Konzistencia hĺbky funkcií

Stanislav Nagy

Katedra pravděpodobnosti a matematické statistiky
Univerzita Karlova v Praze

Robust 2012

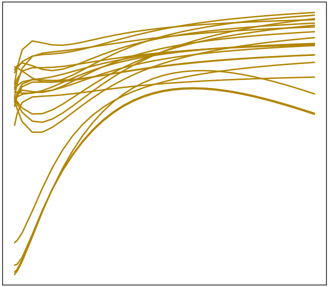
S podporou grantu GA ČR (grant P402/12/G097)

1. Pásová hĺbka a konzistencia
   - Konzistencia
   - Pásové hĺbky funkcií
2. Protipríklad
3. Konzistentné funkcionály
   - Oprava 1: Zospojitenie
   - Oprava 2: Integrálne hĺbky
4. Záver
   - Záver pre pásové hĺbky
   - Vyhliadky do budúcnosti

Funkcionálne dáta

$X \sim P \in \mathcal{P}(\mathcal{C}([0,1]))$ a náhodný výber $X_1, \ldots, X_n$ z P. Uvažujme hĺbku funkcií náhodného výberu voči P (resp. P_n)

$$D : \mathcal{C}([0,1]) \times \mathcal{P}(\mathcal{C}([0,1])) \to [0,1].$$



Silná konzistencia

Hĺbka D je na množine $S \subset \mathcal{C}([0,1])$ **konzistentná**

- **bodove** ak

$$D\left(x;P_n\right) - D\left(x;P\right) \xrightarrow[n\to\infty]{s.i.} 0 \text{ pre každé } x \in S,$$

Silná konzistencia

Hĺbka D je na množine $S \subset \mathcal{C}([0,1])$ **konzistentná**

- **bodove** ak

$$D(x;P_n) - D(x;P) \xrightarrow[n\to\infty]{s.i.} 0 \text{ pre každé } x \in S,$$

- **rovnomerne** ak

$$\sup_{x\in S} \left| D(x;P_n) - D(x;P) \right| \xrightarrow[n\to\infty]{s.i.} 0,$$

Silná konzistencia

Hĺbka D je na množine $S \subset \mathcal{C}([0,1])$ **konzistentná**

- **bodove** ak

$$D(x;P_n) - D(x;P) \xrightarrow[n\to\infty]{s.i.} 0 \text{ pre každé } x \in S,$$

- **rovnomerne** ak

$$\sup_{x\in S} \left|D(x;P_n) - D(x;P)\right| \xrightarrow[n\to\infty]{s.i.} 0,$$

- **univerzálne** ak

$$\sup_{x\in S} \left|D(x;P_n) - D(x;P)\right| \xrightarrow[n\to\infty]{s.i.} 0 \text{ pre každé } P \in \mathcal{P}(\mathcal{C}([0,1])),$$

Silná konzistencia

Hĺbka D je na množine $S \subset \mathcal{C}([0,1])$ **konzistentná**

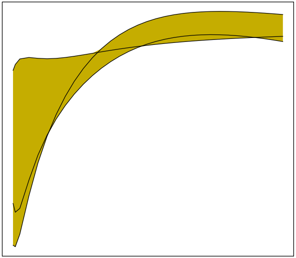
- **bodove** ak
$$D(x;P_n) - D(x;P) \xrightarrow[n\to\infty]{s.i.} 0 \text{ pre každé } x \in S,$$
- **rovnomerne** ak
$$\sup_{x\in S} |D(x;P_n) - D(x;P)| \xrightarrow[n\to\infty]{s.i.} 0,$$
- **univerzálne** ak
$$\sup_{x\in S} |D(x;P_n) - D(x;P)| \xrightarrow[n\to\infty]{s.i.} 0 \text{ pre každé } P \in \mathcal{P}(\mathcal{C}([0,1])),$$
- **uniformne** ak
$$\sup_{P\in\mathcal{P}(\mathcal{C}([0,1]))} \sup_{x\in S} |D(x;P_n) - D(x;P)| \xrightarrow[n\to\infty]{s.i.} 0.$$

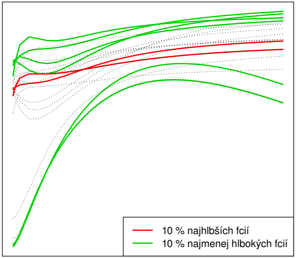
Pásová hĺbka

López-Pintado a Romo [8] pre $J = 2, 3, \ldots$

$$BD^{J)}(x; P) = \frac{1}{J-1} \sum_{j=2}^{J} P\left[G(x) \subset B\left(X_1, X_2, \ldots, X_j\right)\right],$$

kde $G(x)$ je **graf funkcie** x a $B(x_1, x_2, \ldots, x_j)$ je **pás funkcií** $x_1, x_2, \ldots, x_j$

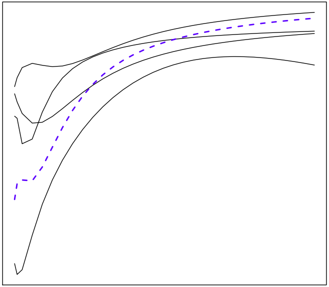




Pásová hĺbka

Výberová verzia je **U-štatistika rádu J**.

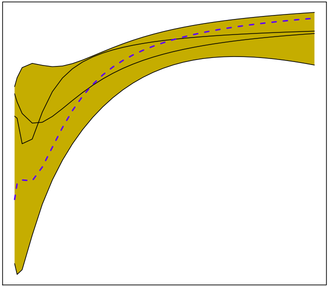
$$BD^{J)}(x;P_n) = \frac{1}{J-1}\sum_{j=2}^{J}\binom{n}{j}^{-1}\sum_{1\leq i_1<i_2<\ldots<i_j\leq n}\mathbb{I}\left[G(x)\subset B\left(X_{i_1},X_{i_2},\ldots,X_{i_j}\right)\right].$$



Pásová hĺbka

Výberová verzia je **U-štatistika rádu J**.

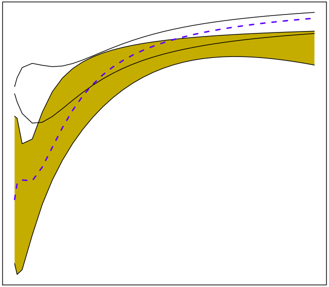
$$BD^{J)}\left(x;P_n\right)=\frac{1}{J-1}\sum_{j=2}^{J}\binom{n}{j}^{-1}\sum_{1\leq i_1<i_2<\ldots<i_j\leq n}\mathbb{I}\left[G(x)\subset B\left(X_{i_1},X_{i_2},\ldots,X_{i_j}\right)\right].$$



Pásová hĺbka

Výberová verzia je **U-štatistika rádu J**.

$$BD^{J)}(x; P_n) = \frac{1}{J-1} \sum_{j=2}^{J} \binom{n}{j}^{-1} \sum_{1 \le i_1 < i_2 < \ldots < i_j \le n} \mathbb{I}\left[G(x) \subset B\left(X_{i_1}, X_{i_2}, \ldots, X_{i_j}\right)\right].$$



Pásová hĺbka

Pásová hĺbka (L-P López-Pintado, R Romo):

- L-P, R: Depth-based classification for functional data (DIMACS 2006)
- L-P, R: Depth-based inference for functional data (CSDA 2007)
- L-P, Jornsten: Functional analysis via extensions of the band depth (IMS Lecture Notes, 2007)
- **L-P, R: On the Concept of Depth for Functional Data (JASA 2009)**
- L-P, R: Robust depth-based tools for the analysis of gene expression data (Biostatistics 2010)
- L-P, R: A half-region depth for functional data (CSDA 2011)
- ...

Konzistencia pásovej hĺbky

López-Pintado a Romo [8, Thm 4]

Tvrdenie:

Nech P je rozdelenie na $\mathcal{C}([0,1])$ s absolútne spojitými marginálnymi rozdeleniami. Potom $BD^{J)}$ je rovnomerne konzistentná na každej rovnako spojitej množine S, t.j.

$$\sup_{x \in S} \left| BD^{J)}(x;P_n) - BD^{J)}(x;P) \right| \xrightarrow[n\to\infty]{s.i.} 0.$$

Konzistencia pásovej hĺbky: Dôkaz

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.

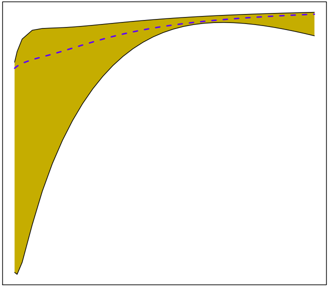
Pretože $BD^{J)}(.;P)$ je pre P s absolútne spojitými marginálnymi rozdeleniami **spojitý funkcionál, stačí dokázať** pre $N \in \mathbb{N}$ pevné

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

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

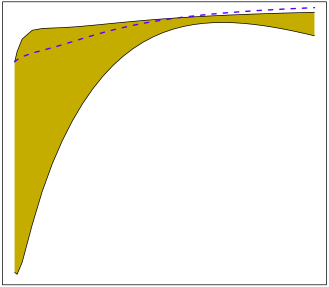
Prečo dôkaz nefunguje?

$BD^{J)}(.; P)$ je spojitá, ale $BD^{J)}(.; P_n)$ **nie je**!



Prečo dôkaz nefunguje?

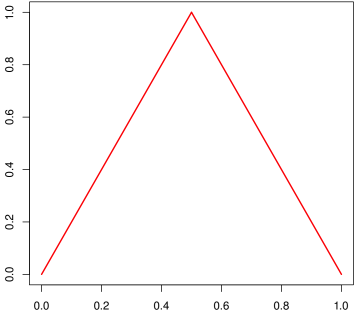
$BD^{J)}(.;P)$ je spojitá, ale $BD^{J)}(.;P_n)$ **nie je**!



Prečo dôkaz nefunguje?

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

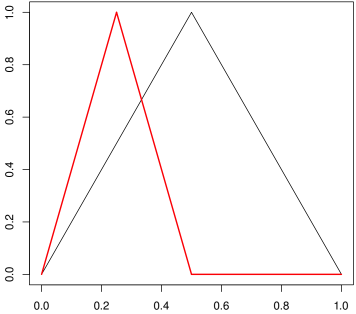
navyše **nedáva rovnomernú konzistenciu**!



Prečo dôkaz nefunguje?

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

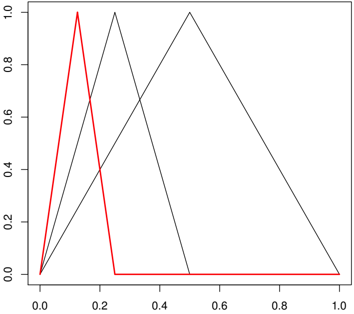
navyše **nedáva rovnomernú konzistenciu**!



Prečo dôkaz nefunguje?

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

navyše **nedáva rovnomernú konzistenciu**!



Je pásová hĺbka konzistentná?

Vyjdime z teórie **empirických procesov** (BÚNO $J = 2$):

- Platnosť

$$\dim_{\mathrm{VC}} \{(x_1,x_2)|G(x) \subset B(x_1,x_2)\}_{x\in S} = \infty$$

  pre kompaktné množiny $S \subset \mathcal{C}([0,1])$ naznačuje, že hĺbka **nebude uniformne konzistentá** (Assouadova veta - [4, Thm 6.4.5]).

- Existencia **boolean σ-nezávislej postupnosti** funkcií v triede

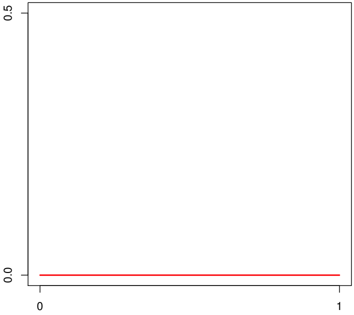
$$\{(x_1,x_2)|G(x) \subset B(x_1,x_2)\}_{x\in S}$$

  naznačuje, že hĺbka **nebude univerzálne konzistentá** (van Handelova veta - [9, Thm 1.3]).

Konzistencia pásovej hĺbky: Protipríklad

Definujme $X \sim P \in \mathcal{P}(\mathcal{C}([0,1]))$ takto:

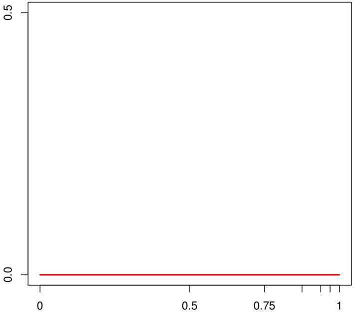
- $P(X(t) = 0$ pre všetky $t \in [0,1]) = 0.5$.



Konzistencia pásovej hĺbky: Protipríklad

Definujme $X \sim P \in \mathcal{P}(\mathcal{C}([0,1]))$ takto:

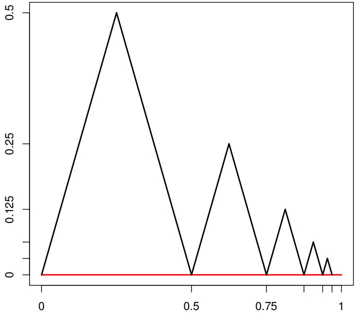
- Rozdeľme interval $[0,1]$ „diadicky“ na disjunktné subintervaly I_j dĺžok $\left\{2^{-j}\right\}_{j\in\mathbb{N}}$.



Konzistencia pásovej hĺbky: Protipríklad

Definujme $X \sim P \in \mathcal{P}(\mathcal{C}([0,1]))$ takto:

- Ak $X \not\equiv 0$, na každom I_j bude X nulová s pst'ou 0.5 alebo nadobudne skok s pst'ou 0.5. Skoky nastávajú nezávisle od seba.



Konzistencia pásovej hĺbky: Protipríklad

Označme ako x_j funkciu s **jediným skokom** práve na intervale I_j, inde 0. Potom platí:

- $BD^{2)}(x_j; P) = 0.25$ pre každé $j \in \mathbb{N}$

Konzistencia pásovej hĺbky: Protipríklad

Označme ako x_j funkciu s **jediným skokom** práve na intervale I_j, inde 0. Potom platí:

- $BD^{2)}(x_j; P) = 0.25$ pre každé $j \in \mathbb{N}$
- Nech n je párne. Ak existuje $j_n \in \mathbb{N}$ tak, že práve $n/2$ funkcií má skok v I_{j_n} a $n/2$ funkcií n.v. je nulových na $[0, 1]$, potom x_{j_n} leží v

$$\binom{n}{2} - 2\binom{n/2}{2} = \frac{n^2}{4}$$

pásoch.

Konzistencia pásovej hĺbky: Protipríklad

Označme ako x_j funkciu s **jediným skokom** práve na intervale I_j, inde 0. Potom platí:

- $BD^{2)}(x_j; P) = 0.25$ pre každé $j \in \mathbb{N}$
- Nech n je párne. Ak existuje $j_n \in \mathbb{N}$ tak, že práve $n/2$ funkcií má skok v I_{j_n} a $n/2$ funkcií n.v. je nulových na $[0,1]$, potom x_{j_n} leží v

$$\binom{n}{2} - 2\binom{n/2}{2} = \frac{n^2}{4}$$

pásoch.

- Pre takéto j_n teda platí

$$BD^{J)}(x_{j_n}; P_n) = \frac{\frac{n^2}{4}}{\binom{n}{2}} = \frac{n}{2(n-1)} \xrightarrow[n\to\infty]{} 0.5.$$

Konzistencia pásovej hĺbky: Protipríklad

Označme ako x_j funkciu s **jediným skokom** práve na intervale I_j, inde 0. Potom platí:

- $BD^{2)}(x_j; P) = 0.25$ pre každé $j \in \mathbb{N}$
- Nech n je párne. Ak existuje $j_n \in \mathbb{N}$ tak, že práve $n/2$ funkcií má skok v I_{j_n} a $n/2$ funkcií n.v. je nulových na $[0, 1]$, potom x_{j_n} leží v

$$\binom{n}{2} - 2\binom{n/2}{2} = \frac{n^2}{4}$$

pásoch.
- Pre takéto j_n teda platí

$$BD^{J)}(x_{j_n}; P_n) = \frac{\frac{n^2}{4}}{\binom{n}{2}} = \frac{n}{2(n-1)} \xrightarrow[n\to\infty]{} 0.5.$$

Konzistencia pásovej hĺbky: Protipríklad

Označme ako x_j funkciu s **jediným skokom** práve na intervale I_j, inde 0. Potom platí:

- $BD^{2)}(x_j;P) = 0.25$ pre každé $j \in \mathbb{N}$
- Nech n je párne. Ak existuje $j_n \in \mathbb{N}$ tak, že práve $n/2$ funkcií má skok v I_{j_n} a $n/2$ funkcií n.v. je nulových na $[0,1]$, potom x_{j_n} leží v
$$\binom{n}{2} - 2\binom{n/2}{2} = \frac{n^2}{4}$$
pásoch.
- Pre takéto j_n teda platí
$$BD^{J)}(x_{j_n};P_n) = \frac{\frac{n^2}{4}}{\binom{n}{2}} = \frac{n}{2(n-1)} \xrightarrow[n\to\infty]{} 0.5.$$

Existuje ale **nekonečne veľa** takých dvojíc (n, j_n)?

Konzistencia pásovej hĺbky: Protipríklad

Existuje ale nekonečne veľa takých dvojíc (n, j_n)? **Áno!**

- Skoro iste existuje nekonečne veľa n takých, že práve $n/2$ funkcií bude nulových na $[0, 1]$ **(stav 0 v symetrickej náhodnej prechádzke je trvalý)**.

Konzistencia pásovej hĺbky: Protipríklad

Existuje ale nekonečne veľa takých dvojíc (n, j_n)? **Áno!**

- Skoro iste existuje nekonečne veľa n takých, že práve $n/2$ funkcií bude nulových na $[0, 1]$ **(stav 0 v symetrickej náhodnej prechádzke je trvalý)**.
- Pre každé takéto n s.i. existuje j_n tak, že všetkých $n/2$ funkcií so skokmi bude mať v intervale I_{j_n} skok **(Borel-Cantelli)**.

Konzistencia pásovej hĺbky: Protipríklad

Existuje ale nekonečne veľa takých dvojíc (n, j_n)? **Áno!**

- Skoro iste existuje nekonečne veľa n takých, že práve $n/2$ funkcií bude nulových na $[0,1]$ **(stav 0 v symetrickej náhodnej prechádzke je trvalý)**.
- Pre každé takéto n s.i. existuje j_n tak, že všetkých $n/2$ funkcií so skokmi bude mať v intervale I_{j_n} skok **(Borel-Cantelli)**.
- Dostávame teda nekonečne veľa n, pričom pre každé existuje funkcia z $\{x_j\}_{j\in\mathbb{N}}$, pre ktorú $BD^{2)}(x_{j_n}; P_n) \approx 0.5$. Preto pre každé $\varepsilon > 0$ a pre nekonečne veľa $n \in \mathbb{N}$ s.i. platí

$$\sup_{j\in\mathbb{N}} \left| BD^{2)}(x_j; P_n) - BD^{2)}(x_j; P) \right| > 0.25 - \varepsilon$$

Konzistencia pásovej hĺbky: Protipríklad

Existuje ale nekonečne veľa takých dvojíc (n, j_n)? **Áno!**

- Skoro iste existuje nekonečne veľa n takých, že práve $n/2$ funkcií bude nulových na $[0,1]$ **(stav 0 v symetrickej náhodnej prechádzke je trvalý)**.
- Pre každé takéto n s.i. existuje j_n tak, že všetkých $n/2$ funkcií so skokmi bude mať v intervale I_{j_n} skok **(Borel-Cantelli)**.
- Dostávame teda nekonečne veľa n, pričom pre každé existuje funkcia z $\{x_j\}_{j\in\mathbb{N}}$, pre ktorú $BD^{2)}(x_{j_n}; P_n) \approx 0.5$. Preto pre každé $\varepsilon > 0$ a pre nekonečne veľa $n \in \mathbb{N}$ s.i. platí

$$\sup_{j\in\mathbb{N}} \left| BD^{2)}(x_j; P_n) - BD^{2)}(x_j; P) \right| > 0.25 - \varepsilon$$

Konzistencia pásovej hĺbky: Protipríklad

Existuje ale nekonečne veľa takých dvojíc (n, j_n)? **Áno!**

- Skoro iste existuje nekonečne veľa n takých, že práve $n/2$ funkcií bude nulových na $[0,1]$ **(stav 0 v symetrickej náhodnej prechádzke je trvalý)**.
- Pre každé takéto n s.i. existuje j_n tak, že všetkých $n/2$ funkcií so skokmi bude mať v intervale I_{j_n} skok **(Borel-Cantelli)**.
- Dostávame teda nekonečne veľa n, pričom pre každé existuje funkcia z $\{x_j\}_{j\in\mathbb{N}}$, pre ktorú $BD^{2)}(x_{j_n}; P_n) \approx 0.5$. Preto pre každé $\varepsilon > 0$ a pre nekonečne veľa $n \in \mathbb{N}$ s.i. platí

$$\sup_{j\in\mathbb{N}} \left| BD^{2)}(x_j; P_n) - BD^{2)}(x_j; P) \right| > 0.25 - \varepsilon$$

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

Oprava prvá: zospojitenie

Problém dôkazu López-Pintadovej a Roma bol v nespojitosti $BD^{J)}(.;P_n)$. Nemerajme teda mieru náležania do pásu indikátorom, ale **vzdialenosťou od pásu**, t.j. pre metriku d na $\mathcal{C}([0,1])$ namiesto

$$P\left[G(x) \subset B\left(X_1, X_2\right)\right] = \mathrm{E}\left[\mathbb{I}\left[G(x) \subset B\left(X_1, X_2\right)\right]\right]$$

v definícii hĺbky používajme

$$\mathrm{E}\left[1 - w\left(d\left(x; B\left(X_1, X_2\right)\right)\right)\right],$$

kde $w : [0,\infty) \to [0,1]$, $w(0) = 1$, $\lim_{t\to\infty} w(t) = 0$ je rovnako spojitá **zhladzujúca funkcia**, napr. e^{-t}.
Obmedzíme sa pre jednoduchosť na suprémovú a L_1 metriku.

Oprava prvá: zospojitenie

Tvrdenie:

Nech w je zhladzujúca funkcia a $S \subset \mathcal{C}([0,1])$ je relatívne kompaktná. Potom pásové hĺbky zhladené w

$$BD^{J)}(.;.,w,d) : \mathcal{C}([0,1]) \times \mathcal{P}(\mathcal{C}([0,1])) \to [0,1]$$

sú tak pre suprémovú, ako pre L_1 metriku uniformne konzistentné na S.

Dôkaz: Zosilnená verzia dôkazu López-Pintadovej a Roma obohatená o rovnakú spojitosť triedy

$$\left\{ BD^{J)}(x;P,w,d) \middle| x \in \mathcal{C}([0,1]), P \in \mathcal{P}(\mathcal{C}([0,1])) \right\}$$

a vlastnosti kanonických U-štatistík (Borovskich a Koroljuk [7, Thm 2.1.4]).

Fraimanova-Munizovej hĺbka

Fraiman a Munizová [5]

$$ID(x;P) = \int_0^1 D(x(t);P_t)\,\mathrm{d}t,$$

kde D je jednorozmerná „hĺbka“ ako

- **polopriestorová hĺbka**

$$D(x(t);P_t) = \min\{F_t(x(t)), 1 - F_t(x(t))\},$$

- **simplexová hĺbka**

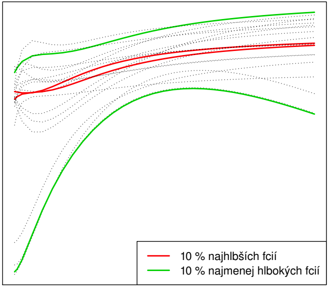
$$D(x(t);P_t) = F_t(x(t))(1 - F_t(x(t))).$$

Na rovnakom princípe fungujú K**-pásové hĺbky** (Hlubinka a Nagy [6]) a **duálne hĺbky** (Cuevas a Fraiman [2]).

Fraimanova-Munizovej hĺbka

Fraiman a Munizová [5]

$$ID\left(x;P\right) = \int_0^1 D\left(x(t);P_t\right)\,\mathrm{d}t,$$



Konzistencia integrálnych hĺbok

Tvrdenie:

Nech výberová verzia jednorozmernej hĺbky $D : \mathbb{R} \times \mathcal{P}(\mathbb{R}) \to [0,1]$ *má tvar U-štatistiky a D je univerzálne konzistentná. Potom funkcionálna hĺbka*

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

je univerzálne konzistentná na $\mathcal{C}([0,1])$.

Dôkaz: Použitím Lebesguovej vety dostávame slabú univerzálnu konzistenciu, ktorá je pre U-procesy ekvivalentná (silnej) univerzálnej konzistencii (de la Peña a Giné [3, p.227]). □

Predpoklady tvrdenia sú splnené napr. voľbou **simplexovej hĺbky** za D.

Záver pre pásové hĺbky

Pásové hĺbky:

- dávajú **zlé výsledky v aplikáciách** (Hlubinka a Nagy [6]),
- **ťažko sa počítajú** ($\mathcal{O}\left(n^J\right)$ proti $\mathcal{O}(n)$ pre integrálne hĺbky),
- **nie sú konzistentné**.

Záver

Nepoužívať pásové hĺbky a voliť integrálne alternatívy.

Vyhliadky do budúcnosti

- Rozšírenie van Handelovej (Assouadovej) vety na U-procesy.
- **Uniformná konzistencia integrálnych hĺbok.**

$$P_\gamma - \dim \{\lambda [t | x(t) \in B(x_1(t), x_2(t))]\}_{x \in S} = \infty \quad \forall \gamma > 0$$

pre kompaktné množiny $S \subset \mathcal{C}([0,1])$ naznačuje, že hĺbka **nebude uniformne konzistentá** (Alonova veta [1, Thm 2.2]).

Literatúra

- Noga Alon, Shai Ben-David, Nicolò Cesa-Bianchi, and David Haussler. Scale-sensitive dimensions, uniform convergence, and learnability. *J. ACM*, 44(4):615–631, 1997.
- Antonio Cuevas and Ricardo Fraiman. On depth measures and dual statistics. a methodology for dealing with general data. *J. Multivar. Anal.*, 100:753–766, April 2009.
- Víctor H. de la Peña and Evarist Giné. *Decoupling: From Dependence to Independence (Probability and its Applications)*. Springer, 1998.
- R.M. Dudley. *Uniform Central Limit Theorems*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1999.
- Ricardo Fraiman and Graciela Muniz. Trimmed means for functional data. *Test*, 10(2):419–440, 2001.
- Daniel Hlubinka and Stanislav Nagy. Functional data depth and classification. *submitted to Comput. Stat.*
- V. S. Koroljuk and Yu. V. Borovskich. *Theory of U-Statistics*. Kluwer, 1994.
- 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.
- Ramon van Handel. The universal Glivenko-Cantelli property. *Probab. Th. Rel. Fields, to appear.*