KONZISTENCIA HĹBKY FUNKCIÍ

STANISLAV NAGY

nagy@karlin.mff.cuni.cz

Univerzita Karlova v Praze, Matematicko-fyzikální fakulta

Katedra pravděpodobnosti a matematické statistiky

ABSTRAKT

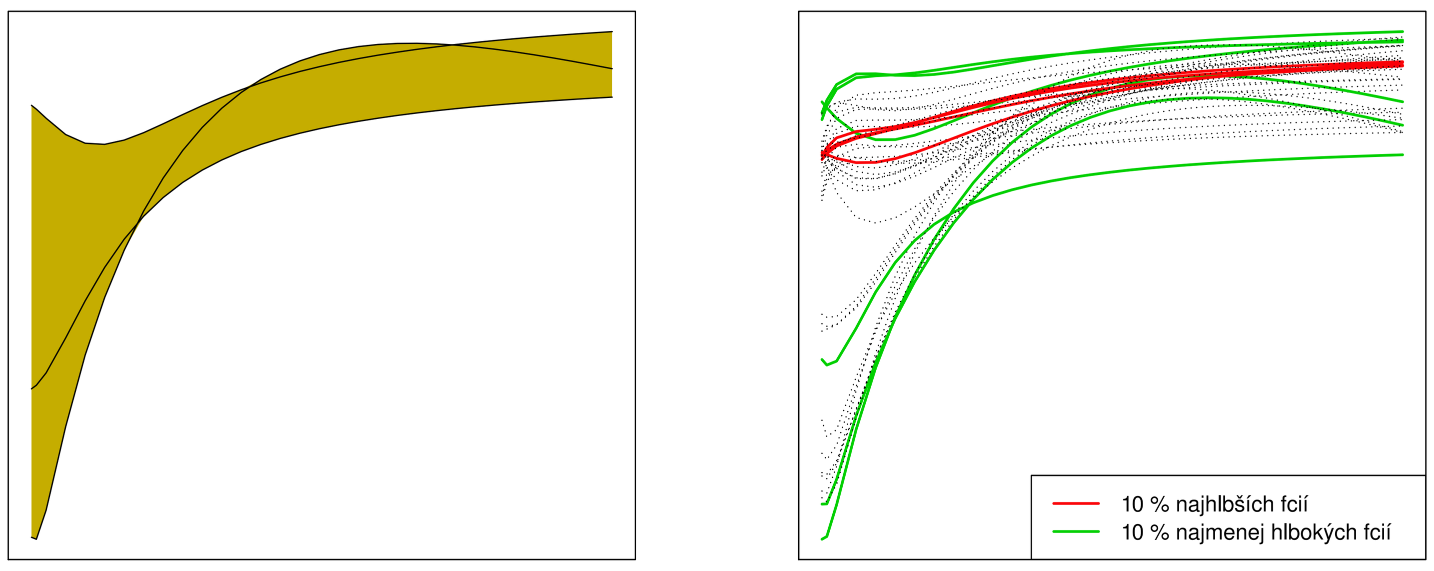
V príspevku sa zameriavame na funkcionálne dáta a niekoľko rôznych prístupov k určovaniu hĺbky funkcií používaných v literatúre. Ukazujeme, že silný výsledok autorov López-Pintado a Romo [4, Thm 4] o rovnomernej konzistencii výberovej pásovej hĺbky neplatí (ani za silnejších predpokladov). Navrhujeme dva rôzne prístupy k úprave pásovej hĺbky tak, aby rovnomernú konzistenciu už bolo možné zaručiť. Pre integrálne hĺbky tým rozširujeme výsledky Fraimana a Munizovej [1, Thm 3.1].

HĹBKA FUNKCIÍ

Štatistická hĺbková funkcia (hĺbka) je zobrazenie

$$D : S \times \mathcal{P}(S) \to \mathbb{R}^+ \quad (\to [0,1]),$$

kde $\mathcal{P}(S)$ označuje triedu pravdepodobnostných rozdelení na priestore S. Hĺbka indukuje na S usporiadanie v smere z „od centra k okrajom“, t.j. typické pozorovania v okolí „centra“ rozdelenia $P \in \mathcal{P}(S)$ majú vysoké hodnoty hĺbky, zatiaľ čo odľahlé a „netypické“ pozorovania voči P majú hĺbku blízku 0. Jedná sa o moderný nástroj neparametrickej analýzy dát, ktorý sa dá považovať za zovšeobecnenie kvantilov pre mnohorozmerné dáta. Výčet používaných hĺbok a ich vlastností pre $S = \mathbb{R}^d$ nájdeme napr. v článku Zua a Serflinga [5].



Obr. 1: Pás troch funkcií vľavo, pásová hĺbka náhodného výberu vpravo ($J = 3$).

V prípade funkcionálnych dát ($S = \mathcal{C}([0,1])$) existujú dva základné prístupy k určovaniu hĺbky. Nech $x \in \mathcal{C}([0,1])$ a $P \in \mathcal{P}(\mathcal{C}([0,1]))$.

- **Pásové hĺbky** (López-Pintado a Romo [4]) pre $J = 2, 3, \ldots$

$$BD^{J)}(x;P) = \frac{1}{J-1}\sum_{j=2}^{J} P\left(G(x) \subset B\left(X_1,\ldots,X_j\right)\right), \qquad (1)$$

  kde $G(x)$ je graf funkcie x a $B(X_1,\ldots,X_j)$ je pás, teda oblasť medzi minimom a maximom náhodných funkcií $X_1,\ldots,X_j$ (pozri obr. 1).
- **Integrálne hĺbky** (Fraiman a Muniz [1], Hlubinka a Nagy [2]) pre hĺbku D na $\mathbb{R}$

$$ID(x;P) = \int_0^1 D\left(x(t); P(t)\right)\, dt,$$

  kde $P(t)$ je marginálne rozdelenie P v t.

KONZISTENCIA HĹBKY

Každá rozumná hĺbka D musí byť konzistentná. Označme $P_n \in \mathcal{P}(S)$ rozdelenie náhodného výberu z P o rozsahu n. Potom D je konzistentná na S

- **bodove** ak $D(x;P_n) - D(x;P) \xrightarrow[n\to\infty]{s.i.} 0$ pre každé $x \in S$,
- **rovnomerne** ak $\sup_{x\in S}|D(x;P_n) - D(x;P)| \xrightarrow[n\to\infty]{s.i.} 0$,
- **univerzálne** ak $\sup_{x\in S}|D(x;P_n) - D(x;P)| \xrightarrow[n\to\infty]{s.i.} 0$ pre každé $P \in \mathcal{P}(S)$,
- **uniformne** ak $\sup_{P\in\mathcal{P}(S)}\sup_{x\in S}|D(x;P_n) - D(x;P)| \xrightarrow[n\to\infty]{s.i.} 0$.

(Ne)konzistencia pásovej hĺbky

Kľúčovým výsledkom práce López-Pintadovej a Roma [4, Thm 4] je

Tvrdenie: Nech $P \in \mathcal{P}(\mathcal{C}([0,1]))$ má absolútne spojité marginálne rozdelenia. Potom $BD^{J)}$ rovnomerne konzistentná na každej rovnako spojitej množine $E \subset \mathcal{C}([0,1])$.

Dôkaz: Pretože $\lim_{\|x\|\to\infty} BD^{J)}(x;P) = 0$, môžeme sa obmedziť sa na množinu $\{\|x\| < M\}$ pre $M > 0$. Rovnako obmedzená množina rovnako spojitých funkcií je podľa Arzéla-Ascoliho vety totálne obmedzená. Pretože $BD^{J)}(.;P)$ je spojitý funkcionál pre P s absolútne spojitými marginálnymi rozdeleniami, stačí dokázať pre $N \in \mathbb{N}$ pevné

$$\max_{\{x_i\}_{i=1}^N \subset E} \left|BD^{J)}(x_i;P_n) - BD^{J)}(x_i;P)\right| \xrightarrow[n\to\infty]{s.i.} 0$$

čo platí, pretože $BD^{J)}(.;P_n)$ je obmedzená U-štatistika. □

Problémy dôkazu sú ale dva: $BD^{J)}(.;P_n)$ je **iba polospojitá** (P_n nemá abs. spojité marg. rozdelenia). Navyše by **nestačila ani spojitosť** $BD^{J)}(.;P_n)$, ak by platila. Aby dôkaz fungoval, nutná by bola rovnaká spojitosť všetkých

$$\left\{BD^{J)}(x;P_n) - BD^{J)}(x;P)\right\}_{x\in E, n\in\mathbb{N}}.$$

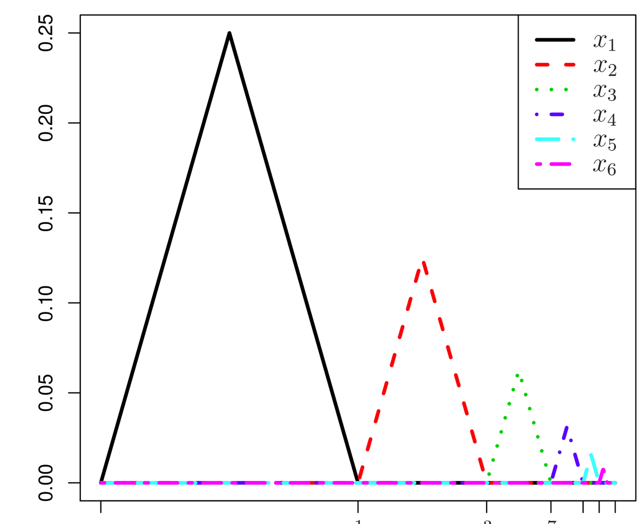
To, že $BD^{J)}$ skutočne nie je univerzálne konzistentná, je možné ukázať na príklade postupnosti (lipschitzovských) funkcií. Definujme $X \sim P \in \mathcal{P}(\mathcal{C}([0,1]))$ takto:

- $P(X(t) = 0 \text{ pre všetky } t \in [0,1]) = 0.5$.
- Rozdeľme interval $[0,1]$ „diadicky“ na disjunktné subintervaly dĺžok $\left\{2^{-j}\right\}_{j\in\mathbb{N}}$. Ak $X \not\equiv 0$, na každom subintervale bude X nulová s psťou 0.5 alebo nadobudne skok s psťou 0.5. Skoky nastávajú nezávisle od seba.
- Pre postupnosť $\{x_j\}_{j\in\mathbb{N}}$ funkcií s práve jedným skokom, inde 0 (obr. 2) platí

$$\sup_{j\in\mathbb{N}}\left|BD^{2)}(x_j;P_n) - BD^{2)}(x_j;P)\right| \geq 0.25 \text{ pre nekonečne veľa } n \in \mathbb{N} \text{ s.i.}$$

(Napríklad) $BD^{2)}$ teda **nie je rovnomerne konzistentná** voči P.

Rozdelenie P nie je marginálne abs. spojité, je však jednoduché modifikovať ho tak, aby bolo a zároveň protipríklad fungoval.



Obr. 2: Prvých šesť funkcií x_j.

Konzistentné funkcionály

Pretože $BD^{J)}$ je obmedzená U-štatistika, je **bodove konzistentná**. Ďalej, nahradením indikátoru v 1 **„lipschitzovskou“ verziou** je $BD^{J)}$ **uniformne konzistentná na kompaktných množinách**. Výsledok môžeme dokázať doplnením dôkazu López-Pintadovej a Roma o vlastnosti U-štatistík [3, Thm 2.1.4]. Univerzálna konzistencia sa dá dokázať aj priamo pomocou Blum-deHardtovho ZVČ.

Inou možnosťou je **prechod k integrálnym hĺbkam**, pre ktoré je možné dokázať univerzálnu konzistenciu bez akýchkoľvek predpokladov, ak je výberová verzia hĺbky U-štatistika. K-pásové hĺbky [2], rovnako ako Fraimanove-Munizovej hĺbky [1], sú teda **vždy univerzálne konzistentné**. Fraiman a Munizová [1, Thm 3.1] dokázali analogickú vetu za ďalších predpokladov na rozdelenie P.

Literatúra

[1] Ricardo Fraiman and Graciela Muniz. Trimmed means for functional data. *Test*, 10(2):419–440, 2001.

[2] Daniel Hlubinka and Stanislav Nagy. Functional data depth and classification. *submitted to Comput. Stat.*

[3] V. S. Koroljuk and Yu. V. Borovskich. *Theory of U-Statistics*. Kluwer, 1994.

[4] Sara López-Pintado and Juan Romo. On the concept of depth for functional data. *Journal of the American Statistical Association*, 104(486):718–734, 2009.

[5] Yijun Zuo and Robert Serfling. General notions of statistical depth function. *Ann. Stat.*, 28(2):461–482, 2000.

This work was supported by Czech Science Foundation (grant P402/12/G097)