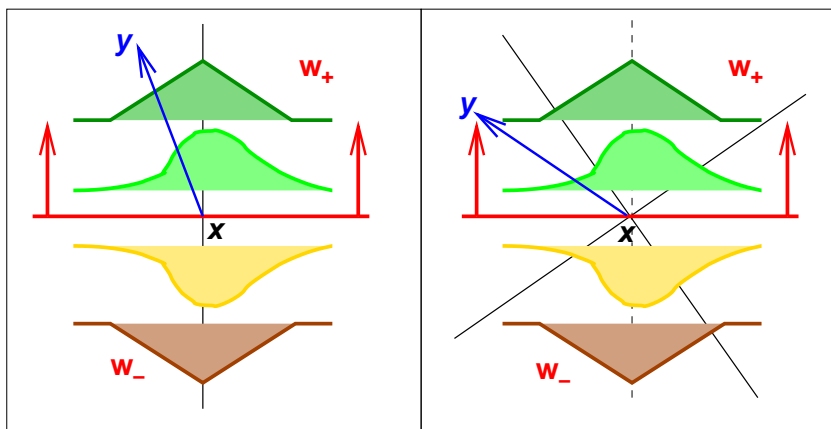
$$w_+(x_1, x_2, \dots, x_{p-1}, x_p) = w_+(-x_1, -x_2, \dots, -x_{p-1}, x_p). \quad (14)$$

Někdy je vhodné předpokládat také *směrovou homogenitu* váhové funkce hloubky

$$w_+(x_1, \dots, x_{p-1}, x_p) = w_+(x_1, \dots, x_{p-1}, kx_p), \text{ pro všechna } k > 0, \quad (15)$$

případně *úhlovou homogenitu* ve tvaru

$$w_+(x_1, \dots, x_{p-1}, x_p) = w_+(kx_1, \dots, kx_{p-1}, kx_p), \text{ pro všechna } k > 0. \quad (16)$$



Obrázek 18: Výpočet vážené hloubky bodu x minimalizací přes všechny ortonormální transformace rozdělení.

Nebudeme-li váhovou funkci specifikovat, budeme dále požadovat její symetrii, po částech spojitost a $w_+(0, \dots, 0, x_p) \geq m > 0$ pro $x_p \geq 0$. Příkladem vhodné váhové funkce může být například *pásová váhová funkce*

$$w_+(x) = \begin{cases} 1 & \sum_1^{p-1} x_i^2 \leq h, x_p \geq 0, \\ 0 & \text{jinak.} \end{cases} \quad (17)$$

Pásová váhová funkce je zřejmě silně symetrická a směrově homogenní ve smyslu (13) a (15).

Vektor $(0, 0, \dots, 0, 1)$ má úlohu *směru*, kterým je zobecněná poloprostorová hloubka orientována. Označme \mathcal{O}_p množinu všech *ortonormálních transformací* (zachovávajících vzdálenosti a úhly) prostoru \mathbb{R}^p .

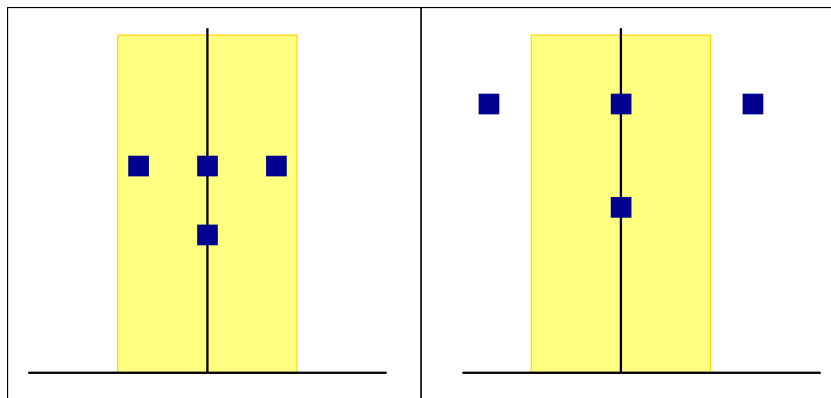
Definice 14. *Bud' P rozdělení pravděpodobnosti na prostoru \mathbb{R}^p a w_+ vhodná váhová funkce hloubky. Definujme hloubku bodu x jako*

$$D(P, x) = \inf_{A \in \mathcal{O}_p} \frac{E_P w_+(A(X - x))}{E_P w_-(A(X - x))}, \quad (18)$$

kde podíl $0/0$ definujeme jako 1.

Na obrázku 18 je znázorněn postup počítání vážené hloubky. V bodě x je počítána hloubka s využitím váhových funkcí w_+ a w_- . Provedeme všechny ortonormální transformace které nezmění bod x (na obrázku je znázorněno otočení bodu y a celého rozdělení) a pro každou takovou ortonormální transformaci A spočítáme podíl vah $E_P w_+(A(X - x)) / E_P w_-(A(X - x))$.

Takto definovaná hloubka je invariantní vůči posunutí i rotacím. Obecně však afinně invariantní není, protože definice váhové množiny v sobě obvyk-



Obrázek 19: Zavedení vah do poloprostorové hloubky kazí afinní invarianci.

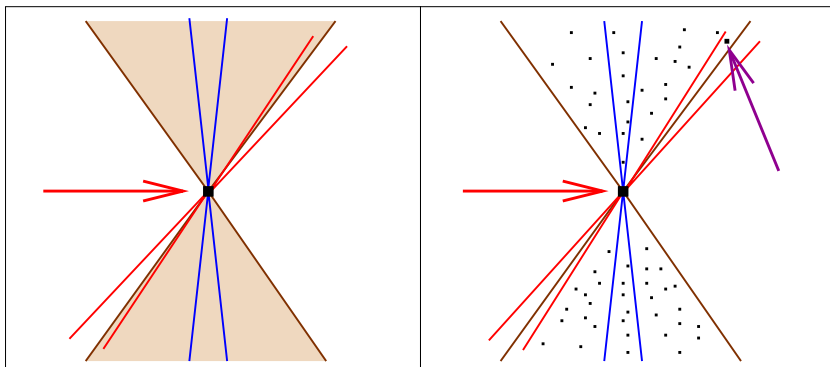
le zahrnuje (byť skrytě) úhly mezi vektory. Například pásová hloubková funkce (17) nemůže definovat hloubku invariantní vůči různému přenásobení os, viz obrázek 19. Proto doporučujeme použít opět standardizaci pomocí vhodné transformace a teprve poté volit vhodnou hloubkovou funkci a počítat hloubku.

5.2 Základní vlastnosti a volba hloubkové funkce

Na rozdíl od většiny hloubek nemusí existovat právě jeden bod s největší váženou hloubkou. V tom je vážená hloubka podobná úrovniovým množinám. Centrální oblasti nemusí být ani konvexní, ani hvězdicovité, ba ani souvislé množiny. Na druhou stranu, při volbě vhodné hloubky (symetrické, případně ještě homogenní) lze snadno ukázat, že střed symetrických rozdělání (případně angulárně symetrických rozdělání) je jediným nejhlubším bodem. V případě středově symetrických rozdělání je i vážená hloubka středově symetrická za předpokladu, že váhová funkce hloubky je symetrická.

Pro absolutně spojitá rozdělání a *rozumné* váhové funkce je hloubka bodu konzistentní. Rozumné váhové funkce jsou takové, které jsou dostatečně spojité (po částech spojitě) a správně měřitelné. příkladem může být funkce (17), případně nějaká její spojitá verze. Toto platí až na některé patologické případy bodů na hranici nosiče hustoty při nevhodně zvolené váhové funkci. Ukázka nekonzistence je na obrázku 20²⁹. Představme si rovnoměrné rozdělání na znázorněné množině. Přírozený střed ve vrcholu kužele teoreticky má, díky symetrii rozdělání, hloubku jedna. Ale pro libovolný náhodný výběr z tohoto rozdělání existuje pozorování y takové, že při vhodném otočení nezasáhne váhová funkce w_+ žádné pozorování, zatímco váhová funkce w_- zasáhne *právě* pozorování y . Empirická hloubka vrcholu kuželu je $0/n^{-1} = 0$ s pravděpodobností jedna, nemůže tak konvergovat k jedné.

²⁹Jde o dost podivný, až patologický příklad



Obrázek 20: Vážená hloubka nemusí být konzistentní ve všech bodech. Teoreticky (vlevo) má vrchol kuželu hloubku jedna. Pro libovolně velký výběr má vrchol s pravděpodobností jedna hloubku nula (vpravo).

Jde zřejmě o velmi umělý příklad. Hloubka všech ostatních bodů v tomto případě konzistentní je. Zde se sešly dvě neblahé skutečnosti; střed (vrchol kuželu) není obklopen nosičem ze všech stran a váhové funkce w_+ s w_- nepokrývají dohromady žádné okolí bodu 0, jen výseč tohoto okolí. Buď se tedy smíříme s tím, že hloubka nemusí být konzistentní na hranici nosiče, nebo zavedeme požadavek *lokálního pokrytí* váhové funkce. Tento požadavek lze vyjádřit formálně

$$\exists r > 0, \varepsilon > 0 \text{ tak, že } \|x\| \leq r, x_p \geq 0 \Rightarrow w_+(x) \geq \varepsilon > 0. \quad (19)$$

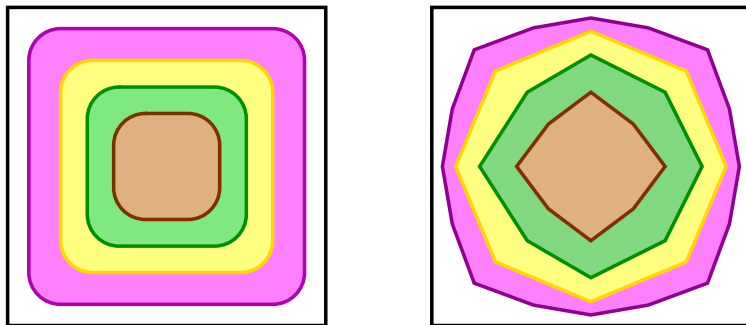
Celkově lze říci, že vážená hloubka oproti poloprostorové hloubce zdvíhá lokální vlastnosti rozdělení. Z toho plyne i návod, jak volit váhovou funkci. Volbou

$$w_+(x) = \begin{cases} 1 & x_p \geq 0, \\ 0 & x_p < 0 \end{cases}$$

dostáváme ekvivalent poloprostorové hloubky. Čím více zužujeme nosič hloubkové funkce k ose x_p (viz například vážená hloubka (17)), tím více zdvíhujeme lokální vlastnosti rozdělení. Na druhou stranu, empirická vážená hloubka počítá jen s pozorováním, které padne do nosiče váhové funkce. Nespojitá váhová funkce má podobné dopady jako nespojitě jádro v odhadu hustoty. Proto lze, zejména pro menší rozsahy výběru, doporučit váhové funkce spojité v $\mathbb{R}^{p-1} \times \mathbb{R}^+$.

5.3 Příklady rozdílů mezi poloprostorovou a váženou hloubkou

Zaměříme se na situace, kdy vážená hloubka dává podstatně jiné výsledky než poloprostorová hloubka. Použijme váhovou funkci (17). Připomeňme, že čím



Obrázek 21: Centrální oblasti vážené (vlevo) a poloprostorové (vpravo) hloubky pro rovnoměrné rozdělení na čtverci.

větší hodnotu h v (17) volíme, tím více se vážená hloubka blíží poloprostorové hloubce a naopak.

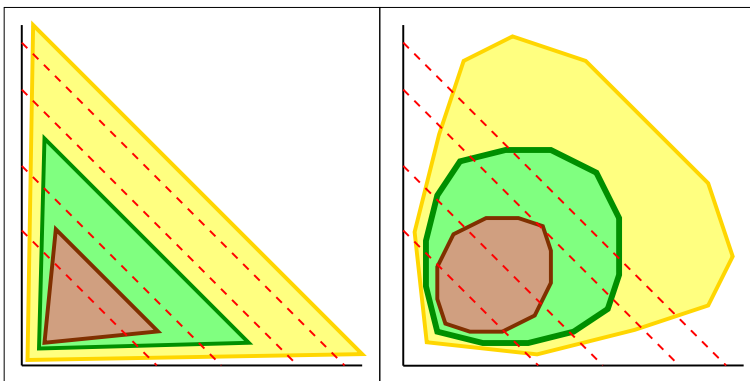
Na obrázku 21 vidíme schematicky znázorněné centrální oblasti pro rovnoměrné rozdělení na čtverci $[0, a]^2$. Volbou váhové funkce (17), kde h je relativně malé vůči délce strany a (například $h = a/20$) dostáváme obrázek vlevo. Na první pohled centrální oblasti vážené hloubky mnohem více kopírují tvar nosiče hustoty. Zaoblení v rozích jsou způsobena volbou šířky pásu h . Standardní poloprostorová hloubka (vpravo) vede na centrální oblasti, které jsou, protože jde o rovnoměrné rozdělení, stejně *velké* v Lebesgueově míře jako odpovídající centrální oblasti vážené hloubky. Tvar centrálních oblastí je hodně proměnlivý a jen pomalu se přibližuje nosiči hustoty. Připomeňme, že úrovnové množiny jsou v tomto případě jen tři, jmenovitě \emptyset , $[0, a]^2$ a \mathbb{R}^2 .

Rovnoměrné rozdělení není možná úplně typické. Na obrázku 22 máme schematicky naznačeny centrální oblasti pro dvourozměrné rozdělení s exponenciálními marginálami a nezávislými složkami. Pro porovnání jsou doplněny vybrané vrstevnice hustoty (hranice úrovnových množin).

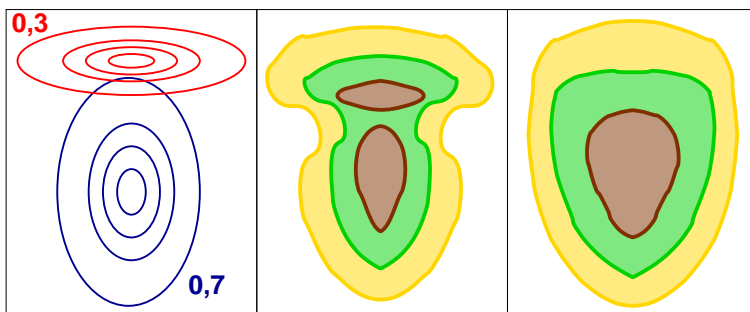
Vážená hloubka opět nabízí centrální oblasti, které jsou mnohem *blíže* struktuře hustoty rozdělení. Tyto centrální oblasti se příliš neliší od úrovnových množin naznačených v obrázku. Centrální oblasti pro poloprostorovou hloubku jsou *příliš vypouklé* a *zbytečně objemné* (v Lebesgueově míře), podle našeho názoru nejsou moc vhodné³⁰. Na tomto příkladu je pěkně vidět, proč o vážené hloubce tvrdíme, že je *kompromisem mezi úrovnovými množinami a poloprostorovou hloubkou*.

Poslední ilustrací budiž případ směsi dvou normálních rozdělení. Na obrázku 23 máme schematicky nakreslené vrstevnice dvou normálních rozdělení, které kombinujeme s vahami 0,3 a 0,7. Uprostřed a vlevo jsou naznačeny vrstevnice vážené, případně poloprostorové hloubky a jejich centrální oblasti.

³⁰Nikomu ale náš názor nevnučujeme. Volba hloubky a posuzování jejích vlastností je velice subjektivní.



Obrázek 22: Centrální oblasti vážené (vlevo) a poloprostorové hloubky pro dvourozměrné exponenciální rozdělení. Čárkovaně jsou naznačeny úrovně množiny.



Obrázek 23: Centrální oblasti vážené (uprostřed) a poloprostorové (vpravo) hloubky pro směs dvou normálních rozdělení. Vlevo jsou naznačeny vrstevnice obou složek a jejich váha ve směsi.

Opět vidíme již známý fenomén. Vážená hloubka více kopíruje hustotu, centrální oblasti poloprostorové hloubky jsou konvexní a mají *zbytečně velkou* plochu. Centrální oblasti vážené hloubky naznačují existenci *dvou oblastí*, podobně jako by tomu bylo u úrovněových množin. Zde se rýsuje možnost použít váženou hloubku pro diskriminační a klasifikační účely.

Přímo se nabízí otázka, zda množina všech kvantilových kontur, nebo ekvivalentně množina všech centrálních oblastí, může charakterizovat rozdělení. Pro poloprostorovou hloubku platí, viz [46], že množina všech *empirických* kvantilových kontur charakterizuje náhodný výběr, potažmo empirickou míru. Podobný výsledek je znám i pro některé další hloubky. Speciálně pro zonoidy [24] a pro simplexovou hloubku [25]. Pro váženou hloubku ještě obdobný výsledek nemáme.

6 Používání hloubky

Hloubka se uplatňuje například v mnohorozměrných obdobích neparametrických testů. Rozsáhlé shrnutí nabízí článek [31], ve kterém se lze dočíst o popisných statistikách, obdobích známých QQ diagramů, nebo o mnohorozměrné variantě krabicových diagramů, slunečních grafech³¹. Jsou definovány pojmy jako měřítko a rozptyl, šikmost, špičatost coby obdoby známých charakteristik jednorozměrné náhodné veličiny. Pomocí grafických metod jsou potom testovány hodnoty těchto charakteristik, případně porovnávány dva výběry. Zde se naskýtá mnoho příležitostí pro odvozování dalších variant a verzí neparametrických mnohorozměrných metod založených na uspořádání hloubkou. Novější výsledky lze nalézt například v článcích [28, 33].

Další z aplikací hloubek je v oblasti klasifikace dat. Velkou výhodou je neparametrická podstata hloubky, čili klasifikační kritéria nejsou založena na (obvyklém) předpokladu normality. Více informací lze získat v článcích [15, 16, 22, 39].

Hloubka je ale zajímavá i pro výzkumníky v oblasti výpočetní geometrie. Zájemcům doporučujeme <http://www.cs.tufts.edu/research/geometry>. Zde lze nalézt také odkaz na software Depth Explorer [21]. Ten pokrývá zejména poloprostorovou hloubku a L_1 hloubku a v současné době by měl už být i ve verzi pro Linux.

Poslední oblastí, kterou zmíníme, je uplatnění hloubky pro funkcionální data, tedy pro nekonečněrozměrné výběrové prostory. V poslední době se objevuje několik prací, zabývajících se touto problematikou, například [8, 9, 34]. I zde je výzkum teprve v počátcích. Proto se domníváme, že v oblasti hloubky a jejích aplikací je spousta zajímavých a nevyřešených problémů, které čekají, až se jich někdo chopí.

Literatura

- [1] Baillo A., Cuesta-Albertas J.A., Cuevas A. (2001). *Convergence rates in nonparametric estimation of level sets*. *Statistics and Probability Letters* **53**, 27–35.
- [2] Breckling J., Chambers R. (1988). *M-quantiles*. *Biometrika* **75**, 761–771.
- [3] Breckling J., Kokic P., Lübke O. (2001). *A note on multivariate M-quantiles*. *Statistics and Probability Letters* **55**, 39–44.
- [4] Burr M.A., Rafalin, E., Souvaine, D.L. (2006). *Simplicial depth: An improved definition, analysis and efficiency for the finite sample case*. In *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications* (Liu, Serfling and Souvaine, Eds.), American Mathematical Society, 1–16.

³¹Sunburst plot

- [5] Chakraborty B., Chaudhuri P. (1996). *On a transformation and retransformation technique for constructing an affine equivariant multivariate median*. Proceedings of the American Mathematical Society **124**, 2539–2547.
- [6] Chakraborty B., Chaudhuri P., Oja H. (1998). *Operating transformation retransformation on spatial median and angle test*. Statistica Sinica **8**, 767–784.
- [7] Chaudhuri P. (1996). *On a geometric notion of quantiles for multivariate data*. Journal of the American Statistical Association **91**, 862–872.
- [8] Cuesta-Albertos J.A., Fraiman, R. (2006) *Impartial trimmed means for functional data*. In Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications (Liu, Serfling and Souvaine, Eds.), American Mathematical Society, 121–146.
- [9] Cuesta-Albertos, J. A. and Nieto-Reyes, A. (2008) *A random functional depth*. In Functional and Operatorial Statistics (Dabo-Niang and Ferraty, Eds.), Physica-Verlag, 121–126.
- [10] DasGupta A., Ghosh J.K., Zen M.M. (1995). *A new general method for constructing confidence sets in arbitrary dimensions: with applications* Annals of Statistics **23**, 1408–1432.
- [11] Donoho D., Gasko M. (1992). *Breakdown properties of location estimates based on halfspace depth and projected outlyingness*. Annals of Statistics **20**, 1803–1827.
- [12] Dudley R.M. (2008). *Uniform Central Limit Theorems*. Cambridge.
- [13] Dümbgen L. (1991). *Limit theorems for the simplicial depth*. Statistics and Probability Letters **14**, 119–128.
- [14] Dyckerhoff R. (2000). *Computing zonoid trimmed regions of bivariate data sets*. In Compstat 2000. Proceedings in Computational Statistics (Bethlehem and van der Heijden, Eds.), Physica-Verlag, 295–300.
- [15] Ghosh A.K., Chaudhuri, P. (2005). *On data depth and distribution free discriminant analysis using separating surfaces*. Bernoulli **11**, 1–27.
- [16] Hartikainen A., Oja H. (2006). *On some parametric, nonparametric and semiparametric discrimination rules*. In Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications (Liu, Serfling and Souvaine, Eds.), American Mathematical Society, 61–70.
- [17] Hasil J. (2005). *Problém kvantilu ve více rozměrech*. Diplomová práce, MFF UK.
- [18] Hettmansperger T.P., Randles R.H. (2002). *A practical affine equivariant multivariate median*. Biometrika **89**, 851–860.
- [19] Hlubinka, D. (2002). *O kvantilech ve více rozměrech*. In ROBUST 2002 (Antoch, Dohnal, Eds.).
- [20] Hlubinka, D., Kotík, L. and Vencálek, O. (2009). *Weighted halfspace depth*. To appear in Kybernetika

- [21] Hugg J., Rafalin E., Seyboth K. Souvaine D. (2006). *An experimental study of old and new depth measures*. In Workshop on Algorithms Engineering and Experiments. Lecture Notes in Computer Sciences, Springer, 51–64.
- [22] Jörnsten, R. (2004). *Clustering and classification based on the L_1 data depth*. Journal of Multivariate Analysis **90**, 67–89.
- [23] Koltchinskii, V. I. (1997). *M-estimation, convexity and quantiles*. Annals of Statistics **25**, 435–477.
- [24] Koshevoy G., Mosler K. (1997). *Zonoid trimming for multivariate distributions*. Annals of Statistics **25**, 1998–2017.
- [25] Koshevoy, G. (1997). *Multivariate depths and underlying distributions: A uniqueness property*. Technical report, Universität zu Köln.
- [26] Kotík, L. (2007). *Periodické regresní kvantily*. Diplomová práce, MFF UK.
- [27] Kotík, L. (2009). *Directional quantiles*. In ROBUST 2008 (Antoch, Dohnal, Eds.)
- [28] Li, J. and Liu, R. Y. (2004). *New nonparametric tests of multivariate locations and scales using data depth*. Statistical Science **19**, 686–696.
- [29] Liu, R. Y. (1988). *On a notion of simplicial depth*. Proceedings of the National Academy of Science USA **85**, 1732–1734.
- [30] Liu, R. Y. (1990). *On a notion of data depth based on random simplices*. Annals of Statistics **18**, 405–414.
- [31] Liu R.Y., Parelius J.M., Singh K. (1999). *Multivariate analysis by data depth: Descriptive statistics, graphics and inference (with discussion)*. Annals of Statistics **27**, 783–858.
- [32] Liu R.Y., Serfling R., Souvaine D.L., Editors (2006). *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications*. American Mathematical Society.
- [33] Liu R.Y., Singh K. (2006). *Rank tests for multivariate scale difference based on data depth*. In Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications (Liu, Serfling and Souvaine, Eds.), American Mathematical Society, 17–36.
- [34] López-Pitardo S., Romo J. (2006). *Depth-based classification for functional data*. In Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications (Liu, Serfling and Souvaine, Eds.), American Mathematical Society, 103–120.
- [35] Mason D., Polonik W. (2008). *Asymptotic normality of plug-in level set estimates*. Preprint.
- [36] Matoušek J. (1991). *Computing the centre of planar point set*. In Discrete and Computational Geometry: Papers from the DIMACS Special Year (Goodman, Pollack and Steiger, Eds.), American Mathematical Society, 221–230.
- [37] Mosler K. (2002). *Multivariate Dispersion, Central Regions and Depth; The Lift Zonoid Approach*. Springer.

- [38] Mosler, K. (2003). *Central regions and dependency*. Methodology and Computing in Applied Probability **5**, 5–21.
- [39] Mosler K., Hoberg R. (2006). *Data analysis and classification with the zonoid depth*. In Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications (Liu, Serfling and Souvaine, Eds.), American Mathematical Society, 49–60.
- [40] Polonik W. (1997). *Minimum volume sets and generalized quantile processes*. Stochastic Processes and their Applications **69**, 1–24.
- [41] Rektorys, K. (2003). *Přehled užití matematiky I (7. vyd.)*. Prometheus.
- [42] Rousseeuw, P. and Ruts, I. (1996). *Algorithm AS307. Bivariate location depth*. Applied Statistics **45**, 516–526.
- [43] Rousseeuw P., Struyf A. (1998). *Computing location depth and regression depth in higher dimensions* Statistics and Computing **8**, 193–203.
- [44] Rousseeuw P., Hubert M. (1999). *Regression depth (with discussion)* Journal of the American Statistical Association **94**, 388–433.
- [45] Serfling R. (2006). *Depth functions in nonparametric multivariate inference*. In Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications (Liu, Serfling and Souvaine, Eds.), American Mathematical Society, 1–16.
- [46] Struyf A., Rousseeuw P. (1999). *Halfspace depth and regression depth characterize the empirical distribution*. Journal of Multivariate Analysis **69**, 135–153.
- [47] Tukey J. (1975). *Mathematics and the picturing of data*. In Proceedings of the International Congress of Mathematicians, Vancouver 1974 (R.D. James, Ed.) **2**, 523–531.
- [48] van den Vaart A., Welner J. (2000). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer.
- [49] Vardi Y., Zhang C.H. (2000). *The multivariate L_1 median and associated data depth*. Proceedings of National Academy of Science USA **97**, 1423–1426.
- [50] Vencálek O. (2008). *Generalization of halfspace depth*. In WDS '08; „Week of Doctoral Students 2008“ Part I (Šafránková and Pavlů, Eds.), Matfyzpress, 100–105.
- [51] Vencálek O. (2009). *Weighted data depth and its properties*. In ROBUST 2008 (Antoch, Dohnal, Eds.)
- [52] Wei Y. (2008). *An Approach to Multivariate Covariate-Dependent Quantile Contours With Application to Bivariate Conditional Growth Charts*. Journal of the American Statistical Association **103**, 397–409.

Poděkování: Článek vznikl s podporou grantu GAČR 201/08/0486. Práce je součástí výzkumného projektu MSM 0021620839 financovaného MŠMT ČR. Autor děkuje Lukáši Kotíkovi a Ondřeji Vencálkovi za pomoc a podporu.

Adresa: D. Hlubinka, MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8 – Karlín

E-mail: daniel.hlubinka@mff.cuni.cz