ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE
Fakulta jaderná a fyzikálně inženýrská

SVOČ 2009

2009 Bc. Jitka Hanousková

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE
Fakulta jaderná a fyzikálně inženýrská
Katedra matematiky

Bc. Jitka Hanousková

Asymptotické vlastnosti odhadů s minimální vzdáleností

Vedoucí práce: Ing. Václav Kůs, PhD.

2009

Poděkování

Na tomto místě bych ráda poděkovala svému školiteli Ing. Václavu Kůsovi, PhD. za poskytnutí literatury, velkou vstřícnost a cenné rady při konzultacích dané problematiky, Bc. Pavlu Hrabákovi za pomoc s grafikou v programu LaTeX a Petru Čapounovi za IT podporu.

Čestné prohlášení

Prohlašuji na tomto místě, že jsem předloženou práci vypracovala samostatně a že jsem uvedla veškerou použitou literaturu.

V Praze dne 29. dubna 2009

. . . .

Jméno Příjmení

Název práce:
Asymptotické vlastnosti odhadů s minimální vzdáleností

Autor: Bc. Jitka Hanousková

Obor: Matematické inženýrství

Vedoucí práce: Ing. Václav Kůs, PhD., Katedra matematiky, Fakulta jaderná a fyzikálně inženýrská, České vysoké učení technické v Praze

Abstrakt: Tato diplomová práce studuje konzistenci a řád konzistence v L_1-normě odhadů pravděpodobnostních hustot s minimální vzdáleností. Zobecníme podmínky kladené na rodinu hustot za kterých je Kolmogorovský odhad konzistentní řádu $n^{-1/2}$ v L_1-normě a také ve střední hodnotě L_1-normy. Jako důsledek odvodíme řád konzistence pro odhady s minimální diskrepanční a Lévyho vzdáleností. Teoretické výsledky ověříme simulací. Na příkladech porovnáme výsledky teorie stupně variace a jejího zobecnění s výsledky Vapnik–Chervonenkisovi teorie a Yatracosovy teorie pro konzistenci odhadů s minimální vzdáleností.

Klíčová slova: řád konzistence, odhady s minimální vzdáleností, stupeň variace.

Obsah

Úvod

Tato práce se zabývá odhady pravděpodobnostních hustot s minimální vzdáleností a řádem jejich konzistence v L_1-normě a její střední hodnotě. Největší část práce se věnuje odhadům s minimální Kolmogorovskou vzdáleností. Dále se zabýváme odhady s minimální zobecněnou Kolmogorov–Smirnovovou, Lévyho Cramer–von Mises a diskrepanční vzdáleností. Kolmogorovské odhady pravděpodobnostních hustot patří mezi neparametrické odhady, z nichž nejznámější jsou histogram nebo jádrový odhad. Pro hustoty s omezeným nosičem je pro histogram dokázána konzistence řádu $n^{-1/3}$ ve střední hodnotě L_1-normy (viz např. [2]). Pro třídu dvakrát diferencovatelných hustot s omezeným nosičem je známa konzistence jádrového odhadu řádu $n^{-2/5}$ v L_1-normě (viz např. [2]). V článku [11] byla dokázána konzistence Kolmogorovských odhadů řádu $n^{-1/2}$ v L_1–normě i její střední hodnotě za předpokladu lokální dominance a dále byla dokázána postačující podmínka pro splnění této dominance. V této práci zobecníme teorii z článku [11], jako důsledky teorie stupně variace a vzájemných vztahů různých vzdáleností mezi pravděpodobnostními mírami odvodíme řád konzistence pro odhady s minimální Lévyho a diskrepanční vzdáleností. Teoretické výsledky ověříme simulací. Následně se budeme věnovat dalším dvěma teoriím poskytujícím podmínky pro konzistenci odhadů s minimální vzdáleností. A to Vapnik-Chervonenkisově teorii a jejím výsledkům pro konzistenci odhadů s minimální zobecněnou Kolmogorov–Smirnovovou vzdáleností a Yatracosově teorii. Výsledky všech přístupů porovnáme na příkladech.

Kapitola 1

Základní pojmy

Zaveďme značení, které budeme v práci používat. Buď $(\mathcal{X}, \mathcal{A})$ měřitelný prostor s množinovou $\sigma-$alaebrou $\mathcal{A}$ a $\mathcal{P}(\mathcal{X}, \mathcal{A})$ množina pravděpodobnostních měr na $(\mathcal{X}, \mathcal{A})$ (dále jen $\mathcal{P}$ pro neparametrický případ, $\{\mathcal{P}_\theta, \theta \in \Theta\}$ pro případ parametrický bude-li to třeba zvýraznit). Nechť λ je $\sigma-$finitní míra na $(\mathcal{X}, \mathcal{A})$. Dále malými písmeny $f, g, \ldots$ budeme značit hustoty pravděpodobnosti a velkými písmeny $F, G, \ldots$ jim odpovídající distribuční funkce, pokud budou existovat.

Buďte $X_1, \ldots, X_n$ stejně a nezávisle rozdělená pozorování na $P \in \mathcal{P}$, symbolem F_n budeme označovat empirickou distribuční funkci založenou na $(X_1, \ldots, X_n)$

$$F_n(x) = \frac{1}{n} \sum_{j=1}^{n} \mathbf{I}_{\{X_j \leq x\}} \tag{1.1}$$

a symbolem $\nu_n(A)$ budeme označovat empirickou distribuci založenou na $(X_1, \ldots, X_n)$

$$\nu_n(A) = \frac{1}{n} \sum_{j=1}^{n} \mathbf{I}_{\{X_j \in A\}}\,, A \in \mathcal{A}\,, \tag{1.2}$$

kde $\mathbf{I}_{\{X_j \leq x\}}$ je indikátor jevu $X_j \in (-\infty, x]$ a $\mathbf{I}_{\{X_j \in A\}}$ je indikátor jevu $X_j \in A$.

Definice 1. Odhad $\widehat{f}_n$ hustoty f, případně odhad $\widehat{\theta}$ parametru θ nazýváme odhadem s minimální vzdáleností d právě tehdy, když platí

$$\widehat{f}_n = \arg\min_{P \in \mathcal{P}} d(P, \nu_n) \qquad \text{s. j.} \tag{1.3}$$

$$\widehat{\theta} = \arg\min_{\theta \in \Theta} d(P_\theta, \nu_n) \qquad \text{s. j.}\,, \tag{1.4}$$

kde d je daná vzdálenost na $\mathcal{P}$ a Θ je příslušný parametrický prostor.

Definice 2. Říkáme, že odhad $\widehat{f}_n$ hustoty $f \in$ je konzistentní v dané ρ_d vzdálenosti, respektive v její střední hodnotě, právě když $\rho_d(\widehat{f}_n, f) \to 0$ skoro jistě, respektive když $E\rho_d(\widehat{f}_n, f) \to 0$. Říkáme, že odhad $\widehat{f}_n$ je konzistentní řádu $r_n \to 0$ ve vzdálenosti ρ_d, respektive v její střední hodnotě právě tehdy, když $\rho_d(\widehat{f}_n, f) = O_p(r_n)$, respektive když $E\rho_d(\widehat{f}_n, f) = O(r_n)$.

A symboly $O(r_n)$ a $O_p(r_n)$ jsou definovány následně.

Definice 3. Buďte funkce $f(x), g(x) : \mathbb{R} \to \mathbb{R}$. Říkáme, že f se chová jako g, značíme $f(x) = O(g(x))$, právě když

$$(\exists K > 0)(\forall\, x \in \mathbb{R}) \left| \frac{f(x)}{g(x)} \right| \leq K\,. \tag{1.5}$$

Definice 4. Buď $(X_n)_1^\infty$ posloupnost náhodných veličin. Řekneme, že je omezená v pravděpodobnosti, označíme $X_n = O_p(1)$ právě tehdy, když pro každé $\varepsilon > 0$ existují konstanty M a n_0 takové, že

$$P(|X_n| > M) < \varepsilon, \qquad \forall\, n > n_0\,. \tag{1.6}$$

Obecněji, pro dvě posloupnosti náhodných veličin $(X_n)_1^\infty$ a $(Y_n)_1^\infty$ symbol $X_n = O_p(Y_n)$ znamená, že $X_n/Y_n = O_p(1)$.

Kapitola 2

Statistické vzdálenosti a divergence

V této kapitole krátce zmíníme některé statistické vzdálenosti a jejich vzájemné vztahy, které se nám budou dále hodit. V této části budeme již specifikovat o jaké měřitelné prostory se jedná. Pokud $\mathcal{X}$ bude metrický prostor budeme jej považovat za měřitelný s Borelovskou $\sigma-$algebrou $\mathcal{B}$.

2.1 Nerovnosti mezi pravděpodobnostními vzdálenostmi

Pro $\mathcal{X}$ omezený metrický prostor označme $\operatorname{diam}(\mathcal{X}) = \sup\{d(x,y) : x, y \in \mathcal{X}\}$, kde d je metrika na $\mathcal{X}$. Pro jednotlivé vzdálenosti používáme následující označení.

označení	vzdálenost
$d_{C–M}$	Cramer–von Mises vzdálenost
d_D	Diskrepance
d_H	Hellingerova vzdálenost
d_I	Relativní entropie
d_K	Kolmogorova metrika
d_L	Lévyho metrika
d_P	Prochorova metrika
d_S	Separační vzdálenost
d_{TV}	Vzdálenost v totální variaci
d_W	Wassersteinova metrika
d_{χ^2}	χ^2-vzdálenost

Tabulka 2.1: Zkratky pro metriky použité ve schématu na Obrázku 2.1

Separační vzdálenost

- definovaná na $\mathcal{X}$ libovolném spočetném prostoru
- definice:

$$d_S(P,Q) := \max_i \left(1 - \frac{P(i)}{Q(i)}\right)$$

- není metrika

Relativní entropie (Kullback-Leiblerova divergence)

- definovaná na $\mathcal{X}$ libovolném měřitelném prostoru
- definice: pokud f, g jsou hustoty odpovídající pravděpodobnostním mírám P, Q s ohledem na dominující míru λ

$$d_I(P,Q) := \int\limits_{\mathcal{X}} f \log\left(\frac{f}{g}\right) \mathrm{d}\lambda\,,$$

  s konvencí $0 \log(\frac{0}{q}) = 0$ pro všechna reálná q a $p \log(\frac{p}{0}) = \infty$ pro všechna reálná nenulová p, kde $\mathrm{supp}(P)$ je nosič míry P na $\mathcal{X}$
- není metrika

χ^2-vzdálenost

- definovaná na $\mathcal{X}$ libovolném měřitelném prostoru
- definice: pokud f, g jsou hustoty odpovídající pravděpodobnostním mírám P, Q s ohledem na dominující míru λ

$$d_{\chi^2}(P,Q) := \int\limits_{\mathcal{X}} \frac{(f-g)^2}{g} \mathrm{d}\lambda\,,$$

  kde $\mathrm{supp}(P)$ a $\mathrm{supp}(Q)$ jsou nosiče měr P a Q na $\mathcal{X}$
- není metrika

Vzdálenost v totální variaci

- definovaná na $\mathcal{X}$ libovolném měřitelném prostoru
- definice:

$$d_{TV}(P,Q) := 2 \sup_{A \in \mathcal{A}} |P(A) - Q(A)|$$

  a pokud f, g jsou hustoty odpovídající pravděpodobnostním mírám P, Q s ohledem na dominující míru λ, pak platí

$$d_{TV}(P,Q) := \int_{\mathcal{X}} |f-g| \,\mathrm{d}\lambda$$

Hellingerova vzdálenost

- definovaná na $\mathcal{X}$ libovolném měřitelném prostoru

- definice: pokud f, g jsou hustoty odpovídající pravděpodobnostním mírám P, Q s ohledem na dominující míru λ

$$d_H(P,Q) := \left[\int_{\mathcal{X}} \left(\sqrt{f} - \sqrt{g}\right)^2 \mathrm{d}\lambda\right]^{1/2} = \left[2\left(1 - \int_{\mathcal{X}} \sqrt{fg}\,\mathrm{d}\lambda\right)\right]^{1/2}$$

Diskrepance

- definovaná na $\mathcal{X}$ libovolném metrickém prostoru
- definice:

$$d_D(P,Q) := \sup_{B \in \boldsymbol{B}} |P(B) - Q(B)|\,,$$

kde $\boldsymbol{B}$ je množina všech uzavřených koulí v $\mathcal{X}$

Prochorovova (nebo Lévy-Prochorovova) metrika

- definovaná na $\mathcal{X}$ libovolném metrickém prostoru
- definice:

$$d_P(P,Q) := \inf\left\{\varepsilon > 0 : P(B) \leq Q(B^\varepsilon) + \varepsilon,\ \text{pro } \forall B \in \mathcal{B}\right\},$$

kde $B^\varepsilon = \{x : \inf\limits_{y \in B} d(x,y) < \varepsilon\}$ a d je metrika daného prostoru $\mathcal{X}$

Wassersteinova (nebo Kantorovichova) metrika

- definovaná na $\mathbb{R}$, nebo $\mathcal{X}$ libovolném metrickém prostoru
- definice: pro $\mathcal{X} = \mathbb{R}$ nechť F, G jsou distribuční funkce příslušející mírám P, Q:

$$d_W(P,Q) := \int\limits_{-\infty}^{\infty} |F(x) - G(x)|\,\mathrm{d}x\,,$$

pro metrický prostor:

$$d_W(P,Q) := \sup\left\{\left|\int_{\mathcal{X}} h\,\mathrm{d}\mu - \int_{\mathcal{X}} h\,\mathrm{d}\nu\right| : h \text{ splňující } |h(x) - h(y)| \leq d(x,y)\right\},$$

kde h je zobrazení $\mathcal{X} \to \mathbb{R}$

Lévyho metrika

- definovaná na $\mathbb{R}$
- definice:

$$d_L(F,G) := \inf\left\{\varepsilon > 0 : G(x-\varepsilon) - \varepsilon \leq F(x) \leq G(x+\varepsilon) + \varepsilon : \forall x \in \mathbb{R}\right\},$$

kde F, G jsou distribuční funkce k P, Q

Pokud P, Q jsou pravděpodobnostní míry na $\mathbb{R}$ je zvykem vyjadřovat následující vzdálenosti jako vzdálenost mezi jejich distribučními funkcemi F, G.

Kolmogorovova metrika

- definovaná na $\mathbb{R}$
- definice:
$$d_K(P,Q) := \sup_{x\in\mathbb{R}} |F(x) - G(x)|\,,$$
kde F, G jsou distribuční funkce k mírám P, Q. Tuto metriku je možno ekvivalentně definovat podobným způsobem jako je definována Lévyho metrika:
$$d_K(P,Q) = \inf\{\varepsilon > 0 : F(x) - \varepsilon \le G(x) \le F(x) + \varepsilon, \forall\, x \in \mathbb{R}\}$$

Cramer–von Mises

- definovaná na $\mathbb{R}$
- definice:
$$d_{C\text{–}M}(F,G) = \int_{\mathbb{R}} \Big(F(x) - G(x)\Big)^2 f(x)\,\mathrm{d}x\,,$$
kde F, G jsou distribuční funkce příslušející k P, Q
- není metrika

Zmiňme nyní omezení platnosti nerovností mezi jednotlivými vzdálenostmi, podrobněji viz článek [6].

Kolmogorova a Lévyho metrika
Pokud je distribuční funkce G absolutně spojitá vzhledem k Lebesqueově míře, pak platí nerovnost:
$$d_K(F,G) \le \left(1 + \sup_x |G'(x)|\right) d_L(F,G).$$

Prochorovova metrika a diskrepance
Buď $\mathcal{X}$ libovolný metrický prostor a Q libovolná pravděpodobnostní míra splňující
$$Q(B^\varepsilon \le Q(B) + \phi(\varepsilon))\,,$$
pro všechny koule B a nějakou zprava spojitou funkci ϕ. Potom pro každou pravděpodobnostní míru P platí, že
$$d_D(P,Q) \le d_P(P,Q) + \phi(d_P(P,Q))\,.$$

Wassersteinova metrika a totální variace
Pokud je $\mathcal{X}$ konečná množina pak platí
$$d_{\min} \cdot d_{TV}(P,Q) \le d_W(P,Q)\,,$$
kde $d_{\min} = \min\{d(x,y) : x, y \in \mathcal{X}, x \ne y\}$ a d je uvažovaná metrika na $\mathcal{X}$.

Relativní entropie a totální variace
Pro spočetný prostor $\mathcal{X}$ platí nerovnast

$$d_{TV}(P,Q) \leq \sqrt{\frac{d_I(P,Q)}{2}}.$$

Hellingerova a χ^2–vzdálenost
Obecně platí

$$d_H(P,Q) \leq \sqrt{2}\sqrt[4]{d_{\chi^2}(P,Q)}\,,$$

pokud míra P dominuje míru Q (ozn. $P \ll Q$) přejde výše zmíněná nerovnost na

$$d_H(P,Q) \leq \sqrt{d_{\chi^2}(P,Q)}\,.$$

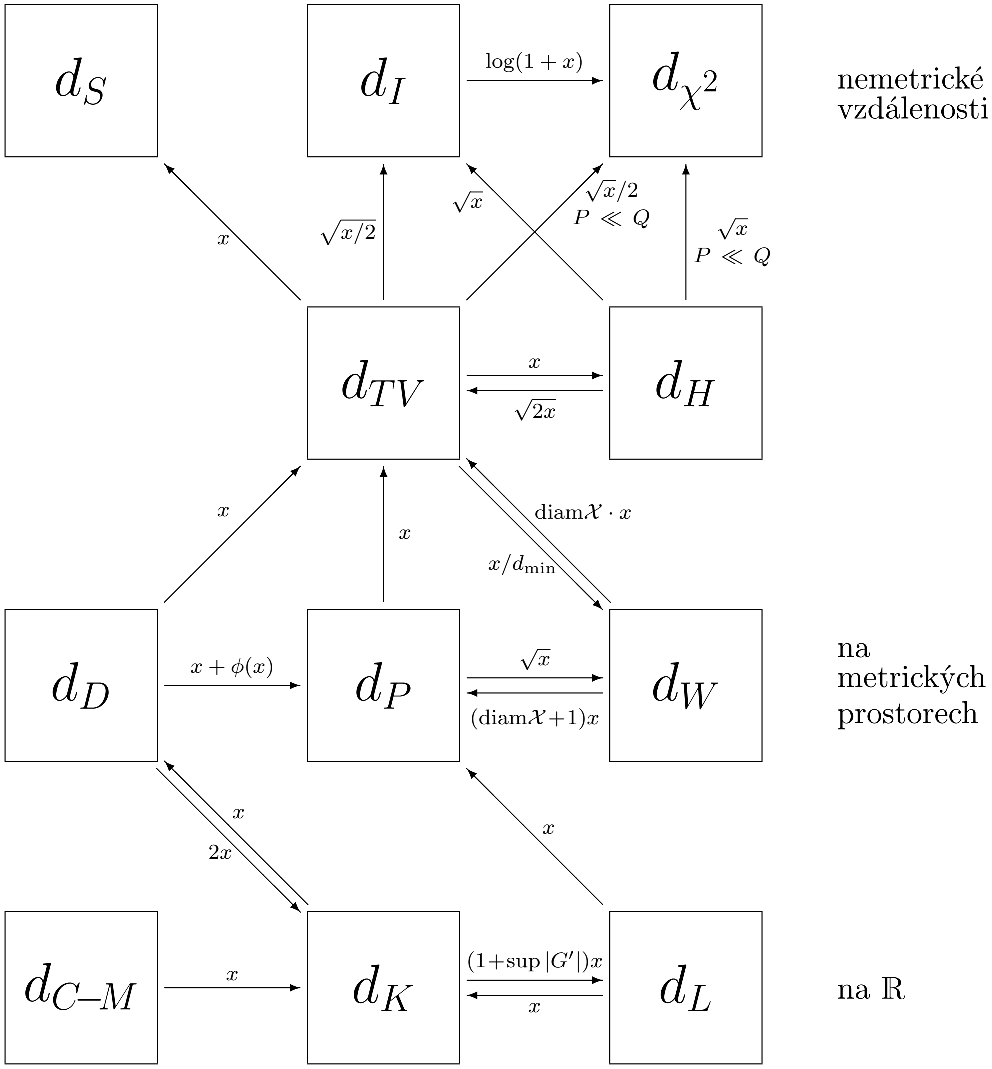
χ^2–vzdálenost a totální variace
Pro spočetný prostor $\mathcal{X}$, nebo pro spojitý prostor $\mathcal{X}$ pokud míra P dominuje míru Q platí

$$d_{TV}(P,Q) \leq \frac{d_{\chi^2}(P,Q)}{2}\,.$$

Kolmogorova a Cramer–von Mises vzdálenost

$$\begin{aligned} d_{C\text{–}M}(F,G) &= \int \Big(F(x)-G(x)\Big)^2 f(x)\,\mathrm{d}x \leq \int \sup_{x\in\mathbb{R}} |F(x)-G(x)| f(x)\,\mathrm{d}x \\ &= d_K(F,G)\int f(x)\,\mathrm{d}x = d_K(F,G)\,. \end{aligned}$$

Ostatní nerovnosti (viz následující schéma) mezi právě zmíněnými vzdálenostmi platí bez omezení (viz. [6]).



Obrázek 2.1: Schéma vztahů mezi pravděpodobnostními vzdálenostmi. Šipka z pole d_A do pole d_B s přiřazenou funkcí $h(x)$ a případným omezením znamená, že za daných omezení platí $d_A \leq h(d_B)$. Argumenty pravděpodobnostních vzdáleností jsou míry P, Q, omezení $P \operatorname{dom} Q$ znamená že daná nerovnost platí pouze pokud míra P dominuje míru Q. Další značení a omezení platnosti jsou zmíněna v popisu jednotlivých vzdáleností.

2.2 ϕ-divergence

Pro úplnost zmiňme ještě celou třídu vzdáleností, tzv. ϕ-divergence. Některé konkrétní ϕ-divergence byly již zmíněny v předchozí části. Zaveďme nyní ϕ-divergence obecně a uveďme některé jejich (pro nás důležité) vlastnosti.

Definice 5. Buďte $P, Q \in \mathcal{P}$ pravděpodobnostní míry absolutně spojité vzhledem k míře λ a označme $f = \mathrm{d}P/\mathrm{d}\lambda$ a $g = \mathrm{d}Q/\mathrm{d}\lambda$ Radon-Nikodymovy derivace P, Q podle λ. Potom ϕ-divergence mezi dvěma pravděpodobnostními mírami P, Q je definována vztahem

$$D_\phi(P,Q) := \int_{\mathcal{X}} g(x)\phi\left(\frac{f(x)}{g(x)}\right) \mathrm{d}\lambda(x), \quad \phi \in \Phi\,, \tag{2.1}$$

kde Φ je třída všech funkcí $\phi : (0,\infty) \to \mathbb{R}$ konvexních na $(0,\infty)$ a ostře konvexních v $t = 1$, s $\phi(1) = 0$. Na hranici otevřené oblasti $p, q > 0$ rozšiřujeme definici takto: $q\phi(p/q) = q \lim_{t\to 0_+} \phi(t)$ pro $p = 0$ a $p \lim_{t\to\infty}(\phi(t)/t)$ pro $q = 0$. Pro další postup označme limitu $\phi(0) = \lim_{t\to 0_+} \phi(t)$

Následující tabulka nabízí přehled významných divergenčních měr, které jsou speciálními případy ϕ-divergencí.

$\phi(x)$	Divergence (autor)
$x \log x - x + 1$	Kullback-Leibler
$-\log x + x - 1$	Minimální informační rozdíl
$(x-1)\log x$	J-divergence
$\frac{1}{2}(x-1)^2$	Pearson, Kagan
$\frac{(x-1)^2}{(x+1)^2}$	Balakrishnan a Sanghvi
$\frac{-x^s + s(x-1) + 1}{1-s}$, $s \neq 1$	Rathie a Kannappan
$\frac{1-x}{2} - \left(\frac{1+x^{-r}}{2}\right)^{-1/r}$, $r > 0$	Harmonický průměr (Mathai a Rathie)
$\frac{(1-x)^2}{2(a+(1-a)x)}$, $0 \leq a \leq 1$	Rukhin
$\frac{ax\log x - (ax+1-a)\log(ax+1-a)}{a(1-a)}$, $a \neq 0, 1$	Lin
$\frac{x^{\lambda+1} - x - \lambda(x-1)}{\lambda(\lambda+1)}$, $\lambda \neq 0, -1$	Cressie a Read
$\lvert 1-x^a \rvert^{1/a}$, $0 < a < 1$	Matusita
$\lvert 1-x \rvert^a$, $a \geq 1$	$\chi-$ divergence řádu a (Vajda); Totální variace pokud $a = 1$ (Saks)

Tabulka 2.2: Významné divergenční míry

V [15] je dokázaný odhad hodnot $\phi-$divergenci v následujícím znění.

Věta 1. *Pro každé dvě pravděpodobnostní P, Q platí*

$$0 \leq D_\phi(P,Q) \leq \phi(0) + \lim_{r\to\infty} \frac{\phi(r)}{r}\,, \tag{2.2}$$

kde

$$D_\phi(P,Q) = \phi(0) + \lim_{r\to\infty} \frac{\phi(r)}{r}\,, \tag{2.3}$$

pokud $\mathrm{supp}P \cap \mathrm{supp}Q = \emptyset$.

Následující věta (viz [12]) poskytuje horní odhad hodnot $\phi-$divergrnce pomocí totální variace.

Věta 2. *Pro všechna $\phi \in \Phi$ platí odhad*

$$0 \le D_\phi(P, Q) \le \left(\phi(0) + \lim_{r \to \infty} \frac{\phi(r)}{r}\right) \cdot \frac{d_{TV}(P, Q)}{2} \,. \tag{2.4}$$

Lemma 2 z článku [7] poskytuje jiný odhad hodnot $\phi-$divergence pomocí totální variace.

Věta 3. *Pokud $\phi \in \Phi_0$ pak pro všechny míry $P, Q \in \mathcal{P}$ platí*

$$0 \le D_\phi(P, Q) \le c_\phi d_{TV}(P, Q) \,,$$

kde $\Phi_0 \subset \Phi$ všech $\phi(x)$ striktně konvexních v bodě $x = 1$, Lipschitzovských na $(0, \infty)$ a takových, že $\phi(x) - x\phi(1/x)$ je lineární na $(0, \infty)$, $c_\phi < \infty$ je lipschitzovská konstanta funkce ϕ.

Pro nás budou zajímavé takové $\phi-$divergence, které splňují předpoklady věty 3, nebo ty, které splňují

$$\left(\phi(0) + \lim_{r \to \infty} \frac{\phi(r)}{r}\right) < \infty \,,$$

protože v podkapitole 4.4 odvodíme konzistenci a řád konzistence Kolmogorovských odhadů v těchto vzdálenostech.

Kapitola 3

Neparametrické odhady hustot konzistentní řádu $n^{-1/2}$ v L_1-normě

V této kapitole shrneme poznatky z článku [11], které budeme v následující kapitole zobecňovat. V následujících kapitolách budeme uvažovat pouze reálný měřitelný prostor s borelovskou $\sigma-$algebrou tedy $(\mathcal{X}, \mathcal{A}) = (\mathbb{R}, \mathcal{B})$ jinak ostatní značení zůstává bezezměny. Symbolem $\mathcal{F}$ označme množinu všech distribučních funkcí odpovídajících mírám z $\mathcal{P}$. ($F(x) = P(X \leq x)$ a $\mathcal{F}_\lambda$ její podmnožinu odpovídající mírám absolutně spojitým vzhledem k míře λ. Označme $\mathcal{D}_\lambda$ množinu v Banachově prostoru $L_1(\mathbb{R}, \mathrm{d}\lambda)$ obsahující hustoty odpovídající distribučním funkcím z $\mathcal{F}_\lambda$, $\mathcal{D}$ její libovolnou neprázdnou podmnožinu a $\mathcal{F}$ podmnožinu $\mathcal{F}_\lambda$, obsahující distribuční funkce odpovídající hustotám z podmnožiny $\mathcal{D}$.

Dále uvažujme libovolnou merickou vzdálenost d na $\mathcal{P}$, ta definuje psedometriku ρ_d na $\mathcal{D}_\lambda$ tímto způsobem: $\rho_d(f, g) = d(P, Q)$, kde $P, Q \in \mathcal{P}$ a jim odpovídající distribuční funkce $F, G \in \mathcal{F}_\lambda$ a $f, g \in \mathcal{D}_\lambda$ jsou odpovídající hustoty. Přejdeme-li k faktorprostoru, jehož prvky jsou třídy ekvivalence ($f \sim g \Leftrightarrow \rho_d(f, g) = 0$), stane se ρ_d metrikou na tomto prostoru. Takto vzniklý metrický prostor budeme i nadále značit $\mathcal{D}_\lambda$.

3.1 Konzistence v L_1-normě

Definice 6. Řekneme, že ρ_K dominuje ρ_{TV} na $\mathcal{D}$ (označíme $\rho_K \succ \rho_{TV}$) právě tehdy, když pro každou posloupnost $f, f_1, f_2, \dots \in \mathcal{D}$ konvergence $f_n \to f$ v ρ_K pro $n \to \infty$ implikuje konvergenci $f_n \to f$ v ρ_{TV}.

Definice 7. Řekneme, že ρ_K stejnoměrně dominuje ρ_{TV} lokálně vzhledem k ρ_K na $\mathcal{D}$ (označíme $\rho_K \succ^u \rho_{TV}/\rho_K$) právě tehdy, když pro každou hustotu $g \in \mathcal{D}$ existuje $c > 0$ a Kolmogorovské okolí hustoty g, $B_K(g) \subset \mathcal{D}$ takové, že $\rho_K(f, g) \geq c\, \rho_{TV}(f, g)$ pro všechny $f \in B_K(g)$.

Poznámka 1. Povšimněme si, že z $\rho_K \succ^u \rho_{TV}/\rho_K \Rightarrow \rho_K \succ \rho_{TV}$ a že opačná implikace neplatí.

Věta 4. *Nechť $\rho_K \succ \rho_{TV}$ na $\mathcal{D}$, potom každý Kolmogorovský odhad hustoty z $\mathcal{D}$ je konzistentní v L_1-normě. Pokud $\rho_K \succ^u \rho_{TV}/\rho_K$ na $\mathcal{D}$, potom každý Kolmogorovský odhad hustoty z $\mathcal{D}$ je konzistentní s rychlostí řádu $n^{-1/2}$ v L_1-normě.*

Věta 5. *Nechť $\rho_K \succ^u \rho_{TV}/\rho_K$ na $\mathcal{D}$, potom všechny Kolmogorovské odhady $\widehat{f_n}$ hustot z $\mathcal{D}$ jsou konzistentní s rychlostí řádu $n^{-1/2}$ ve střední hodnotě L_1-normy.*

Důkazy těchto vět zde nebudeme uvádět, protože v následující kapitole budou dokázány jejich zobecněné podoby. Důkazy lze najít v [11]. Navíc technika důkazů zobecněných a původních vět je podobná.

3.2 Stupeň variace

Následující příklad ukáže, že dominance $\rho_K \succ \rho_{TV}$ a $\rho_K \succ^u \rho_{TV}/\rho_K$ na $\mathcal{D}$ nejsou splněny automaticky na libovolné rodině $\mathcal{D} \subset \mathcal{D}_\lambda$.

Příklad 1. Buď $\mathcal{D} \subset \mathcal{D}_\lambda$ obsahující rovnoměrnou hustotu g na uzavřeném intervalu $[0,2]$. Označme d_n délku subintervalu při rovnoměrném rozdělení intervalu $[0,2]$ na 2^n disjunktních intervalů (tj. $d_n = \frac{1}{2^{n-1}}$). Předpokládejme, že $\mathcal{D}$ obsahuje posloupnost f_n definovanou pro každé $n \in \mathbb{N}$ výrazem

$$f_n(x) = \begin{cases} 2\sin^2\big(\pi(x2^{n-1} - 2k)\big) & x \in [2kd_n, (2k+1)d_n] \\ & k = 0, 1, \dots, 2^{n-1} - 1 \\ 0 & \text{jinde na } \mathbb{R}\,. \end{cases} \tag{3.1}$$

Lze snadno ukázat, že f_n jsou hustoty pravděpodobnosti. Pro vzdálenosti ρ_{TV} a ρ_K platí následující odhady

$$\rho_{TV}(f_n, g) = \int_0^2 \left| f_n(x) - \frac{1}{2} \right| \mathrm{d}x \geq 2^{n-1} \int_{d_n}^{2d_n} \frac{1}{2}\,\mathrm{d}x = \frac{1}{2}, \tag{3.2}$$

$$\rho_K(f_n, g) = \sup_{x \in [0,2]} \left| F_n(x) - \frac{x}{2} \right| \leq \int_0^{d_n} f_n(x)\,\mathrm{d}x = \left(\frac{1}{2}\right)^{n-1}, \tag{3.3}$$

což odporuje oběma dominancím $\rho_K \succ \rho_{TV}$ a $\rho_K \succ^u \rho_{TV}/\rho_K$ na $\mathcal{D}$ současně.

Budeme tedy formulovat podmínky, za kterých rodina $\mathcal{D} \subset \mathcal{D}_\lambda$ vyhoví podmínkám dominance předpokládaným ve větách 4 a 5.

Je známo, že každá dvojice hustot $f, g \in \mathcal{D}_\lambda$ definuje finitní míru ν s hustotou $\frac{\mathrm{d}\nu}{\mathrm{d}\lambda} = f - g$. Tato míra je rozdílem dvou finitních měr na $(\mathbb{R}, \mathcal{B})$, horní variace ν^+ a dolní variace ν^- s hustotami $\frac{\mathrm{d}\nu^+}{\mathrm{d}\lambda} = (f-g)^+ = \max\{0, f-g\}$ a ν^- s hustotou $\frac{\mathrm{d}\nu^-}{\mathrm{d}\lambda} = (f-g)^- = \max\{0, g-f\}$. Lze snadno ověřit, že $\rho_{TV}(f,g) = 2\nu^+(\mathbb{R}) = 2\nu^-(\mathbb{R})\ \forall\, f, g \in \mathcal{D}_\lambda$.

Definice 8. Řekneme, že $A \in \mathcal{B}$ separuje ν^+ a ν^-, právě když platí buď $\nu^+(A) = \nu^+(\mathbb{R})$ a $\nu^-(\mathbb{R} - A) = \nu^-(\mathbb{R})$, a nebo $\nu^+(\mathbb{R} - A) = \nu^+(\mathbb{R})$ a $\nu^-(A) = \nu^-(\mathbb{R})$.

Horní a dolní variace ν^+ a ν^- jsou vždy separovány nosičem horní variace ν^+, neboli množinou $A = \{x \in \mathbb{R} : (f-g)^+ > 0\}$. V následující definici se budeme zajímat o speciální formu nosiče A a $\mathbb{R} - A$.

Definice 9. Buďte $f, g \in \mathcal{D}_\lambda$. Potom stupeň variace $DV(f,g) \in [0, +\infty]$ je definován jako $DV(f,g) = 0$, když A separuje ν^+ a ν^- a $\lambda(A) = 0$. Jinak

$$DV(f,g) = \inf \left\{ m \in \mathbb{N} : \, A = \bigcup_{j=1}^{m} \mathcal{J}_j, \; A \text{ separuje } \nu^+, \, \nu^- \right\}, \tag{3.4}$$

kde $\mathcal{J}_1, \ldots, \mathcal{J}_m$ jsou neprázdné intervaly v $\mathbb{R}$. Je-li minimalizovaná množina prázdná, tj. neexistuje-li žádné m s požadovanými vlastnostmi, pokládáme $DV(f,g) = +\infty$.

Korektnost definice je lehce ověřitelná. Vidíme, že $DV(f,g) = 0$ právě tehdy, když f je rovno g skoro všude vzhledem k míře λ, tj. právě tehdy, když odpovídající distribuční funkce F a G jsou stejné. Definice je symetrická vůči f a g, tj. $DV(f,g) = DV(g,f)$. Případ $DV(f,g) = +\infty$ znamená, že rozdíl $f(x) - g(x)$ mění znaménko nekonečně mnohokrát na $\mathbb{R}$. Pro nenulový konečný stupeň variace $DV(f,g)$ může být počet znaménkových změn rozdílu $f - g$ redukován vhodnou volbou hustot f a g na množinách λ nulové míry na $2DV(f,g)$, pokud všechny intervaly $\mathcal{J}_j$ v definici 9 jsou omezené, nebo na $2DV(f,g)-1$, pokud jeden z intervalů $\mathcal{J}_j$ je neomezený. (Opravené tvrzení z článku [11], detailněji viz Dodatek 6.2.)

Definice 10. Pro danou $f \in \mathcal{D} \subset \mathcal{D}_\lambda$ a $\delta > 0$ definujeme lokální stupeň variace $LDV_\delta(f)$ hustoty f vzhledem ke Kolmogorovské vzdálenosti v $\mathcal{D}$ jako:

$$LDV_\delta(f) = \sup \left\{ DV(f,g) : \, g \in B_{K,\delta}(f) \cap \mathcal{D} \right\}, \tag{3.5}$$

kde $B_{K,\delta}(f)$ je Kolmogorovská koule v $\mathcal{D}$ o poloměru δ se středem v f. Je-li $\delta = +\infty$, interpretujeme $LDV_\infty(f)$ jako stupeň variace $f \in \mathcal{D}$ v rodině $\mathcal{D}$.

Věta 6. *Buď $\mathcal{D} \subset \mathcal{D}_\lambda$ a nechť pro každou hustotu $f \in \mathcal{D}$ existuje $\delta > 0$ takové, že $LDV_\delta(f) < +\infty$. Potom ρ_K stejnoměrně dominuje ρ_{TV} lokálně vzhledem k ρ_K na $\mathcal{D}$ ($\rho_K \succ^u \rho_{TV}/\rho_K$).*

Důkaz této věty zde nebudeme uvádět, protože v kapitole 4 bude dokázána obecnější věta.

Definice 11. Stupněm variace $DV(\mathcal{D})$ rodiny $\mathcal{D} \subset \mathcal{D}_\lambda$ nazveme

$$DV(\mathcal{D}) = \sup \left\{ DV(f,g) : f, g \in \mathcal{D} \right\}. \tag{3.6}$$

Poznámka 2. Pro všechny $f \in \mathcal{D}$ a $\delta > 0$ platí následující nerovnosti mezi jednotlivými stupni variace

$$LDV_\delta(f) \leq LDV_\infty(f) \leq DV(\mathcal{D}). \tag{3.7}$$

Tedy podle dosud uvedené teorie konečnost stupně variace každé hustoty $f \in \mathcal{D}$ nebo konečnost globální charakteristiky $DV(\mathcal{D})$ implikuje, že všechny Kolmogorovské odhady hustot z $\mathcal{D}$ jsou konzistentní řádu $n^{-1/2}$ v L_1-normě a ve střední hodnotě L_1-normy. Všimněme si, že není třeba požadovat žádnou hladkost nebo dodatečné předpoklady na chvosty funkcí (jako např. omezenost nosiče nebo požadavky na rychlost klesání v nekonečnech) v rodině $\mathcal{D}$. Navíc konečnost globální charakteristiky $DV(\mathcal{D})$ nám dovoluje vyhodnotit pravděpodobnost $P(\sqrt{n}\rho_{TV}(\widehat{f}_n, f) \geq k)$ nebo $E(\sqrt{n}\rho_{TV}(\widehat{f}_n, f))$ stejnoměrně pro všechny $f \in \mathcal{D}$, protože konstanta $c > 0$ vystupující ve větě 4 a 5 je nezávislá na skutečné hustotě $f \in \mathcal{D}$.

Kapitola 4

Zobecnění

V předchozí kapitole jsme viděli, že z předpokladu $\rho_K \succ^u \rho_{TV}/\rho_K$ plyne konzistence řádu $n^{-1/2}$ v L_1-normě a její střední hodnotě. Věta 6 nám dává postačující podmínku pro splnění dominance předpokládané ve větách 4 a 5. Tato podmínka nás však omezuje jen na hustoty $g \in B_K(f)$, pro které rozdíl $f - g$ mění znaménko jen konečně mnohokrát na množinách λ nenulové míry. To znamená, že budeme-li ověřovat dominanci pomocí věty 6 získáme konzistenci jen pro Kolmogorovské odhady, pro které rozdíl $\widehat{f}_n - f$ mění znaménko jen konečně mnohokrát.

Cílem této kapitoly je zobecnit teorii předešlé kapitoly tak, abychom v konečném důsledku dokázali, že i Kolmogorovské odhady, pro které $LDV_\delta(f)$ není konečný (tedy rozdíl $\widehat{f}_n - f$ může měnit znaménko nekonečně mnohokrát) jsou konzistentní v L_1-normě a její střední hodnotě. Za tímto účelem zavedeme nové typy dominancí a další teorii budeme budovat podobným způsobem jako v článku [11].

4.1 Konzistence v L_1-normě za obecnějších předpokladů

Definujme zobecnění dominancí z kapitoly 3 ve stejnoměrné i nestejnoměrné podobě.

Definice 12. Řekneme, že ρ_K asymptoticky dominuje ρ_{TV} řádu a_n lokálně s ohledem na ρ_K na $\mathcal{D}$ (označme $\rho_K \gtrsim \rho_{TV}/\rho_K\ (a_n \to 0)$) právě tehdy, když $(\forall f \in \mathcal{D})$ $(\exists B_K(f))$ takové, že $(\forall (f_n)_1^\infty \in B_K(f),\ f_n \to f \text{ v } \rho_K)$ $(\exists\ c > 0)$ tak, že

$$\rho_K(f_n, f) \geq c\rho_{TV}(f_n, f) - a_n\,, \tag{4.1}$$

kde $B_K(f)$ je Kolmogorovské okolí hustoty f a a_n je nezáporná posloupnost splňující $\lim\limits_{n\to+\infty} a_n = 0$.

Definice 13. Řekneme, že ρ_K asymptoticky stejnoměrně dominuje ρ_{TV} řádu a_n lokálně s ohledem na ρ_K na $\mathcal{D}$ (označme $\rho_K \gtrsim^u \rho_{TV}/\rho_K\ (a_n \to 0)$) právě tehdy, když $(\forall\, f \in \mathcal{D})$ $(\exists\ c > 0)$ $(\exists B_K(f))$ $(\forall\, (f_n)_1^\infty \in B_K(f), f_n \to f\ v\ \rho_K)$ tak, že

$$\rho_K(f_n, f) \geq c\rho_{TV}(f_n, f) - a_n\,, \tag{4.2}$$

kde $B_K(f)$ je Kolmogorovské okolí hustoty f a a_n je nezáporná posloupnost a $\lim\limits_{n\to+\infty} a_n = 0$.

Poznámka 3. Je zřejmé, že platí:

$$\rho_K \gtrsim^u \rho_{TV}/\rho_K \; (a_n \to 0) \Rightarrow \rho_K \gtrsim \rho_{TV}/\rho_K \; (a_n \to 0)\,. \tag{4.3}$$

Na první pohled by se mohlo zdát, že předpoklad $f_n \to f$ v ρ_K je zbytečný, ale protože tuto teorii budujeme pro Kolmogorovské odhady, které jej splňují, není pro nás omezující. A navíc kdyby byl tento předpoklad vypuštěn, definice 13 by se stala ekvivalentní s definicí 7. Toto tvrzení dokážeme.

Tvrzení 7. *Výroky A a B jsou ekvivalentní.*
A: $(\forall f \in \mathcal{D})$ $(\exists B_K^{(1)}(f) \subset \mathcal{D})$ a $(\exists\, c_1 > 0)$ taková, že $(\forall (f_n)_1^\infty \in B_K^{(1)}(f))$ platí $\rho_K(f_n, f) \geq c_1 \rho_{TV}(f_n, f) - a_n$, kde a_n je nezáporná posloupnost a $\lim\limits_{n\to+\infty} a_n = 0$.

B: $(\forall\, f \in \mathcal{D})$ $(\exists\, B_K^{(2)}(f) \subset \mathcal{D})$ a $(\exists c_2 > 0)$ taková, že $(\forall\, g \in B_K^{(2)}(f))$ platí $\rho_K(f, g) \;\geq\; c_2\, \rho_{TV}(f, g)$ (označme $\rho_K \succ \rho_{TV}/\rho_K$).

Důkaz. Implikace $B \Rightarrow A$ je zřejmá, za posloupnost a_n ve výroku A můžeme volit například posloupnost konstantně nulovou a $B_K^{(1)} = B_K^{(2)}$, $c_1 = c_2$.
Pro důkaz implikace $A \Rightarrow B$ zvolme libovolně pevně $f \in \mathcal{D}$, posloupnost z výroku A volme konstantní, ale jinak libovolnou.
Tedy $f_n = g$, $n \in \mathbb{N}$, kde $g \in B_K^{(1)}(f) \cap \mathcal{D}$ je libovolná. Podle výroku A k posloupnosti $(f_n)_1^\infty$ existuje $c_1 > 0$, že platí:

$$\rho_K(f_n, f) \geq c_1 \rho_{TV}(f_n, f) - a_n, \quad \forall\, n \in \mathbb{N}. \tag{4.4}$$

To pro naši konkrétní volbu posloupnosti znamená

$$\rho_K(g, f) \geq c_1 \rho_{TV}(g, f) - a_n, \quad \forall n \in \mathbb{N} \tag{4.5}$$

a limitním přechodem $n \to \infty$ získáváme

$$\rho_K(g, f) \geq c_1 \rho_{TV}(g, f), \quad \text{pro libovolnou } g \in B_K^{(1)} \cap \mathcal{D}. \tag{4.6}$$

Položíme-li $B_K^{(1)} = B_K^{(2)}$ a $c_1 = c_2$, je implikace $A \Rightarrow B$ dokázána. □

Vidíme tedy, že tento předpoklad se bez náhrady vypustit nedá. Ale dříve než jsme začali rozebírat předpoklady, bylo třeba ukázat, že definice 12 je zobecněním definice 7. Díky důkazu předešlého tvrzení nám k dokázání obecnějšího charakteru definice 12 stačí zkonstruovat příklad rodiny $\mathcal{D}$, na které není splněna dominance $\rho_K \succ \rho_{TV}/\rho_K$, ale je splněna dominance $\rho_K \gtrsim \rho_{TV}/\rho_K$ $(a_n \to 0)$.

Příklad 2. Nechť rodinu $\mathcal{D}$ tvoří rovnoměrná hustota g na intervalu $[a, b]$ a funkce $f_n^{d,h}$ definované vztahem

$$f_n^{d,h}(x) = \begin{cases} h \sin(n^2 \pi (x - d)) + \frac{1}{b-a} & x \in [d, d + \frac{2}{n}] \\ \frac{1}{b-a} & x \in [a, d) \cup (d + \frac{2}{n}, b] \\ 0 & \text{jinde na } \mathbb{R}\,, \end{cases} \tag{4.7}$$

kde $n \in \mathbb{N}$ a konstanty d, h splňují podmínky $d \in [a, b-2]$, $0 \leq h \leq \frac{1}{b-a}$.

Snadno ověříme, že $f_n^{d,h}$ jsou hustoty pravděpodobnosti. Ukážeme, že na této rodině $\mathcal{D}$ je splněna dominance $\rho_K \succsim \rho_{TV}/\rho_K$ $(a_n \to 0)$, ale není splněna dominance $\rho_K \succ \rho_{TV}/\rho_K$. Totiž

$$\rho_K(f_n^{d,h}, g) = \sup_{x\in\mathbb{R}} \Big| \int_{-\infty}^{x} (f_n^{d,h} - g)\,\mathrm{d}x \Big| = \int_{d}^{d+1/n^2} h \sin(n^2\pi(x-d))\,\mathrm{d}x = \frac{2h}{n^2\pi}\,, \tag{4.8}$$

$$\rho_{TV}(f_n^{d,h}, g) = \int_{\mathbb{R}} |f_n^{d,h} - g|\,\mathrm{d}x = \frac{4h}{n\pi}\,, \tag{4.9}$$

což odporuje dominanci $\rho_K \succ \rho_{TV}/\rho_K$. Ale dominance $\rho_K \succsim \rho_{TV}/\rho_K$ $(a_n \to 0)$ je splněna např. pro volbu posloupnosti $a_n = \frac{4h}{n\pi} - \frac{2h}{n^2\pi}$ a konstanty $c = 1$ v definici 12.

Rodinu $\mathcal{D}$ v příkladu by bylo možné značně rozšířit (aby neobsahovala pouze jednu posloupnost) uvolněním konstant d, h. Rodina by pak obsahovala kromě posloupností konvergujících pro $n \to \infty$ ke g také posloupnosti konvergující pro $d \to \widetilde{d}$ nebo pro $h \to \widetilde{h}$. Pak by se diskuse všech konvergentních posloupností z okolí libovolné hustoty stala značně složitější a příklad by ztratil potřebnou exemplárnost. Nyní vyslovíme a dokážeme obecnější podobu věty 4 a věty 5. K jejich důkazům budeme potřebovat Glivenko–Canteli Teorém (viz např. [3]) a odhad, který dokázali Dvoretzky, Kiefer a Wolfowitz a vylepšil Massart.(viz např. [3]).

Věta 8 (Glivenko-Canteli). *Buďte $X_1, \dots, X_n$ stejně a nezávisle rozdělené reálné náhodné veličiny s distribuční funkcí $F(x) = P(X_1 \le x)$ a $F_n(x)$ je empirická distribuční funkce. Potom pro každé $\varepsilon > 0$ platí*

$$P\left(\sup_{x\in\mathbb{R}} |F(x) - F_n(x)| > \varepsilon\right) \le 8(n+1)\exp\left(\frac{-n\varepsilon^2}{32}\right).$$

Věta 9 (Dvoretzky, Kiefer a Wolfowitz ; Massart)). *Buďte $X_1, \dots, X_n$ stejně a nezávisle rozdělené reálné náhodné veličiny s distribuční funkcí $F(x) = P(X_1 \le x)$ a $F_n(x)$ je empirická distribuční funkce. Potom pro každé $n \in \mathbb{N}$ a $\varepsilon > 0$ platí:*

$$P\left(\sup_{x\in\mathbb{R}} |F(x) - F_n(x)| > \varepsilon\right) \le 2\exp\left(-2n\varepsilon^2\right).$$

Věta 10. *Nechť $\rho_K \succsim \rho_{TV}/\rho_K$ $(a_n \to 0)$ na $\mathcal{D}$, potom každý Kolmogorovský odhad hustoty z $\mathcal{D}$ je konzistentní v L_1-normě. Pokud navíc $a_n = o(n^{-1/2})$, potom Kolmogorovský odhad hustoty z $\mathcal{D}$ je konzistentní řádu $n^{-1/2}$ v L_1-normě.*

Důkaz. 1. Konzistence:
Buď $\widehat{f}_n$ libovolný Kolmogorovský odhad hustoty f z $\mathcal{D}$. Potom

$$\rho_K(\widehat{f}_n, f) = K(\widehat{F}_n, F) \le K(\widehat{F}_n, F_n) + K(F_n, F) \le 2K(F_n, F) \qquad \text{s.j.} \tag{4.10}$$

První nerovnost plyne ze symetrie Kolmogorovské vzdálenosti a trojúhelníkové nerovnosti a druhá nerovnost z definice Kolmogorovského odhadu.
Dále z Glivenko–Canteliho teorému víme, že $K(F_n, F) \to 0$ s.j., a tedy podle poslední nerovnosti $\rho_K(\widehat{f}_n, f) \to 0$ s.j., z čehož plyne,

$$(\forall B_K(f))(\exists n_0)(\forall n > n_0)(\widehat{f}_n \in B_K(f)) \qquad \text{s.j.} \tag{4.11}$$

Použitím předpokladu $\rho_K \gtrsim \rho_{TV}/\rho_K$ $(a_n \to 0)$ dostáváme

$$\rho_K(\widehat{f}_n, f) \;\geq\; c\rho_{TV}(\widehat{f}_n, f) \;-\; a_n, \qquad \forall n > n_0, \tag{4.12}$$

odkud limitním přechodem $n \to \infty$

$$0 \geq c \lim_{n\to+\infty} \rho_{TV}(\widehat{f}_n, f). \tag{4.13}$$

Protože $\forall f, g \in \mathcal{D}$ je $\rho_{TV}(f,g) \geq 0$, získáváme z limity sevřené posloupnosti

$$\lim_{n\to+\infty} \rho_{TV}(\widehat{f}_n, f) = 0, \tag{4.14}$$

což dokazuje konzistenci v L_1-normě.

2. Konzistence řádu $n^{-1/2}$:
S využitím předpokladů a výsledku bodu 1 odhadujme pro každé $n > n_0$ a $k > 0$

$$\begin{aligned} P\left(\sqrt{n}\rho_{TV}(\widehat{f}_n, f) \geq k\right) &\leq P\left(\sqrt{n}(\rho_K(\widehat{f}_n, f) + a_n) \geq ck\right) \\ &\leq P\Big(\sqrt{n}(2K(F_n, F) + a_n) \geq ck\Big) \\ &= P\left(K(F_n, F) \geq \frac{ck - a_n\sqrt{n}}{2\sqrt{n}}\right). \end{aligned} \tag{4.15}$$

Je-li $a_n = o(n^{-1/2})$, pak existuje takové n_1, že pro všechna $n > n_1$ platí, že $\frac{ck - a_n\sqrt{n}}{2\sqrt{n}} > 0$ pro všechna k. Tedy na poslední výraz lze použít odhad z věty 9 (Dvoretzky, Kiefer a Wolfowitz; Massart). Pak získáváme

$$\begin{aligned} P(\sqrt{n}\rho_{TV}(\widehat{f}_n, f) \geq k) &\leq 2\exp\left(-\frac{(ck - a_n\sqrt{n})^2}{2}\right) \\ &= 2\exp\left(-\frac{(ck - o(1))^2}{2}\right) < \varepsilon\,. \end{aligned} \tag{4.16}$$

Odhad platí pro všechna k a pro všechna $n > \max\{n_0, n_1\}$, což dokazuje konzistenci řádu $n^{-1/2}$ v L_1-normě. □

Věta 11. *Nechť $\rho_K \gtrsim \rho_{TV}/\rho_K(a_n \to 0)$ na $\mathcal{D}$. Potom každý Kolmogorovský odhad $\widehat{f}_n$ hustoty f z $\mathcal{D}$ je konzistentní ve střední hodnotě L_1-normy. Pokud navíc $a_n = o(n^{-1/2})$, potom každý Kolmogorovský odhad hustoty f z $\mathcal{D}$ je konzistentní řádu $n^{-1/2}$ ve střední hodnotě L_1-normy.*

Důkaz. 1. Konzistence:
Pro důkaz konzistence ve střední hodnotě L_1-normy je třeba ověřit, že

$$\lim_{n\to+\infty} E\rho_{TV}(\widehat{f}_n, f) = 0. \tag{4.17}$$

Přitom víme, že $\widehat{f}_n \to f$ v ρ_K, tzn. $(\forall B_K(f))(\exists n_0)$ $(\forall n > n_0)(\widehat{f}_n \in B_K)$ s. j. Odhadujme tedy pro každé $n > n_0$

$$E\rho_{TV}(\widehat{f}_n, f) \leq \frac{1}{c}E(\rho_K(\widehat{f}_n, f) + a_n) = \frac{1}{c}E\rho_K(\widehat{f}_n, f) + \frac{a_n}{c}. \tag{4.18}$$

Z důkazu první části věty 10 víme, že $\rho_K(\widehat{f}_n, f) \leq 2K(F_n, F)$ s.j., dále tedy odhadujeme

$$\begin{aligned}\frac{1}{c}E\rho_K(\widehat{f}_n, f) + \frac{a_n}{c} &\leq \frac{2}{c}EK(F_n, F) + \frac{a_n}{c} \\ &= \frac{2}{c}\int\limits_0^\infty P(K(F_n, F) > u)\,\mathrm{d}u + \frac{a_n}{c} \\ &\leq \frac{4}{c}\int\limits_0^\infty \exp\left(-2nu^2\right)\mathrm{d}u + \frac{a_n}{c} = \frac{2}{c}\sqrt{\frac{\pi}{2n}} + \frac{a_n}{c}\,,\end{aligned} \tag{4.19}$$

kde poslední nerovnost plyne z věty 9 (Dvoretzky, Kiefer a Wolfowitz; Massart). A protože vzdálenost v totální variaci je vždy nezáporná, dostáváme z předchozích odhadů nerovnost

$$0 \leq E\rho_{TV}(\widehat{f}_n, f) \leq \frac{2}{c}\sqrt{\frac{\pi}{2n}} + \frac{a_n}{c}\,, \tag{4.20}$$

ze které limitou sevřené posloupnosti získáváme

$$\lim_{n\to+\infty} E\rho_{TV}(\widehat{f}_n, f) = 0\,. \tag{4.21}$$

2. Konzistence řádu $n^{-1/2}$
Pro důkaz konzistence řádu $n^{-1/2}$ je třeba ukázat, že

$$E\rho_{TV}(\widehat{f}_n, f) = O(n^{-1/2}). \tag{4.22}$$

Budeme proto odhadovat stejným způsobem jako v první části důkazu. Tím získáme následující nerovnost:

$$\sqrt{n}E\rho_{TV}(\widehat{f}_n, f) \leq \frac{4}{c}\int\limits_0^\infty \exp\left(-2u^2\right)\mathrm{d}u + \frac{a_n\sqrt{n}}{c} = \frac{2}{c}\sqrt{\frac{\pi}{2}} + \frac{a_n\sqrt{n}}{c}. \tag{4.23}$$

Protože $a_n = o(n^{-1/2})$, existuje konstanta K tak, že

$$|\sqrt{n}a_n| \leq K, \quad \forall\, n \in \mathbb{N}, \tag{4.24}$$

a tedy

$$\sqrt{n}E\rho_{TV}(\widehat{f}_n, f) \leq const., \tag{4.25}$$

což dokazuje konzistenci řádu $n^{-1/2}$ ve střední hodnotě L_1-normy. □

4.2 Parciální stupeň variace

Stejně jako v podkapitole 3.2 zjišťujeme, že dominance $\rho_K \gtrsim \rho_{TV}/\rho_K$ $(a_n \to 0)$ není splněna na libovolné rodině $\mathcal{D} \subset \mathcal{D}_\lambda$. Za příklad, který by to ukázal, můžeme znovu použít příklad 1. Připomeňme si jeho znění.

Hustoty pravděpodobnosti z rodiny $\mathcal{D}$ byly definovány p5edpisem

$$f_n(x) = \begin{cases} 2\sin^2\!\big(\pi(x2^{n-1} - 2k)\big) & \text{pro } x \in [2kd_n, (2k+1)d_n] \\ & k = 0, 1, \ldots, 2^{n-1} - 1 \\ 0 & \text{jinde na } \mathbb{R} \end{cases} \tag{4.26}$$

a g byla rovnoměrná hustota na intervalu $[0,2]$. Pro vzdálenosti ρ_{TV} a ρ_K platí odhady

$$\rho_{TV}(f_n,g) = \int\limits_0^2 \left| f_n(x) - \frac{1}{2} \right| \mathrm{d}x \geq 2^{n-1} \int\limits_{d_n}^{2d_n} \frac{1}{2} \,\mathrm{d}x = \frac{1}{2}, \tag{4.27}$$

$$\rho_K(f_n,g) = \sup_{x\in[0,2]} \left| F_n(x) - \frac{x}{2} \right| \leq \int\limits_0^{d_n} f_n(x)\,\mathrm{d}x = \left(\frac{1}{2}\right)^{n-1}. \tag{4.28}$$

Tedy posloupnost $f_n \to g$ v ρ_K, ale neexistuje taková nezáporná posloupnost $a_n \to 0$ a konstanta c, aby platilo $\rho_K(f_n,f) \geq c\rho_{TV}(f_n,f) - a_n$. To znamená, že na rodině z příkladu 1 není splněna asymptotická dominance $\rho_K \gtrsim \rho_{TV}/\rho_K$ $(a_n \to 0)$.

Nyní budeme formulovat podmínky, za kterých rodina $\mathcal{D} \subset \mathcal{D}_\lambda$ vyhoví podmínkám dominance předpokládaným ve větách 10 a 11. K tomu se nám budou hodit pojmy horní a dolní variace a dále též množiny, které je separují, zavedené v podkapitole 3.2. Definujme zobecněnou podobu stupně variace a podívejme se na jeho vztah k dříve definovanému stupni variace.

Definice 14. Buďte $f,g \in \mathcal{D}_\lambda$ a nechť $a \in [0,+\infty]$. Potom parciální stupeň variace $DV_a(f,g) \in [0,+\infty]$ je definován takto

$$DV_a(f,g) = \inf \left\{ m \in \mathbb{N} \cup \{0\} : \; A = \bigcup_{j=1}^{m} \mathcal{J}_j + I, \; A \text{ separuje } \nu^+, \, \nu^- \right\}, \tag{4.29}$$

kde $\mathcal{J}_1,\dots,\mathcal{J}_m$ jsou neprázdné disjunktní intervaly v $\mathbb{R}$ a $I \subset \mathbb{R}$ množina taková, že $\nu^+(I) \leq a$ a $\nu^-(I) \leq a$ přičemž $\bigcup\limits_{j=1}^{m} \mathcal{J}_j \cap I = \emptyset$. Je-li minimalizovaná množina prázdná, tj. neexistuje-li žádné m a množina I s požadovanými vlastnostmi, potom pokládáme $DV_a(f,g) = +\infty$.

Definice 15. Buď $f \in \mathcal{D} \subset \mathcal{D}_\lambda$ a $\delta > 0$. Definujme lokální parciální stupeň variace $LDV_{\delta,a}(f)$ hustoty f vzhledem ke Kolmogorovské vzdálenosti v $\mathcal{D}$

$$LDV_{\delta,a}(f) = \sup \{ DV_a(f,g) : g \in B_{K,\delta}(f) \cap \mathcal{D} \} \tag{4.30}$$

a parciální stupeň variace rodiny $\mathcal{D}$

$$DV_a(\mathcal{D}) = sup\{DV_a(f,g) : f,g \in \mathcal{D}\}. \tag{4.31}$$

Poznámka 4. Pokud $a > b > 0$, potom $DV_a \leq DV_b \leq DV_0 = DV$.

Vidíme, že parciální stupeň variace je vždy menší nebo roven dříve zavedenému stupni variace DV. Jak bylo uvedeno v předešlé kapitole, stupeň variace nás informuje o počtu znaménkových změn rozdílu $f - g$. Z parciálního stupně variace nezjistíme počet znaménkových změn, víme jen, že pokud $DV_a(f,g) < +\infty$, potom se až na konečný počet výjimek všechny znaménkové změny odehrávají na množině I. Budeme-li volit $a \in [0,+\infty]$ dostatečně velké, bude parciální stupeň variace nulový, v tom případě víme, že veškeré znaménkové změny rozdílu $f - g$ se odehrávají na množině I z definice DV_a. O jejich počtu nemůžeme tvrdit nic, dokonce ani nevíme, je-li jejich počet konečný. Stejně tak vidíme, že pokud je $DV(f,g) = +\infty$, lze volit $a \in [0,+\infty]$ tak, že $DV_a(f,g) < +\infty$. Pro nás zajímavý je případ, kdy je a voleno co nejmenší, takové, aby parciální stupeň variace zůstal konečný.

Nyní vyslovíme a dokážeme postačující podmínku pro splnění dominancí předpokládaných ve větách 10 a 11, neboli postačující podmínku pro konzistenci řádu $n^{-1/2}$ v L_1−normě i její střední hodnotě.

Věta 12. *Buď $\mathcal{D} \subset \mathcal{D}_\lambda$. Nechť pro každou hustotu $f \in \mathcal{D}$ existuje Kolmogorovské okolí $B_K(f)$ a konstanta $K \in [0, \infty)$ a nezáporná posloupnost b_n, taková, že $\lim_{n\to\infty} b_n = 0$ tak, že $(\forall (f_n)_1^\infty \in B_K(f),\ f_n \to f$ v $\rho_K)$ platí, že $\forall\, n \in \mathbb{N}$ je $DV_{b_n}(f_n, f) < K$. Potom ρ_K asymptoticky stejnoměrně dominuje ρ_{TV} řádu a_n lokálně s ohledem na ρ_K na $\mathcal{D}$.*

Důkaz. Zvolme $f \in \mathcal{D}$ libovolně pevně, k této f existuje konstanta K a Kolmogorovské okolí $B_K(f)$ takové, že $\forall (f_n)_1^\infty$ konvergující k f v ρ_K je $DV_{b_n}(f_n, f) < K$, $n \in \mathbb{N}$. Pro každé n označme ν_n^+ a ν_n^- horní a dolní variaci odpovídající hustotám f_n, f a označme A_n množinu z definice 14, která je separuje. Rozdělme $\mathbb{N}$ na dvě disjunktní podmnožiny $N_1,\ N_2$ takové, že $\mathbb{N} = N_1 \cup N_2$ a přitom pro $n \in N_1$ mají všechny množiny A_n vlastnost $\nu_n^+(A_n) = \nu_n^+(\mathbb{R})$ a pro $n \in N_2$ mají všechny množiny A_n vlastnost $\nu_n^-(A_n) = \nu_n^-(\mathbb{R})$.
Je-li $\forall\, n \in \mathbb{N}\ \ DV_{b_n}(f_n, f) = 0$, potom

$$\rho_{TV}(f_n, f) = 2\nu^+(\mathbb{R}) = 2\nu^+(A) = 2\nu^+(I) \le 2b_n \tag{4.32}$$

a tedy

$$\rho_{TV}(f_n, f) - 2b_n \le 0 \le \rho_K(f_n, f). \tag{4.33}$$

Nechť $\forall\ n \in \mathbb{N}$ je $DV_{b_n} = m_n < K < \infty$. Pro indexy $n \in \mathbb{N}$, pro které $m_n = 0$, máme nerovnost dokázanou v předchozím bodě. Pro ostatní indexy je $m_n \ge 1$ a pro $\forall\ n \in N_1$ platí

$$\begin{aligned}\nu_n^+(\mathbb{R}) = \nu_n^+(A) = \nu_n^+\Big(\bigcup_{j=1}^{m_n} \mathcal{J}_j + I\Big) &= \sum_{j=1}^{m_n} \nu_n^+(\mathcal{J}_j) + \nu_n^+(I) \\ &= \sum_{j=1}^{m_n}\Big(\int_{\mathcal{J}_j} \mathrm{d}F_n - \int_{\mathcal{J}_j} \mathrm{d}F\Big) + \nu_n^+(I),\end{aligned} \tag{4.34}$$

kde F_n, F jsou distribuční funkce odpovídající hustotám f_n, f.
Označme $a_j = \inf \mathcal{J}_j,\ b_j = \sup \mathcal{J}_j$ a definujme

$$F_n^*(b_j),\ F^*(b_j) = \begin{cases} F_n(b_j),\ F(b_j) & \text{pokud } b_j \in \mathcal{J}_j \\ F_n(b_j-),\ F(b_j-) & \text{pokud } b_j \notin \mathcal{J}_j\,, \end{cases} \tag{4.35}$$

kde $F(b_j-) = \lim_{b \to b_{j\,-}} F(b)$.

Pak platí:

$$\begin{aligned}\frac{1}{2}\rho_{TV}(f_n, f) = \nu_n^+(\mathbb{R}) &= \sum_{j=1}^{m_n}(F_n^*(b_j) - F_n(a_j)) - \sum_{j=1}^{m_n}(F^*(b_j) - F(a_j)) + \nu_n^+(I) \\ &= \sum_{j=1}^{m_n}(F_n^*(b_j) - F^*(b_j)) + \sum_{j=1}^{m_n}(F(a_j) - F_n(a_j)) + \nu_n^+(I) \\ &\le 2m_n \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| + b_n.\end{aligned} \tag{4.36}$$

Protože $\forall\, n \in \mathbb{N}$ je $m_n < K < \infty$, získáváme odtud nerovnost

$$\forall\, n \in N_1 \quad \rho_{TV}(f_n, f) \leq 4K\rho_K(f_n, f) + 2b_n\,. \tag{4.37}$$

Pro $n \in N_2$ se důkaz provede analogicky a získáme stejnou nerovnost, čímž máme dokázané tvrzení věty:

$$\forall\, n \in \mathbb{N} \quad \rho_K(f_n, f) \geq c\rho_{TV}(f_n, f) - a_n\,, \tag{4.38}$$

kde $c = \frac{1}{4K}$ a $a_n = 2b_n$ je nezáporná posloupnost konvergující k nule. □

Poznámka 5. Pokud rodina $\mathcal{D}$ splňuje, že $(\forall\, f \in \mathcal{D})$ $(\exists\, \delta > 0)$ takové, že $LDV_\delta(f) < +\infty$, potom vyhoví předpokladům věty 12.

Důkaz. Protože

$$LDV_\delta(f) = \sup\left\{DV(f, g) :\, g \in B_{K,\delta}(f) \cap \mathcal{D}\right\}, \tag{4.39}$$

pak pro každou hustotu $g \in B_{K,\delta}(f)$ a každé $a \geq 0$ platí:

$$DV_a(f, g) \leq LDV_\delta(f) = K < \infty\,. \tag{4.40}$$

Tedy nerovnost $DV_{b_n}(f_n, f) \leq K$ platí pro každou posloupnost $(f_n)_1^\infty \in B_{K,\delta}(f)$, kde b_n je libovolná nezáporná posloupnost konvergující k nule. □

4.3 Příklady rodin vyhovujících podmínkám asymptotické dominance

Zkonstruujme příklady rodin hustot s nekonečným stupněm variace, které přitom vyhovují podmínkám asymptotické dominance.

Jako první může být opět zmíněn příklad 2, použitý k demonstraci dobrého smyslu definice asymptotické dominance (definice 12) i s připojenou poznámkou o možnosti jeho rozšíření. Příklad snadno zobecníme následujícím způsobem. Pokud pro každou posloupnost $(f_n)_1^\infty$ z $\mathcal{D}$ takovou, že $f_n \to f$ v ρ_K platí, že znaménkové změny rozdílu hustot $f_n - f$ se odehrávají pouze na množinách A_n s Lebesqueovou mírou jdoucí k nule, pak pro splnění asymptotické dominance stačí, aby hustoty z $\mathcal{D}$ byly stejně omezené.

Další příklad již nebude vyžadovat zmenšování množin znaménkových změn, ale bude mít přísnější požadavky na hustoty, aby bylo možné splnit asymptotickou dominanci.

Příklad 3. Nechť rodinu $\mathcal{D}$ tvoří rovnoměrná hustota g na intervalu $[a, b]$ a funkce $f_n^{d,h}$ definované takto:

$$f_n^{h_n,c} = \begin{cases} \frac{1}{b-a} & \text{na } (c, b] \\ \frac{1}{b-a} + h_n \sin(\frac{2\pi n}{c-a}(x-a)) & \text{na } [a, c] \\ 0 & \text{jinde na } \mathbb{R}\,, \end{cases} \tag{4.41}$$

kde $n \in \mathbb{N}$, $a < c < b$ jsou pevně zvolené konstanty a posloupnost h_n $(h_n < \frac{1}{b-a})$ je pro všechna n klesající posloupnost s nulovou limitou.

Potom například pro posloupnost $a_n = \frac{2(c-a)}{\pi} h_n (1 - \frac{1}{4n})$ platí, že $\rho_K \gtrsim \rho_{TV}$ řádu a_n, neboť

$$\rho_{TV}(g, f_n^{h_n,c}) = 2n \int\limits_a^{a+\frac{c-a}{2n}} h_n \sin\Big(\frac{2\pi n}{c-a}(x-a)\Big)\, \mathrm{d}x = \frac{2(c-a)}{\pi} h_n\,, \tag{4.42}$$

$$\rho_K(g, f_n^{h_n,c}) = \int\limits_a^{a+\frac{c-a}{2n}} h_n \sin\Big(\frac{2\pi n}{c-a}(x-a)\Big)\, \mathrm{d}x = \frac{c-a}{\pi n} h_n\,, \tag{4.43}$$

odkud plyne

$$\rho_K(g, f_n^{h_n,c}) = \frac{c-a}{\pi n} h_n \geq \frac{c-a}{2\pi n} h_n = \rho_{TV}(g, f_n^{h_n,c}) - a_n\,. \tag{4.44}$$

Rodinu v příkladu je možné rozšířit tak, že budeme uvažovat výše definované hustoty pro všechny možné přípustné hodnoty parametrů a, b, c, h_n. Rodina pak bude obsahovat více než jen jednu posloupnost, jak je tomu v prezentovaném příkladu, ale také posloupnosti konvergující pro $n \to \infty$, $c \to \widetilde{c}$, $a \to \widetilde{a}$, $b \to \widetilde{b}$, přičemž asymptotická dominance zůstane i nadále splněná, jen diskuse všech možností se stane značně rozsáhlejší.

V následujícím příkladě se již nebude požadovat ani omezenost intervalu, na kterém se odehrává nekonečný počet znaménkových změn, oproti tomu na hustoty budou kladeny silnější požadavky.

Příklad 4. Nechť rodinu $\mathcal{D}$ tvoří hustota normálního rozdělení $g = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x-\mu)^2\right)$ a hustoty definované následně

$$f_n^{h_n,c}(x) = \begin{cases} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x-\mu)^2\right) & x \in I_1 \\ \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x-\mu)^2\right) + h_n \sin(x-\mu-c) & x \in I_2 \\ \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x-\mu)^2\right) + h_n \sin(\mu-c-x) & x \in I_3 \end{cases} \tag{4.45}$$

$$\begin{aligned} I_1 &= (-\infty, \mu-c-2\pi n) \cup (\mu-c, \mu+c) \cup (\mu+c+2\pi n, -\infty)\,, \\ I_2 &= [\mu+c, \mu+c+2\pi n]\,, \\ I_3 &= [\mu-c-2\pi n, \mu-c]\,, \end{aligned}$$

kde μ, c jsou reálné konstanty splňující $c \geq \mu$ a h_n je reálná klesající posloupnost taková, že $f_n^{h_n,c} \geq 0$ pro všechna přirozená n (např. $h_n \leq \frac{1}{2\sqrt{2\pi}} \exp\left(-\frac{1}{2}(c+2\pi n-\mu)^2\right)$). Potom, pokud se h_n chová alespoň jako $h_n = o(n^{-1})$, je v této rodině $\mathcal{D}$ splněna asymptotická dominance například pro posloupnost $a_n = h_n(8n-1) \to 0$, protože

$$\rho_{TV}(g, f_n^{h_n,c}) = 4n \int\limits_{\mu+c}^{\mu+c+\pi} h_n \sin(x-\mu-c)\, \mathrm{d}x = 8n h_n\,, \tag{4.46}$$

$$\rho_K(g, f_n^{h_n,c}) = \int\limits_{\mu+c}^{\mu+c+\pi} h_n \sin(x-\mu-c)\, \mathrm{d}x = 2h_n \tag{4.47}$$

a tedy platí podmínka asymptotické dominance

$$\rho_K(g, f_n^{h_n,c}) = 2h_n \geq h_n = \rho_{TV}(g, f_n^{h_n,c}) - a_n\,. \tag{4.48}$$

Rodinu z příkladu je možno obohatit stejným způsobem jako v příkladu 3, dále není nutné předpokládat, že znaménkových změn přibývá úměrně n a že periody sinů mají pro všechny hustoty stejnou délku. Stejně tak není nutné konstruovat hustoty $f_n^{h_n,c}$ symetricky. Mohli bychom klidně uvažovat narušení ve tvaru

$$\sin\Big(\frac{1}{l(n)}(x-\mu-c)\Big) \text{ na } (\mu+c, \mu+c+2\pi p(n)l(n))\,, \tag{4.49}$$

kde $p(n), l(n)$ jsou funkce závislé na n a vyjadřují změnu délky periody a přibývání znaménkových změn. Potom stejně jako v příkladu 4 zjistíme, že je splněna podmínka asymptotické dominance, pokud $h_n = o((p(n)l(n))^{-1})$.

Ve všech zde zkonstruovaných případech platí, že $\lim\limits_{n\to\infty} \rho_{TV}(f_n, f) = 0$ pro posloupnosti konvergující v Kolmogorovské vzdálenosti. Je otázkou, je-li možné splnit asymptotickou dominanci (netriviálním způsobem tak, aby rodina neobsahovala žádnou posloupnost hustot konvergující v Kolmogorovské vzdálenosti), aniž by dané posloupnosti nekonvergovaly v totální variaci. To možné není, protože je-li splněno

$$\begin{aligned}&\lim_{n\to\infty} \rho_K(f_n, f) = 0 \wedge \\ &\rho_K(f_n, f) \geq c\rho_{TV}(f_n, f) - a_n \text{ a } \lim_{n\to\infty} a_n = 0\,,\end{aligned} \tag{4.50}$$

potom z limity sevřené posloupnosti získáváme

$$0 = \lim_{n\to\infty} (\rho_K(f_n, f) + a_n) \geq c \lim_{n\to\infty} \rho_{TV}(f_n, f) \geq 0\,, \tag{4.51}$$

a tedy

$$\lim_{n\to\infty} \rho_{TV}(f_n, f) = 0\,. \tag{4.52}$$

Na příkladech jsme viděli rodiny, na kterých je splněna asymptotická dominance, která je postačující podmínkou pro konzistenci řádu $n^{-1/2}$ v L_1-normě (respektive ve střední hodnotě L_1-normy) Kolmogorovských odhadů.

4.4 Konzistence v jiných vzdálenostech

V této části zmíníme některé důsledky podkapitoly 4.1 a 4.2 a vzájemných vztahů mezi jednotlivými vzdálenostmi (viz kapitola 2).

Máme tři typy důsledků:

1. Pokud pro danou vzdálenost d platí nerovnost $d \leq h(d_{TV})$.
2. Pokud pro danou vzdálenost d platí nerovnost $d \leq h(d_K)$.
3. Pokud pro danou vzdálenost d platí obě nerovnosti $d \leq h_1(d_K)$ a $d_K \leq h_2(d)$.

Kde $h : \mathbb{R} \to \mathbb{R}$ je funkce jedné proměnné viz kapitola 2 obrázek 2.1. Pro přehlednost zaveďme základní předpoklady na funkci h, respektive funkce $h_i, i = 1, 2$ a rodinu hustot $\mathcal{D}$. Tyto předpoklady se nám v jednotlivých důsledcích budou neustále opakovat.

(P1) Nechť h je reálná funkce jedné proměnné spojitá v nule s vlastností $h(0) = 0$.

(P2) Nechť rodina hustot $\mathcal{D}$ splňuje předpoklady věty 12.

(P3) Nechť pro funkci h platí $h(x) = Kx$, nebo $h(x) \leq Kx$ alespoň na okolí nuly, kde K je reálná konstanta.

Důsledek 1
Pokud pro vzdálenost d platí, že $d \leq h(d_{TV})$ a předpoklad (P1) a (P2) pak každý Kolmogorovský odhad je konzistentní v dané vzdálenosti d. Pokud navíc platí předpoklad (P3), pak jsou konzistentní i v její střední hodnotě. Ovšem získaný řád konzistence nemusí být $n^{-1/2}$, neboť závisí na funkci h.

Konzistence ve vzdálenosti d a její střední hodnotě: z předpokladů a vlastností Kolmogorovského odhadu plyne, že

$$0 \leq \rho_d(\widehat{f}_n, f) \leq h(\rho_{TV}(\widehat{f}_n, f)) \tag{4.53}$$

$$0 \leq E\rho_d(\widehat{f}_n, f) \leq Eh(\rho_{TV}(\widehat{f}_n, f)) \,, \tag{4.54}$$

z důkazu věty 10 víme, že

$$\lim_{n\to\infty} \rho_{TV}(\widehat{f}_n, f) = 0 \tag{4.55}$$

$$E(\rho_{TV}(\widehat{f}_n, f)) \leq \frac{4}{c}\sqrt{\frac{\pi}{4n}} + \frac{a_n}{c} \tag{4.56}$$

a ze spojitosti funkce a předpokladu $h(0) = 0$ získáváme konzistenci ve vzdálenosti d. Pro konzistenci ve střední hodnotě vzdálenosti d je navíc potřeba předpoklad (P3). Potom kombinací tohoto předpokladu a vztahů (4.54) a (4.56) získáváme

$$0 \leq E\rho_d(\widehat{f}_n, f) \leq Eh(\rho_{TV}(\widehat{f}_n, f)) \leq KE(\rho_{TV}(\widehat{f}_n, f)) \leq \frac{4K}{c}\sqrt{\frac{\pi}{4n}} + \frac{Ka_n}{c} \,, \tag{4.57}$$

odkud již plyne konzistence ve střední hodnotě vzdálenosti d.

Řád konzistence: Pokud platí předpoklady (P1), (P2), (P3) a posloupnost z definice asymptotické dominance $a_n = o(n^{-1/2})$, pak řád konzistence zůstává $n^{-1/2}$.
Z důkazu věty 11 máme následující odhady

$$\begin{gathered}(\forall \varepsilon > 0)(\exists k > 0) \text{ že } P(\sqrt{n}\rho_{TV}(\widehat{f}_n, f) \geq k) < \varepsilon \,, \\ \sqrt{n}E\rho_{TV}(\widehat{f}_n, f) \leq \frac{4}{c}\sqrt{\frac{\pi}{4}} + \frac{\sqrt{n}a_n}{c} \leq const \,.\end{gathered} \tag{4.58}$$

Použitím předpokladů získáváme

$$\begin{aligned} P\left(\sqrt{n}\rho_d(\widehat{f}_n, f) \geq k\right) &\leq P\left(\sqrt{n}h(\rho_{TV}(\widehat{f}_n, f)) \geq k\right) \\ &\leq P\left(\sqrt{n}\rho_{TV}(\widehat{f}_n, f) \geq \frac{k}{K}\right) < \varepsilon \,, \\ \sqrt{n}E\rho_d(\widehat{f}_n, f) \leq \sqrt{n}Eh(\rho_{TV}(\widehat{f}_n, f)) &\leq \sqrt{n}KE\rho_{TV}(\widehat{f}_n, f) \\ &\leq \frac{4K}{c}\sqrt{\frac{\pi}{4}} + \frac{K\sqrt{n}a_n}{c} \,. \end{aligned} \tag{4.59}$$

Odkud již plyne konzistence řádu $n^{-1/2}$ ve vzdálenosti d i její střední hodnotě.

Důsledek 2
Pokud pro vzdálenost d platí, že $d \leq h(d_K)$ a platí předpoklad (P1), pak každý Kolmogorovský odhad $\widehat{f}_n$ hustoty f je konzistentní ve vzdálenosti d. Pokud navíc platí předpoklad (P3), pak je odhad konzistentní i ve střední hodnotě vzdálenosti d, neboť

$$\begin{aligned}
&\rho_d(\widehat{f}_n, f) \leq h(\rho_K(\widehat{f}_n, f)) \leq h(2d_K(F_n, F)),\\
&E\rho_d(\widehat{f}_n, f) \leq Eh(\rho_K(\widehat{f}_n, f)) \leq Eh(2d_K(F_n, F)) \leq 2KEd_K(F_n, F)
\end{aligned} \tag{4.60}$$

odkud již plyne konzistence.

<u>Řád konzistence</u>: Pokud platí předpoklady (P1) a (P3), pak řád konzistence zůstává $n^{-1/2}$, neboť

$$\begin{aligned}
&(\forall \varepsilon > 0)(\exists k > 0) \text{ že platí}\\
&P\left(\sqrt{n}\rho_d(\widehat{f}_n, f) \geq k\right) \leq P\left(\sqrt{n}h(\rho_K(\widehat{f}_n, f)) \geq k\right) \leq P\left(\sqrt{n}\rho_K(\widehat{f}_n, f) \geq \frac{k}{K}\right)\\
&\leq 2\exp\left(\frac{k}{2K^2}\right) < \varepsilon
\end{aligned} \tag{4.61}$$

a pro střední hodnoty

$$\sqrt{n}E\rho_d(\widehat{f}_n, f) \leq \sqrt{n}Eh(\rho_K(\widehat{f}_n, f)) \leq \sqrt{n}KE\rho_K(\widehat{f}_n, f) \leq 4K\sqrt{\frac{\pi}{4}} \tag{4.62}$$

odkud již plyne řád konzistence.

Důsledek 3
Nechť pro vzdálenost d platí nerovnosti $d_K \leq h_1(d)$ a $d \leq h_2(d_K)$ a nechť platí předpoklady (P1), (P2) a navíc jsou funkce h_i rostoucí. Dále pokud vzdálenost d není přímo metrika nechť splňuje alespoň trojúhelníkovou nerovnost. Pak každý odhad s minimální vzdáleností d $\widehat{f}_n$ hustoty $f \in \mathcal{D}$ je konzistentní v L_1-normě.
Z předpokladů plyne

$$\begin{aligned}
0 \leq \rho_{TV}(\widehat{f}_n, f) &\leq \frac{1}{c}\rho_K(\widehat{f}_n, f) + \frac{a_n}{c} \leq \frac{1}{c}h_1\left(\rho_d(\widehat{f}_n, f)\right) + \frac{a_n}{c}\\
&\leq \frac{1}{c}h_1\left(2d(F_n, F)\right) + \frac{a_n}{c} \leq \frac{1}{c}h_1\left(2h_2(\rho_K(F_n, F))\right) + \frac{a_n}{c},
\end{aligned} \tag{4.63}$$

odkud již plyne konzistence odhadů s minimální d vzdáleností v L_1-normě. Pro konzistenci ve střední hodnotě L_1-normy je dále třeba předpoklad (P3). Pak získáváme

$$\begin{aligned}
E\rho_{TV}(\widehat{f}_n, f) &\leq \frac{1}{c}E\rho_K(\widehat{f}_n, f) + \frac{a_n}{c} \leq \frac{1}{c}Eh_1\left(\rho_d(\widehat{f}_n, f)\right) + \frac{a_n}{c} \leq \frac{1}{c}EK_1\rho_d(\widehat{f}_n, f) + \frac{a_n}{c}\\
&\leq \frac{2K_1}{c}Ed(F_n, F) + \frac{a_n}{c} \leq \frac{2K_1}{c}Eh_2\left(\rho_K(F_n, F)\right) + \frac{a_n}{c}\\
&\leq \frac{2K_1K_2}{c}E\rho_K(F_n, F) + \frac{a_n}{c}
\end{aligned} \tag{4.64}$$

odkud již plyne konzistence ve střední hodnotě L_1-normy.

Řád konzistence: Nechť platí dříve vyslovené předpoklady a navíc posloupnost a_n splňuje $a_n = o(n^{-1/2})$, pak každý odhad s minimální d vzdáleností je konzistentní v L_1-normě i její střední hodnotě. Z předpokladů získáváme

$$P\left(\sqrt{n}\rho_{TV}(\widehat{f}_n, f) \geq k\right) \leq P\left(\sqrt{n}\rho_K(\widehat{f}_n, f) + a_n \geq ck\right) \leq P\left(\rho_d(\widehat{f}_n, f) \geq \frac{ck - \sqrt{n}a_n}{K_1\sqrt{n}}\right)$$
$$\leq P\left(d(F_n, F) \geq \frac{ck - \sqrt{n}a_n}{2K_1\sqrt{n}}\right) \leq P\left(d_K(F_n, F) \geq \frac{ck - \sqrt{n}a_n}{2K_2K_1\sqrt{n}}\right), \tag{4.65}$$

odkud již plyne konzistence řádu $n^{-1/2}$ v L_1-normě. Stejným postupem získáme i konzistenci v její střední hodnotě, odhadujeme stejně jako v (4.64) a získáme.

$$\sqrt{n}E\rho_{TV}(\widehat{f}_n, f) \leq \frac{2K_1K_2\sqrt{n}}{c}E\rho_K(F_n, F) + \frac{a_n\sqrt{n}}{c} \tag{4.66}$$

Odkud již plyne konzistence ve střední hodnotě L_1-normy.

Na závěr ještě několik poznámek k předpokladům kladeným na funkci h, respektive na funkce h_i. Pokud bychom opustili požadavek splnění nerovností $h_i(x) \leq K_i x$ na okolí nuly, získali bychom přímo jen konzistenci v dané vzdálenosti. Pokud nahradíme tento předpoklad podmínkou, že funkce h (respektive $h_i i = 1, 2$) jsou takové, že pro ně platí $Eh(X) \leq h(EX)$, kde X je náhodná veličina, získáme i konzistenci ve střední hodnotě dané vzdálenosti. (Z Jensenovy nerovnosti plyne, že podmínku $Eh(X) \leq h(EX)$ splňují všechny konkávní funkce.) S řádem konzistence to již není tak jednoduché viz následující příklad.

Příklad 5. Uvažujme situaci kdy pro vzdálenost d platí $d \leq h(d_{TV})$ a hledejme řád konzistence δ_n Kolmogorovských odhadů ve vzdálenosti d. Předpokládejme, že funkce h je rostoucí taková, že je možné provést následující odhady.

$$\begin{aligned} P\left(\rho_d(\widehat{f}_n, f) \geq k\delta_n\right) &\leq P\left(h(\rho_{TV}(\widehat{f}_n, f)) \geq k\delta_n\right) = P\left(\rho_{TV}(\widehat{f}_n, f) \geq h^{-1}(k\delta_n)\right) \\ &\leq P\left(\rho_K(\widehat{f}_n, f) \geq c{\cdot}h^{-1}(k\delta_n) - a_n\right) \\ &\leq P\Big(d_K(F_n, F) \geq \underbrace{2c{\cdot}h^{-1}(k\delta_n) - a_n}_{\gamma}\Big) \\ &\leq 2\exp\left(-2n\gamma^2\right) \overset{?}{<} \varepsilon \end{aligned} \tag{4.67}$$

Poslední odhad platí pro každé $\gamma > 0$ a každé $n \in \mathbb{N}$. Odtud vyplývá podmínka $\gamma = 2c{\cdot}h^{-1}(k\delta_n) - a_n > 0$ alespoň od jistého n_0, díky čemuž mohou vznikat požadavky na posloupnost a_n. Pokud pro každé $\varepsilon > 0$ existuje $k > 0$ a $n_1 \in \mathbb{N}$, že pro všechna $n > n_1$ je poslední výraz menší než ε, pak jsme získali řád konzistence δ_n. Je tedy vidět, že při tomto způsobu odvozování řád konzistence velmi záleží na funkci h.

Jak je uvedeno v kapitole 2 mezi vzdálenostmi platí následující nerovnosti, rozdělíme je do tří kategorií podle tří typů důsledků.

Vzdálenosti splňující $d \le h(d_{TV})$

- $d_D \le d_{TV}$
- $d_P \le d_{TV}$
- $d_L \le d_P \le d_{TV} \Rightarrow d_L \le d_{TV}$
- $d_W \le \operatorname{diam}(\mathcal{X}) d_{TV}$
- $d_H \le \sqrt{2 d_{TV}}$

konzistence Kolmogorovských odhadů řádu $n^{-1/2}$ ve vzdálenosti d i její střední hodnotě

neplatí $\sqrt{x} \le Kx$

Vzdálenosti splňující $d \le h(d_K)$

- $d_D \le 2d_K$
- $d_L \le d_K$
- $d_{C\text{–}M} \le d_K$

konzistence Kolmogorovských odhadů řádu $n^{-1/2}$ v daných vzdálenostech i jejich středních hodnotách

Vzdálenosti splňující $d \le h_1(d_K)$ a zároveň $d_K \le h_2(d)$

- $d_D \le 2d_K \quad \wedge \quad d_K \le d_D$
- $d_L \le d_K \quad \wedge \quad d_K \le (1 + \sup |G'|) d_L$

konzistence řádu $n^{-1/2}$ odhadů s minimální vzdáleností d v L_1−normě i její střední hodnotě

Do první skupiny vzdáleností patří ještě všechny ϕ−divergenční vzdálenosti, které splňují předpoklady věty 3 a podle věty 2 i všechny divergence, pro které platí

$$\left(\phi(0) + \lim_{r \to \infty} \frac{\phi(r)}{r} \right) \le konst.$$

Na závěr se pokusme vyšetřit konzistenci a její řád v Hellingerově vzdálenosti. Z předchozích úvah přímo plyne konzistence v Hellingerově vzdálenosti (viz. Důsledek 1). Hledejme její řád δ_n. Předpokládejme, že jsou splněny předpoklady věty 12 (tj. Kolmogorovská vzdálenost asymptoticky dominuje vzdálenost v totální variaci řádu a_n, který určíme později z požadavků, které vyplynou v průběhu odvození).

$$\begin{aligned} P\left(\rho_H(\widehat{f}_n, f) \ge k\delta_n\right) &\le P\left(\sqrt{2\rho_{TV}(\widehat{f}_n, f)} \ge k\delta_n\right) = P\left(\sqrt{\rho_{TV}(\widehat{f}_n, f)} \ge \frac{k\delta_n}{\sqrt{2}}\right) \\ &= P\left(\rho_{TV}(\widehat{f}_n, f) \ge \frac{(k\delta_n)^2}{2}\right) \le P\left(\rho_K(\widehat{f}_n, f) \ge \frac{c(k\delta_n)^2 - 2a_n}{2}\right) \\ &\le P\left(d_K(F_n, F) \ge \frac{ck^2\delta_n^2 - 2a_n}{4}\right) \\ &\overset{\mathbf{1}}{\le} 2\exp\left(-2n\frac{ck^2\delta^2 - 2a_n}{4}\right) \overset{\mathbf{2}}{\le} \varepsilon \end{aligned} \tag{4.68}$$

Odhad **1** je možné provést pokud existuje k_0 a n_0 tak, že pro všechna $k > k_0$ a $n > n_0$ platí $\frac{ck^2\delta_n^2 - 2a_n}{4} > 0$, odkud vyplývá, že $a_n = O(\delta_n^2)$.

Odhad **2** je možné provést, pokud existuje n_1 tak, že $2n\frac{ck^2\delta^2-2a_n}{4} \overset{k\to\infty}{\to} \infty$ pro všechna $n > n_1$. Tedy

$$2n\frac{ck^2\delta^2 - 2a_n}{4} = \underbrace{\frac{2n\delta_n^2}{4}}_{\substack{n\to\infty \\ \nrightarrow 0}} \underbrace{\left(ck^2 - 2\frac{a_n}{\delta_n^2}\right)}_{\overset{k\to\infty}{\to}\infty} \tag{4.69}$$

odkud vyplývá, že nejlepší řád konzistence, který je takto možno získat je $\delta_n = n^{-1/2}$ a posloupnost a_n z asymptotické dominance musí splňovat $a_n = O(n^{-1})$.

Nyní zkoumejme konzistenci ve střední hodnotě Hellingerově vzdálenosti. Pro střední hodnoty platí $E\sqrt{X} \leq \sqrt{EX}$, tedy získáme konzistenci ve střední hodnotě Hellingerovy vzdálenosti i její řád. Předpokládejme, že jsou splněny předpoklady věty 12 (tj. Kolmogorovská vzdálenost asymptoticky dominuje vzdálenost v totální variaci řádu a_n, který určíme později z požadavků, které vyplynou v průběhu odvození).

$$E\rho_H(\widehat{f}_n, f) \leq E\sqrt{2\rho_{TV}(\widehat{f}_n, f)} \leq \sqrt{2}\sqrt{E\rho_{TV}(\widehat{f}_n, f)} \leq \sqrt{\frac{8}{c}\sqrt{\frac{\pi}{4n}} + \frac{a_n}{c}}, \tag{4.70}$$

odkud již plyne konzistence ve střední hodnotě Hellingerovy vzdálenosti. Zkoumejme její řád δ_n. Provedeme stejné odhady jako v (4.70).

$$\frac{1}{\delta_n}E\rho_H(\widehat{f}_n, f) \leq \sqrt{\frac{8}{c}\sqrt{\frac{\pi}{4n\delta_n^4}} + \frac{a_n}{\delta_n^2 c}}. \tag{4.71}$$

Odkud vyplývá, že nejlepší možný řád konzistence ve střední hodnotě, který je takto možné odvodit je $\delta_n = n^{-1/4}$ a posloupnost a_n z definice asymptotické dominance musí splňovat $a_n = o(n^{-1/2})$.

4.5 Numerická simulace

V této části uvedeme výsledky numerické simulace odhadů s minimální Kolmogorovskou, diskrepanční, Lévyho a Cramer–von Mises vzdáleností. Pro Kolmogorovské, Lévyho a diskrepanční odhady máme teoreticky odvozený řád konzistence v L_1-normě, který je dobře patrný z následujících grafů. Pro Cramer–von Mises vzdálenost se nepodařilo ukázat nerovnost s Kolmogorovskou vzdáleností, která by podle podkapitoly 4.4 zaručovala konzistenci v L_1-normě a její řád pro odhady s minimální Cramer–von Mises vzdáleností. Nicméně, jak je patrné z numerických výsledků, vykazují odhady s minimální Cramer–von Mises vzdáleností jistou konzistenci v L_1-normě. Výpočetní program vznikl modifikací a doplněním programu, který byl součástí Diplomové práce [5] a který je dále modifikován a rozšiřován v Diplomové práci [1]. Odhady se počítají pro vzorky generované ze známých rozdělení (normální, Laplaceovo, Cauchyho, logistické, rovnoměrné). Implementace rozdělení s nekonečným stupněm variace se nezdařila.

Normální rozdělení $N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Rovnoměrné rozdělení s parametry a, b

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{pro } x \in (a,b) \\ 0 & \text{pro } x \notin (a,b) \end{cases}$$

Cauchyho rozdělení s parametry a, λ

$$f(x) = \frac{1}{\pi}\left[\frac{\lambda}{(x-a)^2+\lambda^2}\right]$$

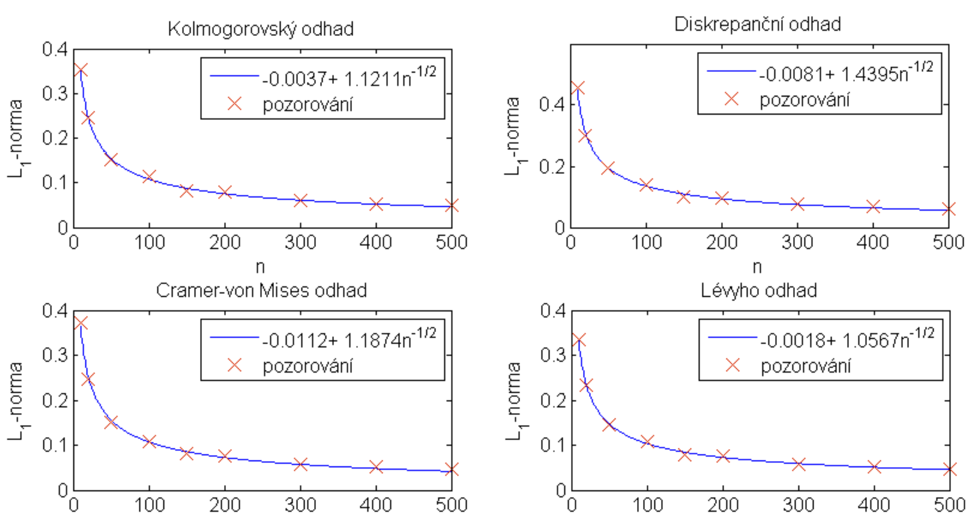
Laplaceovo (dvojtě exponenciální) rozdělení s parametry μ, b

$$f(x) = \frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right)$$

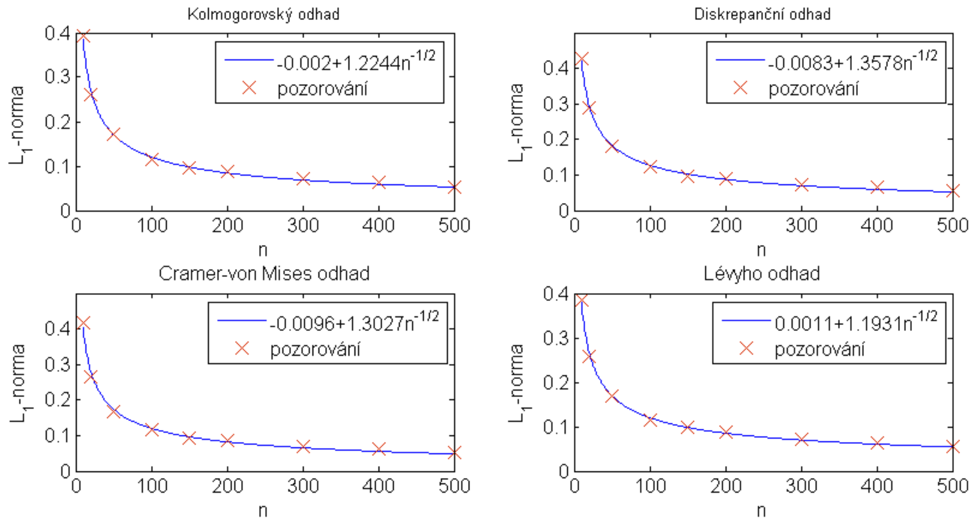
Logistické rozdělení s parametry μ, s

$$f(x) = \frac{\exp\left(-\frac{x-\mu}{s}\right)}{s\left(1+\exp\left(-\frac{x-\mu}{s}\right)\right)^2}$$

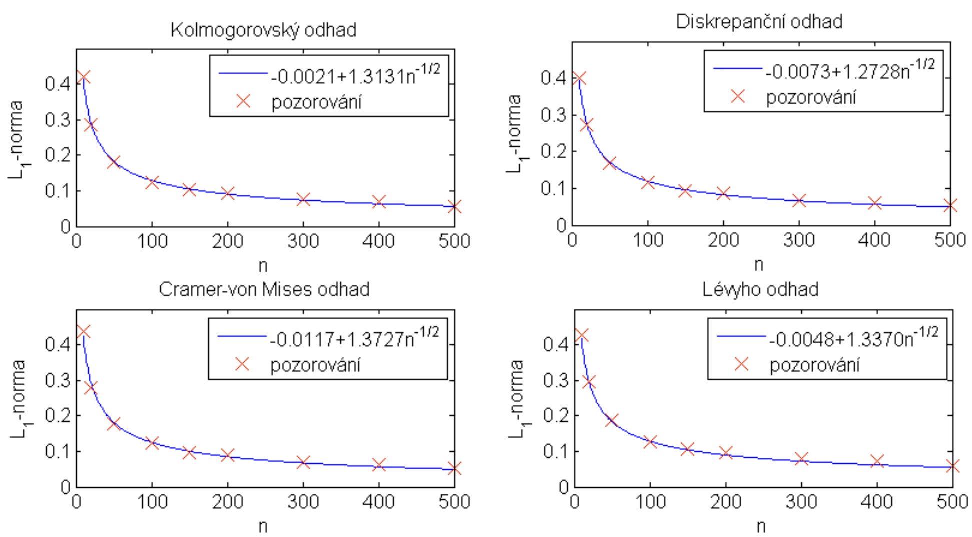
Velikosti vzorků jsou postupně $n = 10, 20, 50, 100, 150, 200, 300, 400, 500$, pro každý vzorek se počítalo 500 odhadů, ve výsledcích jsou uvedeny jejich průměrné hodnoty. Následující grafy znázorňují závislost vzdálenosti v totální variaci (L_1-normu) skutečné hustoty od získaného odhadu. Dále následují tabulky se všemi daty získanými simulací.



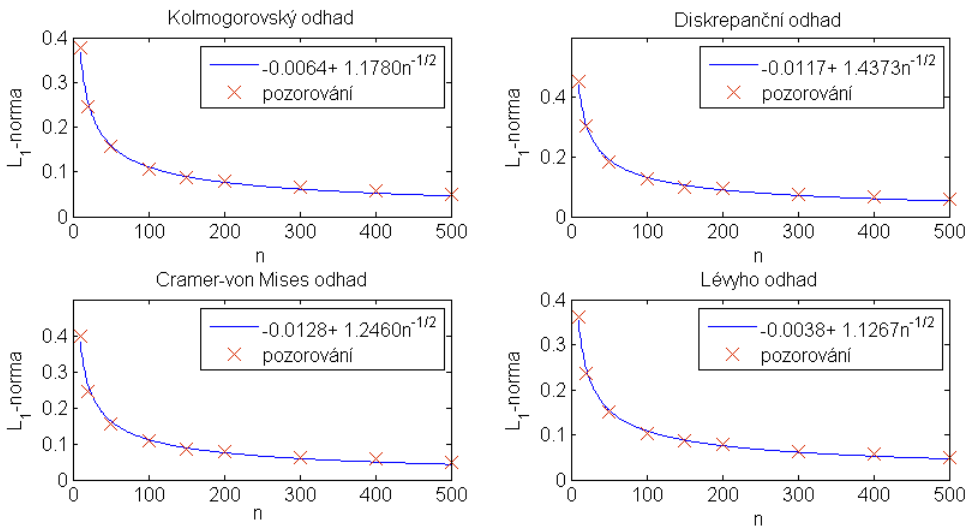
Obrázek 4.1: Odhady z normálního rozdělení s parametry $\mu = 1, \sigma^2 = 1$.



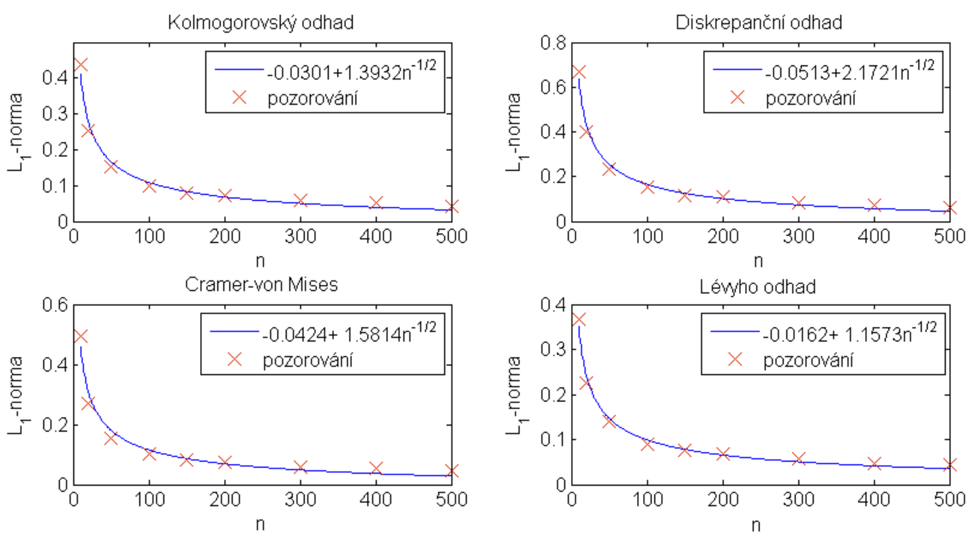
Obrázek 4.2: Odhady z Laplaceova rozdělení s parametry $\mu = 1, b = 1$.



Obrázek 4.3: Odhady z Cauchyova rozdělení s parametry $a = 1, \lambda = 1$.



Obrázek 4.4: Odhady z logistického rozdělení s parametry $\mu = 1, s = 2$.



Obrázek 4.5: Odhady z rovnoměrného rozdělení s parametry $a = 1, b = 1$.

Z výše uvedené teorie (viz podkapitola 4.4) víme, že odhady s minimální Kolmogorovskou, diskrepanční a Lévyho vzdáleností jsou konzistentní řádu $n^{-1/2}$ v L_1-normě. Nasimulovaná data jsme proložili funkcí $f(n) = a_0 + a_1 n^{-1/2}$ (koeficienty jsme získali metodou nejmenších čtverců). Jak je vidět z grafů nasimulovaná data dobře odpovídají teoretickým výsledkům. Pro odhad s minimální Cramer–von Mises vzdáleností jsme neukázali nerovnost, která by podle podkapitoly 4.4 zaručovala konzistenci v L_1-normě. Ale dobré proložení dat získaných simulací funkcí $f(n) = a_0 + a_1 n^{-1/2}$ nás vede k myšlence, že i odhady s minimální Cramer–von Mises vzdáleností jsou konzistentní řádu $\backsim n^{-1/2}$ v L_1-normě. Ovšem z tak malého počtu dat lze jen těžko něco usuzovat.

V tabulkách budou uvedeny numerické hodnoty jednotlivých odhadů. Uvádíme průměrné hodnoty (označené pruhem) a směrodatné odchylky (ozn. $\sqrt{S(\cdot)}$) odhadnutých parametrů, vzdálenost v totální variaci skutečné hustoty od odhadu (ozn $V(f_0)$) a její rozptyl. Pro zajímavost také uvádíme euklidovskou vzdálenost skutečných a odhadnutých parametrů. Prováděly se odhady s minimální Kolmogorovskou, diskrepanční, Cramer–von Mises a Lévyho vzdáleností (v tabulkách ozn.: $K, D, C-M, L$).

n	D	$\overline{\mu}$	$\overline{\sigma^2}$	$\sqrt{S(\mu)}$	$\sqrt{S(\sigma^2)}$	$\overline{V(f_0)}$	$\sigma_V(f_0)$	$\overline{\lvert\mu-\mu_0\rvert}$	$\overline{\lvert\sigma^2-\sigma_0^2\rvert}$
10	K	0.982	0.949	0.308	0.574	0.353	0.202	0.248	0.452
	D	0.968	0.848	0.437	0.578	0.456	0.247	0.359	0.481
	$C-M$	0.832	0.971	0.297	0.581	0.372	0.222	0.280	0.469
	L	0.982	0.976	0.302	0.558	0.335	0.191	0.244	0.433
20	K	0.997	0.997	0.235	0.403	0.245	0.130	0.188	0.326
	D	0.991	0.949	0.322	0.405	0.301	0.158	0.257	0.336
	$C-M$	0.922	1.021	0.231	0.404	0.246	0.138	0.195	0.328
	L	0.997	1.015	0.233	0.384	0.232	0.118	0.188	0.310
50	K	1.010	0.969	0.150	0.246	0.153	0.079	0.122	0.203
	D	1.019	0.944	0.210	0.258	0.193	0.097	0.172	0.216
	$C-M$	0.978	0.977	0.148	0.241	0.150	0.076	0.121	0.200
	L	1.008	0.982	0.147	0.235	0.145	0.075	0.119	0.190
100	K	1.017	0.975	0.117	0.179	0.113	0.061	0.096	0.144
	D	1.019	0.971	0.159	0.194	0.140	0.074	0.131	0.158
	$C-M$	1.001	0.984	0.113	0.177	0.108	0.060	0.092	0.140
	L	1.015	0.977	0.115	0.163	0.107	0.056	0.093	0.131
150	K	0.995	0.995	0.085	0.140	0.081	0.047	0.067	0.109
	D	0.991	0.987	0.113	0.149	0.101	0.054	0.091	0.120
	$C-M$	0.985	0.995	0.084	0.138	0.081	0.046	0.067	0.108
	L	0.995	0.991	0.084	0.133	0.079	0.046	0.065	0.104
200	K	1.007	0.984	0.080	0.127	0.078	0.042	0.065	0.102
	D	1.005	0.982	0.112	0.135	0.097	0.053	0.090	0.109
	$C-M$	1.000	0.990	0.078	0.125	0.075	0.041	0.062	0.099
	L	1.006	0.984	0.078	0.122	0.075	0.041	0.063	0.097
300	K	0.998	0.998	0.064	0.097	0.059	0.033	0.050	0.076
	D	0.995	0.997	0.087	0.105	0.075	0.040	0.070	0.085
	$C-M$	0.993	1.001	0.062	0.096	0.058	0.032	0.049	0.076
	L	0.996	0.995	0.063	0.092	0.058	0.032	0.049	0.071
400	K	1.000	0.990	0.058	0.083	0.053	0.029	0.046	0.067
	D	0.998	0.989	0.080	0.089	0.067	0.035	0.065	0.072
	$C-M$	0.997	0.993	0.056	0.082	0.051	0.029	0.044	0.066
	L	0.999	0.988	0.057	0.081	0.052	0.028	0.044	0.066
500	K	1.006	0.994	0.050	0.079	0.049	0.025	0.040	0.064
	D	1.007	0.993	0.069	0.083	0.060	0.031	0.055	0.067
	$C-M$	1.001	0.997	0.049	0.075	0.047	0.024	0.039	0.061
	L	1.002	0.992	0.049	0.077	0.048	0.023	0.039	0.061

Tabulka 4.1: Normalni rozdělení s parametry $\mu = 1, \sigma^2 = 1$, počet odhadů 500

n	D	$\overline{\mu}$	$\bar{b}$	$\sqrt{S(\mu)}$	$\sqrt{S(b)}$	$\overline{V(f_0)}$	$\sigma_V(f_0)$	$\overline{\lvert\mu-\mu_0\rvert}$	$\overline{\lvert b-b_0\rvert}$
10	K	0.993	0.957	0.400	0.381	0.393	0.207	0.307	0.298
	D	1.001	0.899	0.419	0.393	0.427	0.230	0.313	0.325
	$C-M$	0.831	0.959	0.411	0.386	0.417	0.217	0.339	0.305
	L	0.991	0.976	0.409	0.374	0.386	0.203	0.315	0.291
20	K	1.004	1.009	0.261	0.263	0.262	0.125	0.205	0.211
	D	1.010	0.989	0.294	0.269	0.290	0.140	0.233	0.214
	$C-M$	0.923	1.012	0.258	0.259	0.264	0.132	0.213	0.207
	L	1.001	1.015	0.262	0.257	0.258	0.126	0.209	0.206
50	K	0.998	0.997	0.167	0.167	0.170	0.088	0.135	0.134
	D	0.999	0.990	0.175	0.174	0.179	0.091	0.139	0.139
	$C-M$	0.968	0.999	0.164	0.163	0.168	0.087	0.135	0.129
	L	0.997	0.999	0.173	0.156	0.168	0.091	0.138	0.122
100	K	0.994	1.010	0.113	0.114	0.116	0.061	0.091	0.092
	D	1.003	1.004	0.118	0.123	0.123	0.064	0.095	0.098
	$C-M$	0.979	1.010	0.113	0.115	0.117	0.061	0.093	0.092
	L	0.994	1.010	0.118	0.106	0.116	0.064	0.095	0.085
150	K	1.004	0.988	0.091	0.092	0.096	0.049	0.075	0.076
	D	1.005	0.986	0.092	0.093	0.097	0.051	0.073	0.077
	$C-M$	0.994	0.989	0.090	0.087	0.094	0.048	0.074	0.072
	L	1.002	0.986	0.096	0.089	0.098	0.050	0.079	0.075
200	K	1.001	0.998	0.083	0.086	0.087	0.045	0.068	0.068
	D	1.002	0.999	0.088	0.091	0.091	0.050	0.071	0.073
	$C-M$	0.995	1.001	0.082	0.082	0.086	0.044	0.067	0.066
	L	1.000	0.996	0.086	0.081	0.088	0.044	0.071	0.065
300	K	1.003	0.995	0.066	0.068	0.071	0.036	0.053	0.053
	D	1.007	0.993	0.067	0.072	0.072	0.038	0.055	0.058
	$C-M$	0.997	0.995	0.065	0.065	0.069	0.035	0.052	0.052
	L	0.999	0.993	0.071	0.065	0.073	0.038	0.056	0.052
400	K	1.000	1.001	0.060	0.063	0.064	0.033	0.048	0.050
	D	1.001	0.999	0.060	0.066	0.065	0.034	0.048	0.053
	$C-M$	0.997	1.000	0.058	0.061	0.062	0.032	0.047	0.049
	L	0.998	0.999	0.062	0.061	0.064	0.034	0.050	0.048
500	K	0.999	0.999	0.049	0.055	0.053	0.029	0.039	0.044
	D	1.001	0.997	0.050	0.058	0.055	0.030	0.040	0.047
	$C-M$	0.996	1.000	0.048	0.054	0.052	0.029	0.038	0.044
	L	0.996	0.997	0.052	0.055	0.055	0.030	0.041	0.044

Tabulka 4.2: Laplaceovo rozdělení s parametry $\mu = 1, b = 1$, počet odhadů 500

n	D	$\overline{a}$	$\overline{\lambda}$	$\sqrt{S(a)}$	$\sqrt{S(\lambda)}$	$\overline{V(f_0)}$	$\sigma_V(f_0)$	$\overline{\lvert a-a_0\rvert}$	$\overline{\lvert\lambda-\lambda_0\rvert}$
10	K	0.975	1.063	0.685	0.596	0.419	0.214	0.502	0.432
	D	1.006	0.967	0.550	0.536	0.400	0.210	0.406	0.406
	$C-M$	0.745	1.026	0.695	0.568	0.436	0.221	0.534	0.418
	L	0.969	1.088	0.714	0.610	0.427	0.218	0.526	0.446
20	K	1.006	1.074	0.413	0.401	0.284	0.138	0.319	0.304
	D	1.012	1.027	0.368	0.350	0.273	0.130	0.292	0.270
	$C-M$	0.890	1.057	0.408	0.370	0.279	0.142	0.326	0.280
	L	1.000	1.081	0.429	0.415	0.295	0.142	0.334	0.316
50	K	0.997	1.016	0.253	0.232	0.182	0.095	0.202	0.178
	D	1.001	1.000	0.216	0.216	0.169	0.085	0.173	0.168
	$C-M$	0.955	1.011	0.244	0.216	0.176	0.091	0.199	0.167
	L	0.992	1.019	0.268	0.241	0.189	0.101	0.213	0.182
100	K	0.992	1.021	0.173	0.149	0.124	0.065	0.139	0.120
	D	1.003	1.009	0.151	0.149	0.116	0.061	0.122	0.118
	$C-M$	0.971	1.016	0.167	0.146	0.122	0.064	0.137	0.116
	L	0.986	1.022	0.180	0.152	0.128	0.069	0.146	0.121
150	K	1.006	0.991	0.138	0.120	0.102	0.053	0.112	0.098
	D	1.007	0.988	0.119	0.110	0.092	0.048	0.094	0.091
	$C-M$	0.992	0.989	0.132	0.107	0.096	0.048	0.108	0.088
	L	1.002	0.989	0.147	0.124	0.108	0.055	0.119	0.102
200	K	1.003	1.005	0.126	0.113	0.093	0.048	0.102	0.089
	D	1.003	1.004	0.110	0.110	0.085	0.047	0.088	0.087
	$C-M$	0.992	1.005	0.122	0.103	0.088	0.046	0.099	0.082
	L	1.000	1.001	0.132	0.114	0.097	0.049	0.107	0.091
300	K	1.003	0.997	0.099	0.090	0.075	0.038	0.079	0.071
	D	1.008	0.995	0.086	0.085	0.068	0.035	0.070	0.068
	$C-M$	0.996	0.995	0.095	0.080	0.070	0.035	0.077	0.064
	L	0.997	0.994	0.105	0.092	0.079	0.041	0.084	0.071
400	K	1.001	1.005	0.090	0.083	0.068	0.035	0.073	0.066
	D	1.001	1.001	0.076	0.079	0.061	0.032	0.061	0.063
	$C-M$	0.996	1.002	0.085	0.076	0.064	0.033	0.069	0.061
	L	0.994	1.002	0.096	0.083	0.071	0.036	0.077	0.067
500	K	0.998	1.001	0.074	0.073	0.057	0.032	0.058	0.058
	D	1.001	0.998	0.064	0.069	0.052	0.028	0.050	0.056
	$C-M$	0.994	1.000	0.071	0.066	0.054	0.029	0.056	0.054
	L	0.991	0.999	0.079	0.074	0.060	0.034	0.063	0.059

Tabulka 4.3: Cauchyovo rozdělení s parametry $a = 1, \lambda = 1$, počet odhadů 500

n	D	$\overline{\mu}$	$\overline{s}$	$\sqrt{S(\mu)}$	$\sqrt{S(s)}$	$\overline{V(f_0)}$	$\sigma_V(f_0)$	$\overline{\lvert\mu-\mu_0\rvert}$	$\overline{\lvert s-s_0\rvert}$
10	K	0.994	0.923	0.595	0.315	0.377	0.210	0.469	0.255
	D	1.003	0.860	0.707	0.328	0.451	0.244	0.554	0.290
	$C-M$	0.743	0.927	0.598	0.319	0.398	0.223	0.511	0.259
	L	0.992	0.940	0.590	0.306	0.362	0.199	0.466	0.243
20	K	1.007	0.988	0.396	0.216	0.245	0.119	0.316	0.176
	D	1.022	0.957	0.529	0.224	0.303	0.148	0.428	0.186
	$C-M$	0.876	0.990	0.389	0.213	0.246	0.126	0.328	0.173
	L	1.004	0.996	0.391	0.210	0.237	0.116	0.312	0.171
50	K	1.001	0.989	0.260	0.136	0.156	0.081	0.210	0.109
	D	1.003	0.978	0.325	0.144	0.184	0.093	0.262	0.117
	$C-M$	0.948	0.991	0.255	0.135	0.154	0.081	0.212	0.107
	L	1.000	0.991	0.259	0.131	0.152	0.081	0.208	0.105
100	K	0.992	1.004	0.179	0.095	0.107	0.056	0.145	0.076
	D	1.005	0.996	0.226	0.103	0.127	0.066	0.183	0.083
	$C-M$	0.966	1.005	0.178	0.095	0.107	0.056	0.146	0.076
	L	0.991	1.004	0.178	0.091	0.104	0.056	0.144	0.072
150	K	1.007	0.988	0.143	0.076	0.087	0.045	0.117	0.063
	D	1.009	0.985	0.177	0.078	0.100	0.052	0.142	0.065
	$C-M$	0.989	0.988	0.142	0.073	0.086	0.044	0.117	0.060
	L	1.004	0.986	0.144	0.074	0.087	0.043	0.118	0.062
200	K	1.002	0.995	0.131	0.070	0.079	0.042	0.108	0.056
	D	1.003	0.994	0.172	0.075	0.093	0.052	0.138	0.060
	$C-M$	0.991	0.998	0.130	0.068	0.078	0.039	0.107	0.055
	L	0.999	0.993	0.130	0.068	0.078	0.040	0.108	0.054
300	K	1.003	0.994	0.106	0.056	0.065	0.033	0.085	0.045
	D	1.013	0.992	0.130	0.060	0.074	0.039	0.105	0.049
	$C-M$	0.994	0.994	0.104	0.054	0.063	0.031	0.083	0.044
	L	0.999	0.992	0.106	0.054	0.064	0.033	0.085	0.043
400	K	1.000	1.000	0.095	0.052	0.058	0.030	0.077	0.042
	D	1.001	0.998	0.116	0.055	0.066	0.035	0.093	0.044
	$C-M$	0.995	0.999	0.092	0.051	0.057	0.029	0.075	0.041
	L	0.996	0.999	0.094	0.051	0.057	0.029	0.076	0.041
500	K	0.998	0.998	0.077	0.046	0.049	0.027	0.061	0.037
	D	1.001	0.997	0.096	0.049	0.057	0.030	0.077	0.040
	$C-M$	0.993	0.999	0.077	0.045	0.048	0.026	0.061	0.036
	L	0.993	0.997	0.077	0.045	0.049	0.026	0.061	0.036

Tabulka 4.4: Logisticke rozdělení s parametry $\mu = 1, s = 1$, počet odhadů 500

n	D	$\overline{a}$	$\overline{b}$	$\sqrt{S(a)}$	$\sqrt{S(b)}$	$\overline{V(f_0)}$	$\sigma_V(f_0)$	$\overline{\lvert a-a_0\rvert}$	$\overline{\lvert b-b_0\rvert}$
10	K	1.061	1.934	0.153	0.148	0.437	0.282	0.126	0.123
	D	1.093	1.870	0.222	0.205	0.669	0.372	0.190	0.192
	$C-M$	1.023	1.877	0.162	0.159	0.496	0.355	0.120	0.153
	L	1.038	1.955	0.136	0.134	0.366	0.220	0.113	0.111
20	K	1.022	1.978	0.093	0.096	0.252	0.142	0.076	0.080
	D	1.040	1.945	0.158	0.152	0.399	0.220	0.127	0.130
	$C-M$	0.997	1.950	0.093	0.102	0.271	0.175	0.072	0.087
	L	1.014	1.983	0.085	0.089	0.224	0.121	0.069	0.074
50	K	1.010	1.989	0.064	0.056	0.152	0.088	0.050	0.047
	D	1.022	1.981	0.102	0.098	0.234	0.131	0.084	0.080
	$C-M$	0.999	1.980	0.062	0.058	0.155	0.098	0.047	0.047
	L	1.008	1.990	0.058	0.053	0.140	0.075	0.046	0.045
100	K	0.999	1.996	0.042	0.039	0.099	0.052	0.033	0.032
	D	1.009	1.993	0.070	0.069	0.154	0.089	0.056	0.056
	$C-M$	0.995	1.991	0.041	0.040	0.102	0.061	0.032	0.031
	L	0.996	1.996	0.038	0.037	0.090	0.046	0.031	0.030
150	K	1.005	1.995	0.032	0.031	0.080	0.041	0.026	0.026
	D	1.013	1.995	0.054	0.053	0.117	0.070	0.043	0.042
	$C-M$	1.003	1.992	0.033	0.030	0.082	0.047	0.026	0.024
	L	1.002	1.996	0.030	0.030	0.076	0.039	0.024	0.025
200	K	1.002	1.998	0.029	0.028	0.072	0.037	0.024	0.023
	D	1.010	1.998	0.048	0.050	0.107	0.061	0.039	0.040
	$C-M$	1.000	1.995	0.029	0.028	0.074	0.041	0.023	0.022
	L	1.000	1.997	0.029	0.027	0.068	0.038	0.023	0.022
300	K	1.001	1.997	0.023	0.023	0.059	0.027	0.019	0.020
	D	1.006	1.997	0.038	0.038	0.083	0.048	0.031	0.029
	$C-M$	1.001	1.995	0.022	0.024	0.059	0.033	0.018	0.019
	L	1.000	1.996	0.023	0.023	0.056	0.030	0.018	0.019
400	K	1.000	1.998	0.020	0.020	0.050	0.023	0.017	0.017
	D	1.003	1.996	0.033	0.034	0.073	0.041	0.027	0.027
	$C-M$	1.000	1.997	0.019	0.022	0.052	0.029	0.015	0.017
	L	0.999	1.998	0.019	0.020	0.048	0.026	0.015	0.016
500	K	0.999	1.998	0.016	0.017	0.042	0.022	0.014	0.014
	D	1.005	1.999	0.028	0.028	0.061	0.035	0.022	0.022
	$C-M$	0.999	1.998	0.017	0.018	0.045	0.026	0.013	0.014
	L	0.999	1.998	0.018	0.017	0.043	0.026	0.013	0.014

Tabulka 4.5: Rovnoměrné rozdělení s parametry $a = 1, b = 2$, počet odhadů 500

Kapitola 5

Vapnik Chervonenkisova teorie

V následujících kapitolách se budeme zabývat jinými přístupy, které poskytují podmínky pro konzistenci a řád konzistence odhadů s minimální vzdáleností. V této kapitole zavedeme Vapnik Chrevonenkisovu dimenzi (dále jen VC dimenzi) a pojmy k tomu nezbytné. Následně vyslovíme větu o konzistenci odhadů s minimální zobecněnou Kolmogorov-Smirnovovou (GKS) vzdáleností na rodinách hustot s konečnou VC dimenzí a dokážeme ji.

5.1 Vapnik-Chervonenkisova dimenze

Definice 16. Buď $\mathcal{A}$ třída měřitelných množin. Pro n-tici $(z_1, \ldots, z_n) \in \{\mathbb{R}^d\}^n$ označme $N_{\mathcal{A}}(z_1, \ldots z_n)$ počet různých množin ve třídě

$$\{\{z_1, \ldots, z_n\} \cap A; A \in \mathcal{A}\} . \tag{5.1}$$

Dále n-tý shatter koeficient definujme jako

$$s(\mathcal{A}, n) = \max_{(z_1, \ldots, z_n) \in \{\mathbb{R}^d\}^n} N_{\mathcal{A}}(z_1, \ldots, z_n) . \tag{5.2}$$

To znamená, že shatter koeficient je maximální počet různých podmnožin z n bodů, které mohou být vybrány pomocí třídy množin $\mathcal{A}$.

Je zřejmé, že platí $s(\mathcal{A}, n) \leq 2^n$, neboť n prvková množina má právě 2^n různých podmnožin. Pokud $N_{\mathcal{A}}(z_1, \ldots, z_n) = 2^n$ pro nějaké $(z_1, \ldots, z_n)$, pak říkáme, že $\mathcal{A}$ roztřídí $\{z_1, \ldots, z_n\}$. Pokud $N_{\mathcal{A}}(z_1, \ldots, z_n) < 2^n$, potom každá n prvková množina má takovou podmnoži-nu, že neexistuje množina v $\mathcal{A}$ obsahující právě tuto podmnožinu n prvků. Platí-li tedy $s(\mathcal{A}, k) \leq 2^k$ pro nějaké přirozené k, potom $s(\mathcal{A}, n) \leq 2^n$ pro všechna $n > k$. Případ, kdy poprvé nenastane rovnost, je důležitý.

Definice 17. Buď $\mathcal{A}$ třída množin $|\mathcal{A}| \geq 2$, kde $|\mathcal{A}|$ označuje počet prvků množiny $\mathcal{A}$. Největší přirozené číslo $k \geq 1$, pro které platí, že $s(\mathcal{A}, k) = 2^k$, označme $V_{\mathcal{A}}$ a nazývejme ji Vapnik-Chervonenkisovou (VC) dimenzí třídy $\mathcal{A}$. Pokud $s(\mathcal{A}, n) = 2^n$ pro všechna n, pak položme $V_{\mathcal{A}} = \infty$.

Třídy množin $\mathcal{A}$, pro které $V_{\mathcal{A}} \leq \infty$ jsou nazývány Vapnik-Chervonenkisovi třídy. Pro lepší představu spočtěme Vapnik-Chervonenkisovu dimenzi pro pár nejjednodušších tříd množin.

Příklad 6. Nechť $\mathcal{A}$ obsahuje všechny polopřímky reálné osy $(-\infty, x], x \in \mathbb{R}$, potom $s(\mathcal{A}, 2) = 3 < 2^2$, a tedy $V_\mathcal{A} = 1$.

Příklad 7. Pokud $\mathcal{A}$ obsahuje konečný počet množin, pak $V_\mathcal{A} \leq \log_2 |\mathcal{A}|$ a $s(\mathcal{A}, n) \leq |\mathcal{A}|$ pro všechna přirozená n, protože k rotztřídění n bodů je potřeba minimálně 2^n množin. Druhé tvrzení je jasné.

Příklad 8. Nechť $\mathcal{A}$ je třída všech intervalů $(x, y) \subset \mathbb{R}$, pak $V_\mathcal{A} = 2$. Protože neexistuje interval, který by ze tří libovolných bodů vybral dva krajní a zároveň nevybral prostřední z nich.

Lze dokázat následující tvrzení (důkaz lze nalézt např. v [3]).

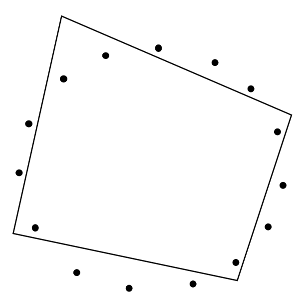
Tvrzení 13. *Buď $\mathcal{A}$ třída všech uzavřených koulí v $\mathbb{R}^d$, tj. podmnožin v $\mathbb{R}^d$ tvaru*

$$\left\{ x = (x^{(1)}, \dots, x^{(d)}) : \sum_{i=1}^{d} |x^{(i)} - a_i|^2 \leq b \right\}, \tag{5.3}$$

kde $a_1, \dots, a_d, b \in \mathbb{R}$ nabývají všech možných hodnot. Potom $V_\mathcal{A} \leq d + 2$.

Příklad 9. Pokud je $\mathcal{A}$ třída všech konvexních mnohoúhelníků v $\mathbb{R}^2$, potom $V_\mathcal{A} = \infty$, neboť každých n bodů tvoří buď mnohoúhelník, nebo leží na kružnici a pro libovolnou podmnožinu množiny bodů $z_1, \dots, z_n \in \mathbb{R}^2$ ležících na kružnici existuje mnohoúhelník, který ji vybere (viz Obrázek 5.1).

Obrázek 5.1: Libovolnou podmnožinu n prvků rozložených na kružnici lze vybrat pomocí mnohoúhelníku



5.2 Důkaz konzistence odhadů s minimální zobecněnou Kolmogorov–Smirnovovou vzdáleností

V této části vyslovíme a dokážeme větu o konzistenci odhadů s minimální zobecněnou Kolmogorov-Smirnovovou vzdáleností. Důkaz provedeme v několika krocích s pomocí tří technických lemmat, která také dokážeme[1].

[1]Celý důkaz je proveden podle velmi stručného návodu uvedeného v [3].

Nejprve vyslovíme větu poskytující odhad, který použijeme v důkazu konzistence odhadů s minimální vzdáleností řádu $n^{-1/2}$ ve střední hodnotě L_1-normy.

Věta 14 (Alexander)**.** *Pro $n\varepsilon^2 \geq 64$ platí následující odhad:*

$$P\Big(\sup_{A\in\mathcal{A}} |\nu_n(A) - Q(A)| > \varepsilon\Big) \leq 16(\sqrt{n}\varepsilon)^{4096V_{\mathcal{A}}} \exp\left(-2n\varepsilon^2\right), \tag{5.4}$$

kde ν_n je empirická distribuce, Q skutečná distribuce a $\mathcal{A}$ množinová $\sigma-$algebra pravděpodobnostního prostoru $(\mathcal{X}, \mathcal{A}, \mathcal{P})$.

Důkaz lze nalézt např. ve [3].

Definice 18. Buď $\mathcal{F}_\Theta = \{f_\theta : \theta \in \Theta\}$ parametrická třída hustot v $\mathbb{R}^d$ ($\Theta \subset \mathbb{R}^k$ parametrický prostor) a $X_1, \ldots, X_n$ stejně nezávisle rozdělená pozorování na $f_\theta \in \mathcal{F}_\Theta$. Definujme třídu množin

$$\mathcal{A} = \Big\{\{x \in \mathbb{R}^d : f_{\theta_1} > f_{\theta_2}\} \ \ \theta_1, \theta_2 \in \Theta\Big\} \tag{5.5}$$

a následně odhad parametru θ s minimální $D_{\mathcal{A}}$ vzdáleností vztahem

$$\widehat{\theta}_n = \arg\min_{\theta\in\Theta} D_{\mathcal{A}}(P_\theta, \nu_n), \tag{5.6}$$

(pokud nějaká taková $\mathbb{X}$ měřitelná statistika splňující $\widehat{\theta}_n = \widehat{\theta}_n(\mathbb{X})$ existuje a je funkcí $\mathbb{X}$) kde P_θ je distribuce odpovídající hustotě f_θ, ν_n je empirická míra založení na $(X_1, \ldots, X_n)$ a $D_{\mathcal{A}}$ je zobecněná Kolmogorov-Smirnovova vzdálenost definovaná jako

$$D_{\mathcal{A}}(P, Q) = \sup_{A\in\mathcal{A}} |P(A) - Q(A)|. \tag{5.7}$$

Věta 15. *Pokud $\mathcal{A}$ má konečnou VC dimenzi pak*

$$E\left(\int |f_{\widehat{\theta}_n}(x) - f_\theta(x)| \,\mathrm{d}x\right) = O(n^{-1/2})\,, \tag{5.8}$$

to znamená, že odhad $f_{\widehat{\theta}_n}$ je konzistentní řádu $(n^{-1/2})$ ve střední hodnotě L_1-normy. Kde všechny symboly jsou zavedeny v předchozí definici.

Pro jednoduchost zápisu označme $\widehat{\theta}_n = \widehat{\theta}$. Připravme tvrzení, která budeme v důkazu potřebovat.

Věta 16 (Scheffeho věta)**.** *Buďte P, Q absolutně spojité pravděpodobnostní míry na $\mathbb{R}^d$ s hustotami f, g. Pak platí:*

$$\sup_{A\in\mathcal{B}} |P(A) - Q(A)| = \frac{1}{2}\int |f(x) - g(x)|\,(d)x, \tag{5.9}$$

kde $\mathcal{B}$ je Borelovská $\sigma-$algebra na $\mathbb{R}^d$.

Důkaz. Pro naše účely nám stačí dokázat Scheffeho větu pouze v této podobě

$$\sup_{A\in\mathcal{A}} |P(A) - Q(A)| = \frac{1}{2}\int |f(x) - g(x)|\,(d)x, \tag{5.10}$$

kde $\mathcal{A}$ je třída množin viz definice 18. Definujme množinu $B = \{x \in \mathbb{R}^d : f(x) > g(x)\}$. Protože $\int (f(x) - g(x))\,\mathrm{d}x = 0$ platí

$$\begin{aligned}\int |f(x) - g(x)|\,\mathrm{d}x &= \int_B |f(x) - g(x)|\,\mathrm{d}x + \int_{B^C} |f(x) - g(x)|\,\mathrm{d}x \\ &= \int_B (f(x) - g(x))\,\mathrm{d}x - \int_{B^C} (f(x) - g(x))\,\mathrm{d}x \\ &= 2\int_B (f(x) - g(x))\,\mathrm{d}x = 2\Big|\int_B f(x)\,\mathrm{d}x - \int_B g(x)\,\mathrm{d}x\Big|.\end{aligned} \tag{5.11}$$

Pro libovolnou množinu $A \in \mathcal{A}$ pak pokračujeme následujícími odhady

$$\begin{aligned}\Big|\int_A f(x)\,\mathrm{d}x - \int_A g(x)\,\mathrm{d}x\Big| &= \Big|\int_{A\cap B} (f(x) - g(x))\,\mathrm{d}x + \int_{A\cap B^C} (f(x) - g(x))\,\mathrm{d}x\Big| \\ &\leq \max\Big(\int_{A\cap B} (f(x) - g(x))\,\mathrm{d}x, \int_{A\cap B^C} (g(x) - f(x))\,\mathrm{d}x\Big) \\ &\leq \max\Big(\int_B (f(x) - g(x))\,\mathrm{d}x, \int_{B^C} (g(x) - f(x))\,\mathrm{d}x\Big) \\ &= \frac{1}{2}\int |f(x) - g(x)|\,\mathrm{d}x\,.\end{aligned} \tag{5.12}$$

Tedy pro supremum platí

$$\sup_{A\in\mathcal{A}}\Big|\int_A f(x)\,\mathrm{d}x - \int_A g(x)\,\mathrm{d}x\Big| = \frac{1}{2}\int |f(x) - g(x)|\,\mathrm{d}x\,, \tag{5.13}$$

přičemž suprema se nabývá na množině $B = \{x \in \mathbb{R}^d : f(x) > g(x)\}$. □

Lemma 17. *Buď ψ reálná, dvakrát diferencovatelná, záporná, klesající konkávní funkce, potom platí*

$$\int_u^\infty \exp(\psi(t))\,\mathrm{d}t \leq \frac{\exp(\psi(u))}{-\psi'(u)}\,. \tag{5.14}$$

Důkaz. Použijeme Taylorův rozvoj funkce ψ se středem v bodě u

$$\begin{aligned}\psi(x) &= \psi(u) + \psi'(u)(x-u) + \frac{1}{2}\psi''(\xi)(x-u)^2\,, \quad \xi \in (u,x) \vee \xi \in (x,u) \\ \psi(x) &\leq \psi(u) + \psi'(u)(x-u)\,,\end{aligned} \tag{5.15}$$

neboť ψ je konkávní, a dále s využitím předpokladů platí

$$\begin{aligned}\int_u^\infty \exp(\psi(x))\,\mathrm{d}x &\leq \int_u^\infty \exp(\psi(u) + \psi'(u)(x-u))\,\mathrm{d}x = \left\{\begin{array}{c} t = \psi'(u)(x-u) \\ \mathrm{d}t = \psi'(u)\mathrm{d}x \end{array}\right\} \\ &= -\frac{\exp(\psi(u))}{\psi'(u)}\int_{-\infty}^0 \exp(t)\,\mathrm{d}t = \frac{\exp(\psi(u))}{-\psi'(u)}\,.\end{aligned} \tag{5.16}$$

□

Lemma 18. *Buď $b > 0$ pevné, potom pro $u > \sqrt{b/2}$ platí*

$$\int\limits_u^\infty t^b \exp\left(-2t^2\right) \mathrm{d}t \le \frac{u^{b-1}\exp\left(-2u^2\right)}{2}\,. \tag{5.17}$$

Důkaz. Pro odhad integrálu použijeme lemma 17, ověřme jeho předpoklady.

$$\begin{array}{rcl} \psi(t) & = & b\ln t - 2t^2 \\ \psi'(t) & = & \frac{b}{t} - 4t \\ \psi'(t) & \le & 0\,, \quad \forall\, t \ge \sqrt{b/4} \\ \psi''(t) & \le & 0\,, \quad \forall\, t\,. \end{array} \tag{5.18}$$

Zápornost funkce na celém intervalu není nutná, viz předchozí důkaz, je jen třeba, aby uvažovaný integrál konvergoval. Tedy podle lemmatu 17 platí

$$\begin{aligned} \int\limits_u^\infty t^b \exp\left(-2t^2\right) \mathrm{d}t = \int\limits_u^\infty \exp\left(b\ln t - 2t^2\right) \mathrm{d}t &\le \frac{u^b\exp\left(-2u^2\right)}{4u - \frac{b}{u}} \\ &\le \frac{u^b\exp\left(-2u^2\right)}{2u} = \frac{u^{b-1}\exp\left(-2u^2\right)}{2}\,, \end{aligned} \tag{5.19}$$

kde poslední odhad platí pro $u > \sqrt{b/2}$. □

Lemma 19. *Buď X kladná náhodná veličina, pro kterou platí*

$$P\{X > u\} \le au^b \exp\left(-2u^2\right) \;\; \textit{pro} \quad u \ge \sqrt{c}\,, \tag{5.20}$$

kde a,b,c jsou kladné konstanty. Pokud $b \ge 2c \ge e$, potom platí

$$EX \le \frac{a}{2} + \sqrt{\frac{b\ln b}{2}}\,. \tag{5.21}$$

Důkaz. Protože pro kladnou náhodnou veličinu platí $EX = \int\limits_o^\infty P\{x > t\}\,\mathrm{d}t$ a dále platí tyto nerovnosti $\sqrt{\frac{b\ln b}{2}} > \sqrt{\frac{b}{2}} > \sqrt{c}$, můžeme provést následující odhady

$$\begin{aligned} EX = \int\limits_0^\infty P\{X > t\}\,\mathrm{d}t &= \int\limits_0^{\sqrt{\frac{b\ln b}{2}}} P\{X > t\}\,\mathrm{d}t + \int\limits_{\sqrt{\frac{b\ln b}{2}}}^\infty P\{X > t\}\,\mathrm{d}t \\ &\le \int\limits_0^{\sqrt{\frac{b\ln b}{2}}} 1\,\mathrm{d}t + \int\limits_{\sqrt{\frac{b\ln b}{2}}}^\infty at^b\exp\left(-2t^2\right)\mathrm{d}t \le \sqrt{\frac{b\ln b}{2}} + \frac{a}{2}\left(e^{-b}\sqrt{\frac{(b\ln b)^{b-1}}{2}}\right) \\ &\le \sqrt{\frac{b\ln b}{2}} + \frac{a}{2}\,, \end{aligned} \tag{5.22}$$

kde jsme použili lemma 18 a v poslední nerovnosti odhadli výraz v závorce jedničkou. □

A nyní již máme připravená všechna tvrzení potřebná k důkazu věty 15.

Důkaz (věty 15). Pomocí Scheffeho věty, trojúhelníkové nerovnosti a vlastností zobecněné Kolmogorov-Smirnovovy vzdálenosti odhadujeme

$$
\begin{aligned}
\int |f_{\widehat{\theta}}(x) - f_\theta(x)| \,\mathrm{d}x = 2 \sup_{A\in\mathcal{A}} |P_{\widehat{\theta}}(A) - P_\theta(A)| &= 2D_{\mathcal{A}}(P_{\widehat{\theta}}, P_\theta) \\
&\le 2(D_{\mathcal{A}}(P_{\widehat{\theta}}, \nu_n) + D_{\mathcal{A}}(\nu_n, P_\theta)) \\
&\le 4D_{\mathcal{A}}(\nu_n, P_\theta)\,.
\end{aligned}
\tag{5.23}
$$

Dále odhadujeme střední hodnotu

$$
\begin{aligned}
\sqrt{n}E\Big\{\int |f_{\widehat{\theta}}(x) - f_\theta(x)| \,\mathrm{d}x\Big\} &\le E\{4\sqrt{n}D_{\mathcal{A}}(\nu_n, P_\theta)\} \\
&= 4E\{\sqrt{n}\sup_{A\in\mathcal{A}} |\nu_n(A) - P_\theta(A)|\} \\
&\le 4\Big(8 + \sqrt{2048V_{\mathcal{A}}\ln(4096V_{\mathcal{A}})}\Big)\,,
\end{aligned}
\tag{5.24}
$$

kde poslední odhad získáme použitím lemmatu 19. Ověřme jeho předpoklady.

$$
\begin{aligned}
P\Big\{\sqrt{n}\sup_{A\in\mathcal{A}} |\nu_n(A) - P_\theta(A)| > u\Big\} &= P\Big\{\sup_{A\in\mathcal{A}} |\nu_n(A) - P_\theta(A)| > \frac{u}{\sqrt{n}}\Big\} \\
&\le 16u^{4096V_{\mathcal{A}}}\exp(-2u^2)\,, \quad \text{pro } u \ge 8\,,
\end{aligned}
\tag{5.25}
$$

kde odhad je dán větou 14 (Alexander). A nerovnost mezi konstantami b, c z lemmatu 19 je také splněn neboť $b/2 = 2048V_{\mathcal{A}} \ge 64 = c \ge e$.
Tím je důkaz dokončen. □

Pomocí přístupu přes Vapnik-Chervonenkisovu dimenzi se nám také podařilo dokázat konzistenci odhadů s minimální vzdáleností s rychlostí řádu $n^{-1/2}$ ve střední hodnotě L_1-normy.

5.3 Vapnik-Chervonenkisova dimenze a stupeň variace

Na příkladech porovnejme oba zatím prozkoumané přístupy a podívejme se na vzájemný vztah jim odpovídajících charakteristik: stupně variace, parciálního stupně variace a Vapnik-Chervonenkisovy dimenze. Nejprve poznamenejme, že Vapnik-Chervonenkisova dimenze je citlivá na změny hustot na množinách nulové míry, zatímco stupně variace ne. Proto v této části uvažujme rodiny hustot $\mathcal{D}$ obsahující hustoty různící se na množinách nenulové míry.

Na jednoduchém příkladě snadno ověříme, že v obecném případě z konečnosti Vapnik-Chervonenkisovy dimenze neplyne konečnost stupně variace.

Příklad 10. Uvažujme konečnou rodinu hustot $\mathcal{D} = \{f_1, \ldots, f_p\}$ na $\mathbb{R}$ s nekonečným stupněm variace. Spočtěme VC dimenzi třídy množin

$$
\mathcal{A} = \Big\{\{x \in \mathbb{R} : f_i > f_j\} \;\; i, j \in \widehat{p}\Big\}\,. \tag{5.26}
$$

Z Příkladu 7 víme, že $V_{\mathcal{A}} \le p < \infty$.

Příklad lze velmi snadno rozšířit i na nekonečnou rodinu hustot, jak ukáže následující příklad 11.

Příklad 11. Nechť rodina hustot $\mathcal{D}$ obsahuje rovnoměrnou hustotu g intervalu $(0,1)$ a hustoty f_n definované takto

$$f_n = \begin{cases} \frac{n}{n+1} & x \in (\frac{1}{2^{2i}}, \frac{1}{2^{2i-1}}),\ i = 1,2,\dots \\ \frac{n+3}{n+1} & x \in (\frac{1}{2^{2i+1}}, \frac{1}{2^{2i}}),\ i = 1,2,\dots \\ 1 & x \in (\frac{1}{2}, 1) \end{cases} \tag{5.27}$$

Potom $DV(\mathcal{D}) = \infty$, protože například $DV(g, f_1) = \infty$, zatímco $V_{\mathcal{A}} < \infty$, protože třída množin $\mathcal{A}$ obsahuje pouze dvě různé množiny

$$\begin{aligned} A_1 &= \{x \in \mathbb{R} : g(x) > f_j(x)\} &&= \bigcup_{i=1}^{\infty} \left(\tfrac{1}{2^{2i}}, \tfrac{1}{2^{2i-1}}\right),\ \forall j \in \mathbb{N} \\ A_2 &= \{x \in \mathbb{R} : f_j(x) > g(x)\} &&= \bigcup_{i=1}^{\infty} \left(\tfrac{1}{2^{2i+1}}, \tfrac{1}{2^{2i}}\right),\ \forall j \in \mathbb{N} \\ A_3 &= \{x \in \mathbb{R} : f_j(x) > f_i(x)\} &&= A_1,\ \text{pro } j > i \\ A_4 &= \{x \in \mathbb{R} : f_j(x) > f_i(x)\} &&= A_2,\ \text{pro } i > j \end{aligned} \tag{5.28}$$

Tento příklad ukazuje situaci, kdy je možné použít Vapnik-Chervonenkisovu teorii, ale teorii založenou na stupni variace ne. Podívejme se ještě, je-li možné použít teorii parciálního stupně variace vybudovanou v kapitole 4. Snadno zjistíme, že existuje posloupnost a_n, $\lim\limits_{n\to\infty} a_n = 0$ taková, že $DV_{a_n}(f_n, g) < K < \infty$, $\forall n \in \mathbb{N}$, protože

$$\lim_{n\to\infty} \nu_n^+\left(\bigcup_{i=1}^{\infty}\left(\tfrac{1}{2^{2i}}, \tfrac{1}{2^{2i-1}}\right)\right) = 0 \ \text{ a také } \lim_{n\to\infty} \nu_n^-\left(\bigcup_{i=1}^{\infty}\left(\tfrac{1}{2^{2i}}, \tfrac{1}{2^{2i-1}}\right)\right) = 0,$$

kde ν_n^+ a ν_n^- jsou horní a dolní variace definované hustotami g, f_n, a tedy je možné tuto teorii použít.

V následujících dvou příkladech zkonstruujeme rodinu hustot, na které bude možné použít obě teorie (tj. $DV(\mathcal{D}) < \infty \wedge V_{\mathcal{A}} < \infty$), a rodinu hustot, na které obě teorie selžou (tj. $DV(\mathcal{D}) = \infty \wedge V_{\mathcal{A}} = \infty$).

Příklad 12. Uvažujme rodinu $\mathcal{D}$ všech hustot Gaussova normálního rozdělení ve tvaru $f(\sigma, \mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$. Potom zřejmě $DV(\mathcal{D}) = 1$ a $V_{\mathcal{A}} = 3$, protože třída množin

$$\mathcal{A} = \{x \in \mathbb{R} : f(\mu_1, \sigma_1) > f(\mu_2, \sigma_2),\ \mu_1, \mu_2 \in \mathbb{R},\ \sigma_1, \sigma_2 > 0\} \tag{5.29}$$

obsahuje pouze množiny typu

$$A_1 = (x,y), A_2 = (-\infty, x) \cup (y, \infty), A_3 = (-\infty, x), A_4 = (y, \infty),\ x, y \in \mathbb{R}\,, \tag{5.30}$$

pomocí kterých jdou roztřídit maximálně tři body $z_1, z_2, z_3 \in \mathbb{R}$.

Příklad 13. Zvolme $k \in \mathbb{N}$ libovolně a vyberme k různých bodů $z_1, \dots, z_k$ v intervalu $(2k, 2k+1)$ a označme $z_{(1)}, \dots, z_{(k)}$ jejich vzestupné přerovnání. Pro $j = 1, \dots, k$ definujme intervaly

$$\begin{aligned} U_j = (u_{j-1}, u_j) \subset (2k, 2k+1)\,,\ \text{kde } u_j &= \frac{z_{(j+1)} + z_{(j)}}{2} \qquad \text{pro } j = 1, \dots, k-1 \\ u_0 &= 2k,\ u_k = 2k+1\,. \end{aligned} \tag{5.31}$$

Existuje 2^k různých podmnožin množiny $\{1,\ldots,k\}$, označme je M_i^k, $i = 1,\ldots,2^k$ a definujme hustoty

$$f_{M_i^k} = \begin{cases} \lambda\Big(\bigcup\limits_{j\in M_i^k} U_j\Big)^{-1} & \text{na } \bigcup\limits_{j\in M_i^k} U_j \\ 0 & \text{jinde na } \mathbb{R}\,, \end{cases} \tag{5.32}$$

kde $\lambda(A)$ je Lebesqueova míra množiny A.

Definujme rodinu hustot $\mathcal{D} = \{f_{M_i^k} : i = 1,\ldots 2^k, k \in \mathbb{N}\}$.
Z konstrukce je zřejmé, že pro každé $k \in \mathbb{N}$ existuje množina $\{z_1,\ldots,z_k\}$, která je roztříděna třídou množin $\mathcal{A} = \{x \in \mathbb{R} : f_{M_i^k} > f_{M_j^k}, i,j \in \{1,\ldots 2^k\}, k \in \mathbb{N}\}$, a tedy $V_{\mathcal{A}} = \infty$. Ovšem v tomto případě také $DV(\mathcal{D}) = \infty$.

Následující příklad ukáže, že obecně neplatí ani obrácená implikace, tedy že z konečnosti stupně variace neplyne konečnost Vapnik-Chervonenkisovy dimenze. Jeho konstrukce bude podobná jako u příkladu 13, jen značně opatrnější, aby stupeň variace zůstal konečný.

Příklad 14. [2] Zvolme $k \in \mathbb{N}$ libovolně a vyberme k různých bodů $z_1,\ldots,z_k$ v intervalu $(2k-1, 2k)$ a označme $z_{(1)},\ldots,z_{(k)}$ jejich vzestupné přerovnání a $Z_k = \{z_1,\ldots,z_k\}$. Pro $j = 1,\ldots,k$ definujme intervaly

$$\begin{aligned} U_i = (u_{i-1}, u_i) \subset (2k-1, 2k)\,,\ \text{kde } u_i &= \frac{z_{(j+1)} + z_{(j)}}{2} \text{ pro } i = 1,\ldots,k-1 \\ u_0 &= 2k-1,\ u_k = 2k\,. \end{aligned} \tag{5.33}$$

Existuje 2^k různých podmnožin $S_j \subset Z_k\, j = 1,\ldots,2^k$. Pro $j = 1,\ldots,2^k$ a $i = 1,\ldots,k$ definujme hustoty

$$g_j^k(x) = \begin{cases} 1 - \frac{1}{2^j} & \text{pro } x \in U_i,\ \text{když } U_i \cap S_j = \emptyset \\ 1 - \frac{1}{2^{j+1}} & \text{pro } x \in U_i,\ \text{když } U_i \cap S_j \neq \emptyset \\ 1 - \sum_{i=1}^k \alpha_j^i d_i & \text{pro } x \in (2k-2,\ldots 2k-1)\,, \end{cases} \tag{5.34}$$

kde $d_i = |u_i - u_{i-1}|$ je délka intervalu U_i a

$$\alpha_j^i = \begin{cases} 1 - \frac{1}{2^j} & \text{když } U_i \cap S_j = \emptyset \\ 1 - \frac{1}{2^{j+1}} & \text{když } U_i \cap S_j \neq \emptyset\,. \end{cases} \tag{5.35}$$

Definujme rodinu hustot

$$\mathcal{D} = \{g_j^k : j = 1,\ldots,2^k, k \in \mathbb{N}\} \cup \{f_k, k \in \mathbb{N}\}, \tag{5.36}$$

kde f_k je rovnoměrná hustota na intervalu $(2k-1, 2k)$. Vidíme, že $DV(\mathcal{D}) = 1$, zatímco VC dimenze třídy množin $\mathcal{A} = \{x \in \mathbb{R} : f(x) > g(x), f, g \in \mathcal{D}\}$ je $V_{\mathcal{A}} = \infty$, protože z konstrukce je jasné, že pro každé $k \in \mathbb{N}$ existuje množina $Z_k = \{z_1,\ldots,z_k\}$, která je roztříděna třídou množin $\mathcal{A}$.

Je tedy vidět, že podmínka $DV(\mathcal{D}) < \infty$ pro Kolmogorovské odhady je přímo neporovnatelná s podmínkou $V_{\mathcal{A}} < \infty$ pro odhady s minimální GKS vzdáleností bez omezujících požadavků na rodiny $\mathcal{D}$. Pro konstrukci podmínek, za kterých bude konečnost stupně variace implikovat konečnost VC dimenze se inspirujme příkladem 14. Tedy je třeba zabránit situaci kdy $f_1 - f_2$ na nějakém intervalu (a,b), $-\infty \le a < b \le +\infty$ nemění znaménko a zároveň $\{x : f_1 > f_2\}$ je sjednocení libovolného počtu intervalů.

[2] Příklad je převzat z [11].

Věta 20. *Nechť pro každé $f_i, f_j \in \mathcal{D}$ platí, že pokud existuje interval $(a,b) \subset \mathbb{R}$ takový, že pro $\forall x \in (a,b)$ $f_i(x) = f_j(x)$ a existuje $c < a$ pro které $f_i(c) \neq f_j(c)$, potom $f_i(x) = f_j(x)$ pro $\forall x > a$. Pak konečnost stupně variace implikuje konečnost VC dimenze.*

Poznámka 6. Existence $c < a$ pro které $f_i(c) \neq f_j(c)$ v předpokladech věty 20 je proto, aby v rodině $\mathcal{D}$ mohly být i hustoty, které se na intervalu $(-\infty, d)$, $d \in \mathbb{R}$ rovnají a teprve v bodě d se jejich hodnoty rozcházejí.

Důkaz. Nechť $DV(\mathcal{D}) < \infty$, potom počet znaménkových změn rozdílu libovolných dvou hustot z $\mathcal{D}$ je nanejvýš roven $2DV(\mathcal{D})$ a díky předpokladu množiny tvořící třídu $\mathcal{A}$ jsou sjednocení maximálně $DV(\mathcal{D}) + 1$ intervalů. To znamená, že třída množin $\mathcal{A}$ neroztřídí $2DV(\mathcal{D}) + 3$ bodů, protože nedokáže vybrat každý druhý. Tedy $V_{\mathcal{A}} < 2DV(\mathcal{D}) + 3$. □

Poznámka 7. Implikaci nelze obrátit. Pokud je $V_{\mathcal{A}} < \infty$ pak existuje $l \in \mathbb{R}$, že třída množin $\mathcal{A}$ neroztřídí žádnou $l-$tici bodů $\{z_1, \ldots, z_l\}$. Důvody k tomu mohou být dva. Buď $|\mathcal{A}| < 2^l$ nebo $|\mathcal{A}| \geq 2^l$, ale množiny jsou nešikovné. Ovšem ani v jednom případě není zaručeno, že ve třídě $\mathcal{A}$ není množina typu $\bigcup_{i=1}^{\infty}(a_i, b_i)$, která při platnosti předpokladu věty 20 už nutně musí pocházet od dvou hustot s nekonečným stupněm variace.

Na závěr poznamenejme, že zobecněný Kolmogorovský odhad je výpočetně značně náročnější než Kolmogorovský odhad, neboť minimalizaci provádí přes mnohem větší třídu množin. Současně ověření podmínky $DV(\mathcal{D}) < \infty$ je snažší, než ověřování podmínky $V_{\mathcal{A}} < \infty$.

Kapitola 6

Yatracosovo kritérium

V této kapitole krátce shrneme poznatky článku [16] o konzistenci odhadů s minimální vzdáleností. Na příkladech porovnáme výsledky dosažené pomocí této teorie s výsledky teorie založené na stupni variace.

6.1 Konstrukce odhadu

Na měřitelný prostor $(\mathcal{X}, \mathcal{A})$ nejsou kladena žádná omezení, nechť množina pravděpodobnostních měr $\mathcal{P}(\mathcal{X}, \mathcal{A})$ je dána parametricky $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$. O struktuře parametrického prostoru Θ se nepředpokládá nic, ani jestli je konečněrozměrný, nebo ne. Metrizujme $\mathcal{P}$ totální variací d_{TV} mezi mírami, která je vlastně L_1-normou mezi odpovídajícími hustotami, pokud existují. (Což nás opravňuje říkat že $\mathcal{P}$ je metrizována L_1-normou).

Definice 19. Buď (Y, d) totálně omezený metrický prostor. Pro každé a označme $N(a)$ nejmenší počet $d-$koulí o poloměru a (tj. množin tvaru $\{x \in Y : d(x, x_0) < a\}$, kde $x_0 \in Y$ je střed), které pokrývají Y. Funkci $\log_2 N(a)$ nazýváme Kolmogorovská entropie prostoru Y.

Lemma 21. *Buď $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ L_1-totálně omezená monožia pravděpodobnostních měr na $(\mathcal{X}, \mathcal{A})$. Pak pro každé $a > 0$ existuje třída množin $F_a \subseteq \mathcal{A}$ mohutnosti $|F_a| \leq N^2(a)$ tak, že $d_{TV}(P_{\theta_1}, P_{\theta_2}) \leq 4a + \sup\{|P_{\theta_1}(A) - P_{\theta_2}(A)|, A \in F_a\}$*

Věta 22. *Pokud je $\mathcal{P}$ L_1-totálně omezená, pak existuje stejnoměrně konzistentní odhad $\widehat{\theta}_n$ parametru θ jehož řád konvergence splňuje rovnici $a_n = [\log(N(a))/n]^{1/2}$.*

6.2 Yatracosův přístup a stupeň variace

Předpokládejme, že k mírám z $\mathcal{P}$ existují hustoty (množinu odpovídajících hustot označme $\mathcal{D}$), potom odhad $\widehat{f}_n$ hustoty f získaný podle věty 22 je přímo jeden ze středů L_1 koulí pokrývajících $\mathcal{D}$. Všimněme si, že $N(a) \to \infty$ když $n \to \infty$ pro posloupnost a_n klesající k nule. A tedy odhad $\widehat{f}_n$ nemusí být konzistentní řádu $n^{-1/2}$.

Některé z následujících příkladů jsou převzaty z [16].

Příklad 15. Uvažujme rodinu hustot tvaru:

$$\mathcal{D} = \left\{ f : [0,1] \to \mathbb{R}^+ : |f^{(s)}x - f^{(s)}(y)| \leq L|x - y| \text{ na } [0,1] \right\} \tag{6.1}$$

Yatracos ukázal že odpovídajíjí odhad s minimální vzdáleností daný větou 22 je konzistentní řádu $a_n = n^{-s+1/2s+3}$ v L_1-normě. Tedy můžeme získat odhad s řádem konzistence libovolně blízko $n^{-1/2}$, když s bude vybráno libovolně velké. Stejný výsledek byl získán již dříve pro Lipschitzovskou třídu hustot viz [2]. V tomto případě nemusí být stupeň variace ani parciální stupeň variace konečný.

Příklad 16. Uvažujme parametrickou rodinu hustot:

$$\mathcal{D} = \left\{ f_\theta(x) = \frac{1}{2} \exp\left\{-|x-\theta|\right\} :\in \mathbb{R}, \theta \in [0,1] \right\} \tag{6.2}$$

$\mathcal{D}$ je L_1-totálně omezená a platí, že $N(a_n) \sim 1/a_n$ pro a_n malé. Pak z věty 22 obdržíme odhad konzistentní řádu $a_n = [(\log n)/n]^{1/2}$ v L_1-normě. Zatímco použitím vět 10 a 11 získáme silnější výsledek. Neboť $DV(\mathcal{D}) = 1$ a tedy Kolmogorovské odhady jsou konzistentní řádu $n^{-1/2}$ v L_1-normě i její střední hodnotě.

Příklad 17. Definujme rodinu hustot

$$\begin{aligned} &\mathcal{D} = \left\{ f(x) = \frac{1}{K} \exp\left\{\frac{1}{1-(x-s)^2}\right\} : x \in [s-1, s+1], s \in \mathbb{R} \right\} \\ &\text{kde } K = \int_{-1}^{1} \exp\left\{\frac{1}{1-x^2}\right\}. \end{aligned} \tag{6.3}$$

Vidíme, že rodina $\mathcal{D}$ nemůže být L_1 totálně omezená, protože obsahuje podmnožinu $\mathcal{D}^* = \{f(x) : x \in [s-1, s+1], s \in \mathbb{Z}\}$, která není L_1 totálně omezená díky posunům nosičů, zatímco $DV(\mathcal{D}) = 1$. Máme tedy příklad rodiny hustot na které nejsou splněny předpoklady Yatracosova kritéria (věta 22), ale teorie stupně variace je použitelná.

Snadno získáme i příklad rodiny hustot s nekonečným stupněm variace, ale s konečným parciálním stupněm, na které nebudou splněny předpoklady Yatracosova kritéria (věta 22). Stačí předchozí příklad (17) modifikovat.

Příklad 18. Nechť rodinu hustot $\mathcal{D}$ obsahuje hustoty

$$f_n^s = \begin{cases} \frac{1}{K} \exp\left\{\frac{1}{1-(x-s)^2}\right\} & x \in [s-1, c] \cup [d, s+1],\ \ s \in \mathbb{R} \\ \frac{1}{K} \exp\left\{\frac{1}{1-(x-s)^2}\right\} + h_n \sin\left(\frac{2\pi n}{d-c}(x-c)\right) & x \in [c, d],\ \ s \in \mathbb{R} \end{cases} \tag{6.4}$$

kde $K = \int_{-1}^{1} \exp\left\{\frac{1}{1-x^2}\right\}$ a pro konstanty c, d platí, že $s-1 \leq c \leq d \leq s+1$ a posloupnost h_n je taková nezáporná posloupnost s nulovou limitou, že $h_n < \min\{f(c), f(d)\}$. Potom například pro posloupnost $a_n = 2h_n \frac{d-c}{\pi}\left(1 - \frac{1}{4n}\right)$ je na rodině $\mathcal{D}$ splněna asymptotická dominance řádu a_n, ale ze stejných důvodů jako v příkladu 17 nenjsou splněny předpoklady pro použití Yatracosovy teorie.

Z příkladů je vidět, že nenajdeme přímé propojení Yatracosovy teorie a teorie stupně variace. Neboť existují rodiny hustot (viz výše) sloužící jako protipříklady.

Dodatky

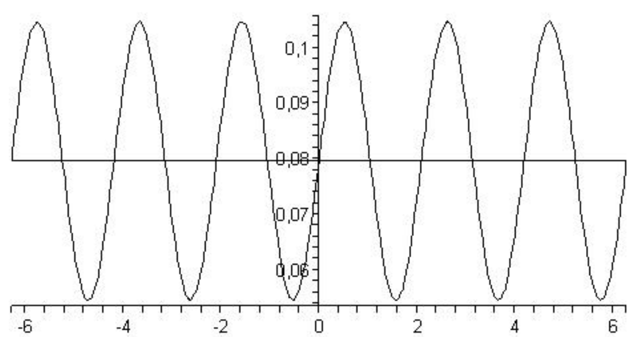
Stupeň variace a počet znaménkových změn

V článku [11] se objevuje tvrzení:

Pro nenulový konečný stupeň variace $DV(f,g)$ můžeme vhodnou volbou hustot f a g na množinách λ nulové míry redukovat počet znaménkových změn rozdílu $f-g$ na $DV(f,g)+2$, když všechny intervaly $\mathcal{J}_j$ v definici 9 jsou omezené, nebo na $DV(f,g)+1$, pokud jeden z intervalů $\mathcal{J}_j$ je neomezený.

Toto tvrzení však neplatí. Demonstrujme to jednoduchým příkladem. Uvažujme f rovnoměrnou hustotu na intervalu $(-2\pi, 2\pi)$ a hustotu g na stejném nosiči definovanou následně $g(x) = \frac{1}{4\pi} + \frac{1}{40}\sin(3(x+2\pi))$.

Obrázek 1:



Z obrázku 6.2.1 je vidět, že počet znaménkových změn rozdílu $f-g$ je 11 a $DV(f,g) = 6$, což je v rozporu s výše uvedeným původním tvrzením.

Literatura

[1] Demut, R., *Robustnost odhadů hustot s minimální divergencí Diplomová práce* FJFI ČVUT Praha

[2] Devroye, L., Györfi, L., *Nonparametric Densiti Estimation: The L_1-View*, Wiley, New York 1985

[3] Devroye, L., Györfi, L., Lugosi, G., *A Probabilistic Theory of Pattern Recognition*, Springer New York 1996

[4] Friedrich, L., Vajda, I., *On Divergences and Informations in Statistics and Information Theory*, IEEE Transactions on Information Theory, Vol. 52, No. 10, 2006, s.4394-4412

[5] Frýdlova, I., *Odhady pravděpodobnostních hustot s minimální Kolmogorovskou vzdáleností Diplomová práce* FJFI ČVUT Praha, 2004

[6] Gibbs,A. L., Su F.,E., *On choosing and bounding probability metrics*, International Statistical Review, 70, 2002, 419–435.

[7] Györfi, L., Vajda I., van der Meulen, E. C.,*Family of point estmates yielded by L_1-consistent density estimate* L1-Statistical Analysis and Related Methods, 1992, s. 415-430

[8] Györfi, L., Vajda I., van der Meulen, E. C., *Minimum Kolmogorov Distance Estimates of Parameters and Parametrized Distributions*, Metrika, 43, 1996, s.237-255

[9] Izenman, A. J., *Recent Developments in Nonparametric Density Estimation*, Journal of the American Statistical Association, 86, 1991

[10] Kirmani, S. N. U. A., *On the Relation between Matusita's and Kolmogorov's measures of distance*, Ann. Inst. Statist. Math., 31, 1979, s. 289-291

[11] Kůs, V., *Nonparametric density estimates consistent of the order of $n^{-1/2}$ in the L_1-norm*, Metrika, 51, 2004

[12] Kůs, V., Morales, D., Frýdlová, I., *Minimum disparity estimator–existence, consistency and computer simulations*, zasláno do Metrika 2009

[13] Matusita, K., *On the notion of affinity of several distributions and some of its applications*, Ann. Inst. Statist. Math., 19, 1967, s. 181-192

[14] Pardo, L., *Statistical Interference Based on Divergence Measures*, Taylor and Francis Group, LLC, Boca Raton - London - New York 2006

[15] Vajda, I., *Information - Theoretic Methods in Statistics*, Research report No 1834 March 1995, UTIA AV ČR, Praha 1995

[16] Yatracos, Y. G.,*Rates of Convergence of Minimum Distance Estimators and Kolmogorov's Entropy* Annals of Statistics 13, s. 768-774