EMPIRICKÉ PROTĚJŠKY KLASICKÝCH OPTIMÁLNÍCH ODHADŮ

Igor Vajda, ÚTIA ČSAV

1. Úvod a přehled výsledků

V práci uvažujeme určité míry divergence pravděpodobnostních distribucí. Ukážeme, že pomocí těchto měr lze zobecnit pojem ztrátové funkce v obecném statistickém problému odhadu. Ukážeme dále, že klasické optimální (stejnoměrně nejlepší, minimaxní, resp. Bayesovy) odhady minimalizují divergenci mezi odhadnutým a neznámým teoretickým rozdělením. Empirické protějšky těchto odhadů minimalizují divergenci mezi odhadnutým a známým empirickým rozdělením. Empirické odhady se tedy liší od klasických pouze v tom, že ve ztrátové funkci (divergenci), která figuruje v příslušné definici, nahradíme neznámé teoretické rozdělení známým empirickým. Každý klasický odhad (pokud existuje) má tedy svůj empirický protějšek (pokud existuje) a naopak.

Zmíněná záměna rozdělení v definici má závažné důsledky pokud jde o existenci, výpočet a robustnost jedněch a druhých odhadů. Napřed několik slov k existenci. Jak známo z literatury (např. Zacks [5], Wald [6]), k existenci klasických odhadů nestačí běžné podmínky regularity (kompaktnost či kompaktifikovatelnost parametrického prostoru a spojitost divergence a teoretické rodiny rozdělení). Stejnoměrně absolutně nejlepší odhad existuje jen pro singulární teoretická rozdělení. Pro netriviální teoretické rodiny mohou existovat jen stejnoměrně relativně nejlepší, například stejnoměrně nejlepší nevychýlené odhady. Nechme stranou minimaxní odhad, který vyžaduje vedle existence Bayesova odhadu i existenci nejméně příznivého apriorního rozdělení. Když se omezíme na euklidovské parametrické prostory a stejnoměrně nejlepší nevychýlené a Bayesovy odhady, pak se k existenci vyžaduje, aby ztrátová funkce (divergence) byla kvadratická nebo alespoň konvexní a aby teoretická rodina splňovala určité v praxi nepříliš snadno ověřitelné podmínky jako existenci úplné postačující statistiky resp. určité vlastnosti rizikové funkce (viz kap. 3 a 6 v [6]). Naproti tomu ukážeme, že empirické odhady za zmíněných běžných podmínek regularity vždy existují.

Nyní několik slov k výpočtům odhadů. Výpočet předepsané hodnoty empirického odhadu pro daný výběr je úloha matematického (dosti často konvexního) programování. Toto obecně neplatí pro klasické odhady. Rozhodnout bez znalosti předepsané hodnoty o tom, zda jedna hodnota parametru je lepší aproximací nežli hodnota jiná je v případě stejnoměrně nejlepšího nevychýleného odhadu úloha, jejíž algoritmizace je mimo rámec veškerých představ. V případě Bayesova odhadu jde o úlohu algoritmizovatelnou, avšak ukážeme, že složitost algoritmu je mnohonásobně vyšší, nežli složitost podobného algoritmu v případě empirického odhadu. Tím se vysvětluje, proč není možné resp. vhodné počítat klasické odhady iterativně, postupným přibližováním k předepsaným hodnotám. To že empirické odhady lze počítat iterativně je patrně hlavní příčinou jejich mnohem širšího praktického uplatnění.

Další příčinou je robustnost, která sice není zcela univerzální, ale nicméně je dosti typickou vlastností empirických odhadů. Ukážeme, že přímo z definice lze tuto skutečnost celkem jednoduše nahlédnout. Lze také odvodit číselnou míru citlivosti divergence na malé modifikace teoretických rozdělení, která je současně mírou robustnosti příslušného empirického odhadu. Uvidíme, že pro známé robustní odhady je tato citlivost konečná, zatím co pro maximálně věrohodný odhad je tato citlivost za obvyklých podmínek nekonečná.

Do empirických odhadů vedle maximálně věrohodného odhadu patří některé M- a L-odhady a D-odhady uvažované v dřívějších pracích [2,3,4], s poněkud modifikovaným pravidlem výběru předepsané hodnoty ve víceznačných případech.

2. Základní označení a definice

V práci $\mathcal{N}$ označuje množinu přirozených čísel, $\mathcal{R}$ reálnou přímku a $(\mathcal{X},\mathcal{A})$ výběrový měřitelný prostor, přičemž se předpokládá, že $\mathcal{X}$ je Hausdorffův topologický prostor s topologií $\mathcal{A}_0$, že $\mathcal{A}$ je σ-algebra borelovských podmnožin $\mathcal{X}$ generovaná topologií $\mathcal{A}_0$ a že $\{x\}\in\mathcal{A}$ pro všechna $x\in\mathcal{X}$. Dále $\mathcal{P}$ označuje rodinu všech rozdělení na $(\mathcal{X},\mathcal{A})$ a $1_{\{x\}}$ element $\mathcal{P}$ s veškerou pravděpodobností v bodě $x\in\mathcal{X}$. $(\mathcal{X}^n,\mathcal{A}^n)$, P^n pro $P\in\mathcal{P}$ jsou obvyklé kartézské součiny a $\underline{x}=(x_1,\dots,x_n)\in\mathcal{X}^n$ je výběr rozsahu $n\in\mathcal{N}$. Pro každý výběr $\underline{x}\in\mathcal{X}^n$, $n\in\mathcal{N}$, definujeme empirické rozdělení

(1) $$P_n=\frac{1}{n}\sum_{i=1}^{n}1_{\{x_i\}}\in\mathcal{P}.$$

Třídu všech empirických rozdělení označujeme $\mathcal{P}_{emp}$.

O parametrickém prostoru Θ předpokládáme, že je lokálně kompaktní Hausdorffův topologický prostor se spočetnou basí $\mathcal{B}_1$. Symbolem $\mathcal{B}_0\supset\mathcal{B}_1$ označíme topologii a symbolem $\mathcal{B}\supset\mathcal{B}_0$ σ-algebru borelovských podmnožin prostoru Θ. $\mathcal{P}_\Theta\subset\mathcal{P}$ označuje teoretickou rodinu pomocí které statistik popisuje experiment, který produkuje výběrové údaje x_i a $\mathcal{Q}_\Theta$ výběrovou rodinu, která ve skutečnosti popisuje uvedený experiment (v tom smyslu, že $\underline{x}=(x_1,\dots,x_n)$ je realizací náhodného vektoru $\underline{X}=(X_1,\dots,X_n)$ s výběrovým pravděpodobnostním prostorem $(\mathcal{X}^n,\mathcal{A}^n,Q_\theta^n)$ pro některé $Q_\theta\in\mathcal{Q}_\Theta$). V dalším se omezujeme na teoretické rodiny $\mathcal{P}_\Theta$ s $P_{\tilde\theta}\neq P_\theta$ pro $\tilde\theta\neq\theta$.

Pod odhadem $T=\{T_n:\ n\in\mathcal{N}\}$ rozumíme posloupnost $(\mathcal{A}^n,\mathcal{B})$-měřitelných zobrazení $T_n:\mathcal{X}^n\to\Theta$. Pokud výslovně nebude uveden jiný kvantifikátor, ve všech výrocích a formulích této práce obsahujících parametr $n\in\mathcal{N}$ předpokládáme kvantifikátor "pro všechna $n\in\mathcal{N}$". Symbolem $\mathcal{T}$ označíme třídu všech odhadů v problému s rodinou $\mathcal{P}_\Theta$. Funkce

$$R_n(\theta\,|\,T,\mathcal{P}_\Theta)=E_{P_\theta^n}L(T_n,\theta)\quad\text{pro }\theta\in\Theta,\ T\in\mathcal{T}$$

nebo obecněji

(2) $$R_n(\theta\,|\,T,\mathcal{P}_\Theta)=E_{P_\theta^n}D(P_{T_n},P_\theta)\quad\text{pro }\theta\in\Theta,\ T\in\mathcal{T}$$

je riziková funkce odhadu T při rodině $\mathcal{P}_\Theta$. Zde

$$L:\Theta\times\Theta\to[0,\infty]$$

je $\mathcal{B}\times\mathcal{B}$-měřitelná ztrátová funkce s $L(\tilde\theta,\theta)=0$ právě když $\tilde\theta=\theta$ a

$$D: \mathcal{P}_\Theta \times \mathcal{P}_0 \to [0,\infty] \quad ; \mathcal{P}_\Theta \cup \mathcal{P}_{emp} \subset \mathcal{P}_0$$

je divergence jejíž restrikce $D(P_{\tilde\theta},P_\theta)$ je $\mathcal{B}\times\mathcal{B}$ -měřitelná a rovna nule právě když $P_{\tilde\theta} = P_\theta$. Pro teoretické rodiny s $P_{\tilde\theta} \neq P_\theta$ pro $\tilde\theta \neq \theta$ je první uvedená riziková funkce speciálním případem (2) při divergenci $D(P_{\tilde\theta},P_\theta) = L(\tilde\theta,\theta)$ pro $(P_{\tilde\theta},P_\theta) \in \mathcal{P}_\Theta \times \mathcal{P}_\Theta$ a $D(P_{\tilde\theta},Q) = \infty$ při $(P_{\tilde\theta},Q) \in \mathcal{P}_\Theta \times (\mathcal{P} - \mathcal{P}_\Theta)$.

$T \in \mathcal{T}$ je <u>klasický</u> (stejnoměrně nejlepší v $\mathcal{T}^* \subset \mathcal{T}$, minimaxní nebo Bayesův pro apriorní rozdělení W na $(\Theta,\mathcal{B})$) jestliže pro funkci (2) platí buď

(3) $\qquad R_n(\theta|T,\mathcal{P}_\Theta) \le R_n(\theta|T^*,\mathcal{P}_\Theta)$ pro všechna $\theta\in\Theta$ a $T^*\in\mathcal{T}^*$

nebo

(4) $\qquad \sup_{\theta\in\Theta} R_n(\theta|T,\mathcal{P}_\Theta) \le \sup_{\theta\in\Theta} R_n(\theta|T^*,\mathcal{P}_\Theta)$ pro všechna $T^*\in\mathcal{T}$

anebo

(5) $\qquad E_W R_n(\theta|T,\mathcal{P}_\Theta) \le E_W R_n(\theta|T^*,\mathcal{P}_\Theta)$ pro všechna $T^*\in\mathcal{T}$.

Z následující lemmy vyplývá, že je-li odhad stejnoměrně nejlepší ve třídě všech odhadů, pak je také minimaxní a univerzálně (pro všechna W) Bayesův. Dále je z ní vidět, proč se v (3) obecně musíme omezovat na vlastní podtřídy $\mathcal{T}^* \subset \mathcal{T}$.

LEMMA 1. Nechť $R_n(\theta|T,\mathcal{P}_\Theta)$ je libovolná funkce na $\Theta\times\mathcal{T}$ a nechť pro některý odhad $T\in\mathcal{T}$ platí (3) s $\mathcal{T}^*=\mathcal{T}$. Potom pro tento odhad platí jak (4) tak také (5) pro všechna rozdělení W. Splňuje-li některý odhad T při funkci (2) podmínku (3) s $\mathcal{T}^*=\mathcal{T}$, potom $\mathcal{P}_\Theta$ je singulární rodina ve smyslu $P_{\tilde\theta} \perp P_\theta$ pro všechna $\tilde\theta \neq \theta \in \Theta$.

<u>Důkaz:</u> První tvrzení je zřejmé, protože nerovnosti v (4) a (5) vyplývají ze soustavy nerovností uvažovaných v (3). Pokud jde o druhé tvrzení, definujme třídu odhadů $\{T^\theta: \theta\in\Theta\} \subset \mathcal{T}$ podmínkou $T_n^\theta(\underline{x}) = \theta$ pro všechna $\underline{x}\in\mathcal{X}^n$. Z (3) vyplývá, že pro odhad T pak platí $R_n(\theta|T,\mathcal{P}_\Theta) \le R_n(\theta|T^\theta,\mathcal{P}_\Theta)$ pro všechna $\theta\in\Theta$. Z (2) a z předpokládané identity $D(P_\theta,P_\theta) = 0$ plyne $R_n(\theta|T^\theta,\mathcal{P}_\Theta) = 0$ pro všechna $\theta\in\Theta$. Z poslední nerovnosti a z (2) tedy plyne, že pro odhad T platí

$$E_{P_\theta^n} D(P_{T_n},P_\theta) = 0 \quad \text{pro všechna } \theta\in\Theta .$$

Odsud a z předpokladu $D(P_{\tilde\theta},P_\theta) > 0$ pro $\tilde\theta \neq \theta$ plyne $P_\theta^n(T_n \neq \theta) = 0$ pro všechna $\theta\in\Theta$. Množiny $A_\theta = \{\underline{x}: T_n(\underline{x})=\theta\}$ jsou měřitelné s vlastností $P_\theta^n(A_\theta) = 1$ pro všechna $\theta\in\Theta$. Proto platí pro každou pevnou dvojici $\tilde\theta \neq \theta$ jednak $P_\theta^n(A_\theta) = 1$ a také $P_{\tilde\theta}^n(A_\theta) = 0$ odkud plyne $P_{\tilde\theta} \perp P_\theta$ C.B.D.

Nyní přikročíme k definici hvězdičkovaného protějšku ztrátové funkce (2) následujícím vztahem

(6) $\qquad R_n^*(\theta|T,\mathcal{P}_\Theta) = E_{P_\theta^n} D(P_{T_n},P_n)$ pro $\theta\in\Theta$, $T\in\mathcal{T}$.

Jak plyne z lemmy 2 a věty 1, v následující definici si můžeme dovolit použít podmínku, která je podle lemmy 1 nejsilnější z podmínek optimality (3) - (5).

Řekneme, že odhad T je <u>*-klasický</u>, jestliže pro funkci (6) platí

(7) $\qquad R_n^*(\theta|T,\mathcal{P}_\Theta) \le R_n^*(\theta|T^*,\mathcal{P}_\Theta)$ pro všechna $\theta\in\Theta$ a $T^*\in\mathcal{T}$.

Dále řekneme, že odhad T je <u>empirický</u>, jestliže

(8) $\qquad T_n(\underline{x}) \in \mathrm{Arg}\,(\min D(P_\theta,P_n)) \subset \Theta$ pro všechna $\underline{x}\in\mathcal{X}^n$ (viz (1)).

Z (6) a (7) plyne, že odhad totožný s.j. $[\mathcal{P}_\Theta]$ s $*$-klasickým je $*$-klasický. Dále platí pro všechna $a(\underline{x}) > 0$, $b(\underline{x}) \in \mathcal{R}$

$$\mathrm{Arg}(\min D(P_\theta,P_n)) = \mathrm{Arg}(\min(a(\underline{x})D(P_\theta,P_n)+b(\underline{x}))) \text{ pro všechna } \underline{x} \in \mathcal{X}^n .$$

Proto je rozumné definici empirického odhadu (7) rozšířit na všechny před-divergence $D: \mathcal{P}_\Theta \times \mathcal{P}_0 \to [-\infty,\infty]$ s $\mathcal{P}_\Theta \cup \mathcal{P}_{emp} \subset \mathcal{P}_0$, pro které existují divergence D_k s vlastnostmi požadovanými v definici (2) a funkce $a_n^k(\underline{x}) > 0$, $b_n^k(\underline{x}) > 0$, $\underline{x} \in \mathcal{X}^n, k,n \in \mathcal{N}$ takové, že platí

(9) $$D(P_\theta,P_n) = \lim_{k\to\infty} (a_n^k(\underline{x})\, D_k(P_\theta,P_n) + b_n^k(\underline{x})) \text{ s.j. } [P_\theta^n] \quad \text{pro všechna } \theta \in \Theta .$$

LEMMA 2. Každý empirický odhad je $*$-klasický.

Důkaz: Jasné z (6) - (8).

3. Příklady empirických odhadů

Uvedeme tři příklady, které nejsou ilustrativní, nýbrž jen poukazují na souvislost empirických odhadů s některými třídami divergenčních odhadů, o kterých se zmiňují naše první práce na téma empirického odhadování [2,3,4].

PŘÍKLAD 1. Nechť $f: [0,\infty] \to (-\infty,\infty]$ je spojitá a konvexní a navíc ryze konvexní v bodě 1 s $f(1) = 0$. Za podmínek $0f(0/0) = 0$, $0f(p/0) = p \lim_{u\to\infty} f(u)/u$ definujeme výrazem

$$D_f(P,Q) = E_\lambda\, qf(p/q),\ p = \frac{dP}{d\lambda},\ q = \frac{dQ}{d\lambda} \text{ pro } \lambda \gg P,Q$$

tzv. f-divergenci rozdělení $P, Q \in \mathcal{P}$. Jak známo (lemma 2.1 v [3]), jde o nezápornou funkci definovanou dobře a nezávisle na λ a to na celém oboru $\mathcal{P}\times\mathcal{P}$. Ze zmíněné lemmy také vyplývá, že její zúžení na libovolném $\mathcal{P}_\Theta \times \mathcal{P}$ má vlastnosti požadované v (2) s $\mathcal{P}_0 = \mathcal{P}$. Empirické odhady příslušné této divergenci nazveme ve shodě s [2,3] standardními divergenčními odhady (f-odhady).

PŘÍKLAD 2. Nechť $\mathcal{A}_* = \{A_x : x \in \mathcal{X}\} \subset \mathcal{A}$, $\mathcal{P}_* = \{P \in \mathcal{P}: F(x) = P(A_x)\ \mathcal{A}\text{-měřitelná}\}$, $g(u,v) = vf(u/v)+(1-v)f(1-(1)/(1-v))$ pro $u,v \in [0,1]$, kde f je jako v příkladě 1. Lze ukázat, že za podmínek uvažovaných v Lemmě 2.1 práce [3] je g spojitá konvexní funkce na $[0,1] \times [0,1]$ s hodnotami na rozšířené $\mathcal{R}$. Na $\mathcal{P}_* \times \mathcal{P}_*$ definujme při některé míře λ na $(\mathcal{X},\mathcal{A})$ slabou f-divergenci rozdělení P, Q se zobecněnými distribučními funkcemi F(x), G(x) vztahem

$$WD_f(P,Q) = E_\lambda\, g(F,G).$$

Předpokládejme, že existuje třída $\mathcal{P}_{emp} \subset \mathcal{P}_0 \subset \mathcal{P}_*$ s $WD_f(P,Q) > 0$ pro všechna $P,Q \in \mathcal{P}_0$. Potom pro každou $\mathcal{P}_\Theta$ má zúžení slabé divergence na $\mathcal{P}_\Theta \times \mathcal{P}_0$ vlastnosti požadované v definici (2). Empirické odhady příslušné této divergenci nazveme ve shodě s [2,3,4] slabě divergenčními odhady ((λ,g)-odhady).

Následující příklad je založen na následující lemmě, která se opírá o tuto definici. Řekneme, že míra λ na $(\mathcal{X},\mathcal{A})$ je U-spojitá jestliže existují postupně zjemňované $\mathcal{A}$-měřitelné spočetné rozklady $\mathcal{D}_1, \mathcal{D}_2, \dots$ prostoru $\mathcal{X}$, jejichž sjednocení generuje $\mathcal{A}$ a $\lambda(A) = 1/k$ pro všechna $A \in \mathcal{D}_k$, $k \in \mathcal{N}$. Je zřejmé, že U-spojitá míra λ je σ-konečná a platí pro ni $\lambda(\{x\}) = 0$ pro všechna $x \in \mathcal{X}$ a že posloupnost $\mathcal{A}_k$, $k \in \mathcal{N}$, pod-σ-algeber generovaných rozklady $\mathcal{D}_k$ je filtrace na $(\mathcal{X},\mathcal{A})$.

LEMMA 3. Nechť pro f z příkladu 1 $\lim_{u\to\infty} f(u)/u = 0$, $0f(0) = \lim_{u\to 0} uf(u) = 0$,

(10) $f(uv) = \varphi(v)\, f(u) + f(v)$ pro nějakou $\varphi : [0,\infty] \to [-\infty,\infty]$ a všechna $u,v \in [0,\infty]$

a nechť $\mathcal{P}_\Theta \ll \lambda$ pro nějakou U-spojitou míru λ. Potom před-divergence

(11) $D(P_\theta,Q) = E_Q f(p_\theta)$ $p_\theta = dP_\theta/d\lambda$ pro všechna $P_\theta \in \mathcal{P}_\Theta$, $Q \in \mathcal{P}$

je dobře definovaná na $\mathcal{P}_\Theta \times \mathcal{P}$ a platí pro ni (9) s $D_k(P_\theta,P_n) = D_f(P_\theta^{(k)},P_n^{(k)})$, kde $P_\theta^{(k)}$, $P_n^{(k)}$ jsou restrikce P_θ, P_n na výše definované pod-σ-algebře $\mathcal{A}_k$.

Důkaz: Každá funkce f je za podmínek uvažovaných v příkladě 1 zdola ohraničená a proto střední hodnota v (11) je dobře definovaná. Řekneme, že $\underline{x} \in \mathcal{X}^n$ má vlastnost $\mathcal{V}_r$ jestliže r je nejmenší přirozené číslo pro které existují vzájemně různé $A_i \in \mathcal{D}_r$ pro které $x_i \in A_i$, $i = 1,\dots,n$. Označíme $\mathcal{D}_r(\underline{x}) = \{A_1,\dots,A_n\}$. Uspořádaná podtřída $\mathcal{D}_r(\underline{x})$ je vektorem $\underline{x}$ s vlastností $\mathcal{V}_r$ určena jednoznačně, protože když např. $x_1 \in A_1$, nemůže již patřit do žádné jiné množiny rozkladu $\mathcal{D}_r$. Dále označme symbolem $A^{(n)}$ množinu těch vektorů $\underline{x} \in \mathcal{X}^n$ jejichž alespoň dvě souřadnice x_i, x_j jsou totožné. Protože jednoprvkové množiny $\{x\}$ jsou $\mathcal{A}$-měřitelné, je také $A^{(n)}$ $\mathcal{A}^n$-měřitelná a protože $\lambda(\{x\}) = 0$, je také $\lambda^n(A^{(n)}) = 0$. Dále každé dva různé body $x_i,x_j \in \mathcal{X}$ lze pokrýt dvěma různými množinami z $\mathcal{D}_k$ pro všechna k od jistého počínaje (opak vede ke sporu s předpokladem, že $\mathcal{X}$ je Hausdorffův). Odsud plyne jednak že každý bod $\underline{x} \in \mathcal{X}^n - A^{(n)}$ má vlastnost $\mathcal{V}_r$ pro některé $r \in \mathcal{N}$ a také že pro každé $\underline{x} \in \mathcal{X}^n - A^{(n)}$ a pro všechna $k > r$ existuje právě jedna množina $\mathcal{C}_k(\underline{x}) = \{A_1,\dots,A_n\}$ vzájemně různých $A_i \in \mathcal{D}_k$ s vlastností $x_i \in A_i$ pro $i = 1,\dots,n$ (tedy $\mathcal{C}_r(\underline{x}) = \mathcal{D}_r(\underline{x})$). Protože jak přirozených čísel r tak i různých tříd $\mathcal{D}_r(\underline{x})$ je nejvýše spočetně mnoho, množina $\mathcal{X}^n - A^{(n)}$ se rozpadá na nejvýše spočetně mnoho množin A^* s vlastností $\mathcal{V}_r$ a s $\mathcal{D}_r(\underline{x}) = \{A_1,\dots,A_n\} \subset \mathcal{D}_r$ pro všechna $\underline{x} \in A^*$. Protože

$$A^* = \bigtimes_{i=1}^{n} A_i,$$

platí $A^* \in \mathcal{A}^n$. V dalším budeme uvažovat libovolnou pevnou množinu A^*, označíme

$$\tilde{A} = \bigcup_{i=1}^{n} A_i \in \mathcal{A},$$

a definujeme $\tilde{P} \in \mathcal{P}$ vztahem

$$\tilde{P}(A) = \frac{\lambda(A \cap \tilde{A})}{\lambda(\tilde{A})} = \frac{r}{n}\,\lambda(A \cap \tilde{A}) \text{ pro všechna } A \in \mathcal{A}.$$

Pro divergenci D_k uvažovanou v Lemmě 3 a všechna $\underline{x} \in A^*$ a $k > r$ zřejmě platí (viz definici f-divergence v příkladě 1)

$$D_k(P_\theta,P_n) = \sum_{A \in \mathcal{D}_k} P_n(A)\, f\Big(\frac{P_\theta(A)}{P_n(A)}\Big) = \sum_{A \in \mathcal{C}_k(\underline{x})} P_n(A)\, f\Big(\frac{P_\theta(A)}{P_n(A)}\Big).$$

Podle (10) tedy také platí

$$D_k(P_\theta,P_n) = \sum_{A \in \mathcal{C}_k(\underline{x})} P_n(A)\Big[\varphi\Big(\frac{\tilde{P}(A)}{P_n(A)}\Big)\, f\Big(\frac{P_\theta(A)}{\tilde{P}(A)}\Big) + f\Big(\frac{\tilde{P}(A)}{P_n(A)}\Big)\Big].$$

Protože pro $A \in \mathcal{C}_k(\underline{x}) \subset \mathcal{D}_k$ platí $\lambda(A \cap \tilde{A}) = \lambda(A) = 1/k$, platí též $\tilde{P}(A) = r/(nk)$ a samozřejmě $P_n(A) = 1/n$. Platí tudíž též

$$D_k(P_\theta,P_n) = \sum_{A \in \mathcal{C}_k(\underline{x})} P_n(A)\Big[\varphi\Big(\frac{r}{k}\Big) f\Big(\frac{P_\theta(A)}{\tilde{P}(A)}\Big) + f\Big(\frac{r}{k}\Big)\Big] = \varphi\Big(\frac{r}{k}\Big) \sum_{A \in \mathcal{C}_k(\underline{x})} P_n(A) f\Big(\frac{P_\theta(A)}{\tilde{P}(A)}\Big) + f\Big(\frac{r}{k}\Big).$$

S využitím konvencí zavedených v Lemmě 3 a příkladě 1 můžeme tedy psát

$$D_k(P_\theta,P_n) = \varphi(\tfrac{r}{k}) \sum_{A\in\mathcal{D}_k} P_n(A)\, f\Big(\frac{P_\theta(A\cap\tilde{A})}{\tilde{P}(A)}\Big) + f(\tfrac{r}{k}).$$

Jestliže označíme $\tilde{p}_\theta = dP_\theta(.\cap\tilde{A})/d\tilde{P}$, pak zřejmě $P_\theta(A\cap\tilde{A})/\tilde{P}(A)$ koinciduje s podmíněnou střední hodnotou $E_{\tilde{P}}[\tilde{p}_\theta|\mathcal{A}_k]$ na $A\in\mathcal{D}_k$. Dále zřejmě platí

$$\tilde{p}_\theta = \lambda(\tilde{A})p_\theta = \frac{n}{r}\, p_\theta \quad \text{na } \tilde{A},$$

takže též

$$E_{\tilde{P}}[\tilde{p}_\theta|\mathcal{A}_k] = \frac{n}{r}\, E_{\tilde{P}}[p_\theta|\mathcal{A}_k] \quad \text{na } \mathcal{X}.$$

Dokázali jsme tedy, že platí

$$D_k(P_\theta,P_n) = \varphi(\tfrac{r}{k})\, E_{P_n} f(E_{\tilde{P}}[\tilde{p}_\theta|\mathcal{A}_k]) + f(\tfrac{r}{k}) =$$

$$= \varphi(\tfrac{r}{k})\big[\varphi(\tfrac{n}{r})\, E_{P_n} f(E_{\tilde{P}}[\tilde{p}_\theta|\mathcal{A}_k]) + f(\tfrac{n}{r})\big] + f(\tfrac{r}{k}) \qquad \text{(viz (10)).}$$

Zde $\{E_{\tilde{P}}[p_\theta|\mathcal{A}_k]: k\in\mathcal{N}\}$ je regulární martingál na $(\mathcal{X},\mathcal{A},\tilde{P})$, který tudíž s. j. $[\tilde{P}]$ konverguje k p_θ. Nechť A_0 je podmnožina těch $x\in\tilde{A}$ pro které zmíněná konvergence neplatí. Protože $\lambda(A_0) = n\tilde{P}(A_0)/r = 0$, platí též

$$\lambda(A^* - {}^*A) = 0 \quad \text{kde} \quad {}^*A = \mathop{\times}_{i=1}^{n} (A_i - A_0), \quad A^* = \mathop{\times}_{i=1}^{n} A_i.$$

Protože dále pro všechna $\underline{x}\in{}^*A$ zřejmě platí

$$E_{P_n} f(E_{\tilde{P}}[p_\theta|\mathcal{A}_k]) \longrightarrow E_{P_n} f(p_\theta) = \frac{1}{n}\sum_{i=1}^{n} f(p_\theta(x_i)),$$

dokázali jsme, že pro skoro všechny $[\lambda^n]$ výběry $\underline{x}\in A^*$ existují funkce $d_k(\theta,\underline{x})$, které konvergují k $D(P_\theta,P_n) = E_{P_n} f(p_\theta)$ při $k\to\infty$ pro všechna $\theta\in\Theta$, přičemž platí

(12) $$D_k(P_\theta,P_n) = \varphi(\tfrac{r}{k})\,\varphi(\tfrac{n}{r})\, d_k(\theta,\underline{x}) + \varphi(\tfrac{r}{n})\, f(\tfrac{n}{r}) + f(\tfrac{r}{k}).$$

Protože $\lambda^n(A^{(n)}) = 0$ a $\mathcal{X}^n - A^{(n)}$ se rozkládá na nejvýše spočetně mnoho různých množin A^*, poslední tvrzení platí pro skoro všechny $[\lambda^n]$ výběry $\underline{x}\in\mathcal{X}^n$. Kdyby v (10) platilo $\varphi(v) = 0$ pro některé $v > 0$, plynulo by z (10), že f je konstanta na $[0,\infty]$ což by byl spor s předpokladem, že f je ryze konvexní v bodě 1. Tudíž $\varphi(v) > 0$ pro všechny $v\in(0,\infty]$. Odsud a z (12) plyne, že (9) platí při

$$a_{n,k}(\underline{x}) = \Big[\varphi\Big(\frac{r(\underline{x})}{k}\Big)\varphi\Big(\frac{n}{r(\underline{x})}\Big)\Big]^{-1}, \qquad b_{n,k}(\underline{x}) = -a_{n,k}(\underline{x})\Big[\varphi\Big(\frac{r(\underline{x})}{n}\Big) f\Big(\frac{n}{r(\underline{x})}\Big) + f\Big(\frac{r(\underline{x})}{k}\Big)\Big]$$

pro všechna $\theta\in\Theta$ a skoro všechna $[\lambda^n]$ $\underline{x}\in\mathcal{X}^n$. Odsud a z předpokladu $\mathcal{P}_\Theta \ll \lambda$ již evidentně plyne i tvrzení lemmy 3.

PŘÍKLAD 3. Nechť $\mathcal{P}_\Theta$ je libovolná rodina dominovaná některou U-spojitou mírou na $(\mathcal{X},\mathcal{A})$ (je-li $\mathcal{X} = R^k$, pak Lebesgova míra $\lambda = \mu^k$ je zřejmě U-spojitá). Funkce

$$f(u) = \frac{1-u^\alpha}{\alpha}, \quad \varphi(u) = u^\alpha \qquad \text{pro } \alpha\in[0,1)$$

zřejmě vyhovují podmínkám lemmy 3 (při $\alpha = 0$ se spojitě dodefinovává $f(u) = -\log u$). Odsud, z (8) a z lemmy 3 již vyplývá, že všechny odhady $T = T^\alpha$ splňující při všech $\underline{x}\in\mathcal{X}^n$ podmínky

$$T_n^\alpha(\underline{x}) \in \mathrm{Arg}\Big(\max \Big\langle \begin{matrix} E_{P_n}\, p_\theta^\alpha & \text{pro } \alpha\in(0,1) \\ E_{P_n} \ln p_\theta & \text{pro } \alpha = 0 \end{matrix}\Big)$$

jsou empirické.

Ve shodě s [2,3] nazveme tyto odhady řízené divergenční odhady ((λ,α)-odhady). Maximálně věrohodný odhad je $(\lambda,0)$-odhad, pokud je definován pro U-spojitou dominující míru λ.

Rozdíl mezi divergenčními odhady zde a v pracích [2,3,4] spočívá v tom, že zde připouštíme obecná pravidla výběru $(\underline{x},T(\underline{x})) \to \tau(\underline{x},T(x))$ $T_n(\underline{x})$ z množin $T(\underline{x})$ = Arg(min $D(P_\theta,P_n)$) zatím co tam se předpokládalo speciální pravidlo $(\underline{x},T(\underline{x})) \to \tau(T(\underline{x})) = T_n(\underline{x})$ s určitou vlastností spojitosti, kterou lze splnit jen v triviálních parametrických prostorech. Asymptotické výsledky zmíněných prací však nevyužívaly zmíněného pravidla výběru a platí pro každé pravidlo zaručující měřitelnost odhadu. Přenášejí se tedy i na zde uvažované varianty bez jakékoliv změny.

4. Existence empirických odhadů

Uvažujme jednobodovou kompaktifikaci parametrického prostoru Θ pomocí některého $\theta^* \notin \Theta$, označme $\Theta^* = \Theta \cup \{\theta^*\}$ příslušný kompaktní prostor a předpokládejme (aniž dodefinováváme objekt P_{θ^*} jako takový), že pro každé $\underline{x} \in \mathcal{X}^n$ existuje

$$D(P_{\theta^*},P_n) = \lim_{\theta \to \theta^*} D(P_\theta,P_n).$$

VĚTA 1. Je-li pro každé $\theta \in \Theta$ funkce $D(P_\theta,P_n): \mathcal{X}^n \to [-\infty,\infty]$ $\mathcal{A}^n$-měřitelná a pro každé $\underline{x} \in \mathcal{X}^n$ funkce $D(P_\theta,P_n): \Theta \to [-\infty,\infty]$ spojitá a shora ostře ohraničená hodnotou $D(P_{\theta^*},P_n)$, tj.

(13) $$D(P_\theta,P_n) < D(P_{\theta^*},P_n) \quad \text{pro všechna } \theta \in \Theta,\ \underline{x} \in \mathcal{X}^n,$$

pak existuje empirický odhad T.

__Důkaz:__ Označme symbolem $D^*(P_\theta,P_n)$ rozšířenou funkci $D(P_\theta,P_n): \Theta^* \to [-\infty,\infty]$. Z předpokladu spojitosti funkce D a z definice $D(P_{\theta^*},P_n)$ vyplývá, že D^* je spojitá na Θ^* při každém $\underline{x} \in \mathcal{X}^n$. Podle posledního tvrzení ve větě 1.9 Pfanzagla [1] je pro každou kompaktní $B \subset \Theta^*$ funkce

$$\underline{x} \to \inf_{\theta \in B} D^*(P_\theta,P_n)$$

$\mathcal{A}^n$-měřitelná. Do třetice, ze spojitosti D^* a kompaktnosti Θ^* vyplývá, že

$$\text{Arg}(\min D^*(P_\theta,P_n)) \subset \Theta^*$$

je neprázdná pro každé $\underline{x} \in \mathcal{X}^n$. Podle věty 3.10 Pfanzagla [1] existuje tedy $\mathcal{A}^n$-měřitelné zobrazení $T_n: \mathcal{X}^n \to \Theta^*$ s hodnotami $T_n(\underline{x})$ v Arg(min $D^*(P_\theta,P_n)$). Protože z (13) plyne

$$\text{Arg}(\min D^*(P_\theta,P_n)) = \text{Arg}(\min D(P_\theta,P_n)) \text{ pro všechna } \underline{x} \in \mathcal{X}^n,$$

plyne odsud a z (8) tvrzení věty 1.

5. Výpočet empirických a klasických odhadů

Jak vyplývá z definice (8), výpočet hodnoty $T_n(\underline{x}) \in \Theta$ empirického odhadu při pevném výběru $\underline{x} \in \mathcal{X}^n$ je úloha matematického programování: hledání jedné (kterékoliv) hodnoty argumentu minimalizující funkci $d(\theta|\underline{x}) = D(P_\theta,P_n)$ (otázka měřitelnosti nemá žádný patrný praktický význam). K jejímu řešení za určitých předpokladů o funkci $d(\theta|\underline{x})$ a kroku 1 a 5 vede následující program.

Uložení $1 \to i$, $\theta_0 \to \theta$, $d(\theta_0|\underline{x}) \to d(\theta|\underline{x})$.

1. Výpočet θ_i.
2. Výpočet $d(\theta_i|\underline{x})$.
3. Porovnání $d(\theta_i|\underline{x})$, $d(\theta|\underline{x})$.
4. $\theta_i \to \theta$, $d(\theta_i|\underline{x}) \to d(\theta|\underline{x})$ když $d(\theta_i|\underline{x}) < d(\theta|\underline{x})$.
5. $i + 1 \to i$ a rozhodování o zastavení.
6. Tisk θ, stop.

Jak jsme naznačili v úvodu, za následujících předpokladů je též výpočet klasického Bayesova odhadu úlohou matematického programování: Pro každé $\underline{x} \in \mathcal{X}^n$ existuje aposteriorní rozdělení $P_{\underline{x}}$ na $(\Theta, \mathcal{B})$ a existuje marginální rozdělení $\bar{P}$ na $(\mathcal{X}, \mathcal{B})$. Úlohu funkce $d(\theta|\underline{x})$ pak přebírá aposteriorní riziko

(14) $$d_*(\theta|\underline{x}) = \int_{\Theta} D(P_\theta, P_{\tilde\theta})\, dP_{\underline{x}}(\tilde\theta) \doteq \sum_{i=1}^{M} D(P_\theta, P_{\theta_i})\, P_{\underline{x}}(A_i),$$

kde $\mathcal{D} = \{A_1, \dots, A_k\} \in \mathcal{A}$ je rozklad parametrického prostoru. Tuto úlohu řeší za stejných předpokladů výše uvedený program, pouze funkce d je nahražena funkcí d_*. Rozdíl je v tom, že krok 2:"výpočet nové hodnoty d_*" znamená zde realizaci následujícího podprogramu.

Uložení $1 \to i$, $0 \to \Sigma$.

O. Výpočet A_i a $P_{\underline{x}}(A_i)$.
I. Výpočet θ_i.
II. Výpočet $d(\theta_i|\underline{x})$.
III. Výpočet $d(\theta_i|\underline{x})\, P_{\underline{x}}(A_i)$.
IV. $d(\theta_i|\underline{x})\, P_{\underline{x}}(A_i) + \Sigma \to \Sigma$.
V. $i + 1 \to i$ a rozhodování zda $i + 1 > M$.
VI. Tisk Σ, stop.

Kroky I-VI tohoto podprogramu jsou v podstatě identické s kroky 1-6 výše uvedeného programu, krok O je navíc. Jestliže tedy S vyjadřuje počet operací programu pro empirický odhad, pak počet operací programu pro klasický odhad bude přibližně M(S+K) kde K je počet operací v kroku O. Protože k dosažení potřebné přesnosti v (14) jsou zpravidla nutná velká M, vidíme, že výpočet klasického odhadu je zpravidla mnohonásobně složitější než výpočet empirického odhadu. Přitom není vždy jasné, jak na počítači realizovat výpočet $P_{\underline{x}}(A_i)$. Výjimku představují jen případy, kdy rozdělení $P_{\underline{x}}$ lze zadat analytickým předpisem (normální rodina, normální apriorní rozdělení).

6. Robustnost empirických odhadů

První část diskuse bude intuitivní. Klasické odhady při dané (před-)divergenci D a rodině $\mathcal{P}_\Theta$ minimalizují podle (3)-(5) rizikové funkce $R_n(\theta|T,\mathcal{P}_\Theta) = E_{P_\theta^n} D(P_{T_n}, P_\theta)$. $*$-klasické odhady minimalizují podle (7) rizikové funkce $R_n^*(\theta|T,\mathcal{P}_\Theta) = E_{P_\theta^n} D(P_{T_n}, P_n)$. Předpokládejme nyní, že výběr je generován rodinou $\mathcal{Q}_\Theta$, jejíž rozdělení Q_θ jsou blízká P_θ stejnoměrně na Θ. Vhodná spojitost (před-)divergence spolu s Glivenko--Cantelliho větou pak implikují při $n \to \infty$ $R_n^*(\theta|T,\mathcal{P}_\Theta) \doteq E_{P_\theta^n} D(P_{T_n}, Q_\theta)$. Porovnejme nyní rizikovou funkci $E_{Q_\theta^n} D(Q_{T_n},Q_\theta)$, která poskytuje klasický optimální odhad pro neznámou rodinu $\mathcal{Q}_\Theta$, s rizikovými funkcemi $E_{P_\theta^n} D(P_{T_n},P_\theta)$, $E_{P_\theta^n} D(P_{T_n},Q_\theta)$ klasického a empirického odhadu při rodině $\mathcal{P}_\Theta$ (viz Lemma 2). Z definice (8) vyplývá, že empirický odhad minimalizuje také funkci $E_{Q_\theta^n} D(P_{T_n},Q_\theta)$. Zatím co riziková funkce klasického odhadu $E_{P_\theta^n} D(P_{T_n},P_\theta)$ se s rostoucím n nijak neadaptuje na neznámou rodinu $\mathcal{Q}_\Theta$, riziková funkce empirického odhadu $E_{Q_\theta^n} D(P_{T_n},Q_\theta)$ bude blízká rizikové funkci $E_{Q_\theta^n} D(Q_{T_n},Q_\theta)$ pokud z předpokládané blízkosti P_{T_n} a Q_{T_n} na $\mathcal{X}^n$ vyplývá stejnoměrná vzájemná blízkost (před-) divergencí $D(P_{T_n},Q_\theta)$ a $D(Q_{T_n},Q_\theta)$ na $\mathcal{X}^n$. Protože pro konzistentní odhady $T_n \doteq \theta$, bude nás zajímat citlivost (před-)divergence $D(P,Q)$ na změny rozdělení P v okolí Q.

Nechť $\mathcal{P}_1 \subset \mathcal{P}$ je určitá třída kontaminačních rozdělení. Citlivost (před-)divergence $D: \mathcal{P}_0 - \mathcal{P}_{emp} \times \mathcal{P}_0 \to [-\infty,\infty]$ v rozdělení $Q \in \mathcal{P}_0 - \mathcal{P}_{emp}$ definujeme takto

$$C_D(Q) = \sup_{P\in\mathcal{P}_1} \limsup_{\varepsilon\downarrow 0} \left| \frac{D((1-\varepsilon)Q+\varepsilon P,Q)-D(Q,Q)}{\varepsilon}\right| .$$

Řekneme dále, že odhad příslušný divergenci s citlivostí C_D je robustní vůči kontaminacím z $\mathcal{P}_1$ (pro rodiny $\mathcal{P}_\Theta \subset \mathcal{P}_0 - \mathcal{P}_{emp}$) jestliže C_D je konečná na $\mathcal{P}_0 - \mathcal{P}_{emp}$.

PŘÍKLAD 4. Lze ukázat, že pro před- divergenci příslušnou řízeným divergenčním odhadům ((λ,α)-odhadům) platí při $\mathcal{P}_0 - \mathcal{P}_{emp} \ll \lambda$, $\mathcal{P}_1 \ll \lambda$,

$$C_D(Q) = \sup_{P\in\mathcal{P}_1} \left| E_\lambda \frac{p-q}{q^{1-\alpha}}\right|, \quad p = \frac{dP}{d\lambda}, \quad q = \frac{dQ}{d\lambda} .$$

Platí zřejmé tvrzení: (λ,α)-odhad je robustní právě když

$$\sup_{P\in\mathcal{P}_1} E_\lambda \frac{p}{q^{1-\alpha}} < \infty .$$

Protože při $\alpha < \tilde{\alpha}$ platí, že $E_\lambda(p/q^{1-\tilde{\alpha}}) \to \infty$ (pro orientovaný systém hustot p) implikuje $E_\lambda(p/q^{1-\alpha}) \to \infty$, platí také, že je-li $\tilde{\alpha}$-odhad nerobustní, pak jsou nerobustní všechny α-odhady s $\alpha \in [0,\tilde{\alpha}]$ a je-li $\tilde{\alpha}$-odhad robustní, pak jsou robustní i všechny odhady s $\alpha \in [\tilde{\alpha},1)$. Maximálně věrohodný odhad ($(\lambda,0)$-odhad) je nerobustní zřejmě vždy když λ je nekonečná míra a $\mathcal{P}_1 \cap \mathcal{P}_0 - \mathcal{P}_{emp} \neq \emptyset$. Tímto se exaktně dokazuje teze vyslovená na základě čistě intuitivních úvah v [2,3], že totiž robustnost α-odhadů roste s rostoucím $\alpha \in [0,1)$.

7. Odkazy

[1] J. Pfanzagl: On the measurability and consistency of minimum contrast estimates. Metrika 14(1969), 249-272.

[2] I. Vajda: Minimum divergence principle in statistical estimation. Statistics and Decisions 2(1984) (v tisku).

[3] I. Vajda: Motivation, existence and equivariance of D-estimators. Kybernetika 20(1984), 189-208.

[4] I. Vajda: Minimum weak divergence estimators of structural parameters. Proc. 3rd Prague Symp. on Asympt. Statist., (Editors P.Mandl,M.Hušková),Elsevier, Amsterdam 1984.

[5] A. Wald: Statistical Decision Functions. J. Willey, New York 1950.

[6] S. Zacks: Theory of Statistical Inference. J. Willey, New York 1971.