

POSUDOK OPONENTA BAKALÁRSKEJ PRÁCE

Názov: Teoretické a empirické kvantily a ich využitie pri konštrukcii predikčných intervalov
Autor: Jakub Šimičák

ZHRNUTIE OBSAHU PRÁCE

Cieľom predloženej bakalárskej práce bolo oboznámiť sa s kvantilmi a ich využitím pri konštrukcii tzv. predikčných intervalov.

V kapitole 1 sú definované (empirické) kvantily a je diskutovaný význam predikčných intervalov. Kapitola 2 predstavuje metódy konštrukcie presných predikčných intervalov pomocou pivotálnych kvantíl, a asymptotickej metódy popísanej v článku Lawless a Fredette (2005). V kapitole 3 je zavedený konformný predikčný interval. Záverečná kapitola 4 obsahuje malú simulačnú štúdiu.

CELKOVÉ HODNOTENIE PRÁCE

Téma práce. Nosné témy práce sú dve: (a) predstavenie (empirických) kvantilov, a (b) konštrukcia predikčných intervalov. Z týchto dvoch tém, iba druhá presahuje obsah základných kurzov bakalárskeho štúdia na KPMS MFF.

Vlastný príspevok. Autor prezentuje drobnú simulačnú štúdiu v kapitole 4 práce.

Matematická úroveň. Práca obsahuje niekoľko definícií. Jediné matematické odvodenia v práci nájdeme v príkladoch 2 a 3. Príklad 3 sa však zdá byť z matematického hľadiska problematický, viz body 2 a 3 nižšie.

Práca so zdrojmi. Zoznam literatúry uvádza štyri položky:

- poznámky k prednáške *Matematická statistika I*,
- odborné články Lawless a Fredette (2005, ďalej [LF]) a
- Shafer a Vovk (2008, [SV]), a
- pdf súbor prezentácie Tibshiraniho (2009, [T]).

Z článku [LF] má vychádzať sekcia 2.2 práce. Podľa všetkého je však v sekcii 2.2 uvedená iba značne naivná “plug-in” metóda spomenutá v [LF] iba okrajovo. Článok [SV] je citovaný iba na str. 17 práce. Tam je uvedené, že v [SV] je dokázané, že zameniteľnosť náhodných veličín implikuje ich rovnaké rozdelenie. To ale plynie triviálne z definície zameniteľnosti. Odvodenie konformného predikčného intervalu je podľa všetkého založené iba na dvoch stránkach prezentácie [T], viz obr. 1 nižšie. Táto referencia je z textu niekoľkokrát nesprávne odkazovaná ako článok [T].

Formálna úprava. Formálna úprava práce je pomerne slabá. Text obsahuje veľké množstvo gramatických chýb (už napr. v samotnom názve práce má byť *konštrukcii* namiesto *konštrukcií*). Ďalšie chyby uvádzam s pripomienkami nižšie.

Starting simple (cont.)

The trick is to use **symmetry**. By exchangeability, the rank of Y_{n+1} among Y_1, \dots, Y_{n+1} is discrete uniform over $\{1, \dots, n+1\}$, i.e., let

$$\pi(y) = \frac{1}{n+1} \sum_{i=1}^n 1\{Y_i \leq y\} + \frac{1}{n+1}$$

then $\pi(Y_{n+1})$ is discrete uniform over $\{1/(n+1), \dots, 1\}$. Thus

$$\mathbb{P}\left(\pi(Y_{n+1}) \leq \frac{[(n+1)(1-\alpha)]}{n+1}\right) \geq 1-\alpha$$

and we can **invert this** to get

$$\mathbb{P}(Y_{n+1} \leq \hat{q}_n) \geq 1-\alpha$$

for $\hat{q}_n = Y_{[(n+1)(1-\alpha)]} = \text{Quantile}(1-\alpha; \{Y_i\}_{i=1}^n \cup \{\infty\})$

3

Obr. 1: Referencia [T] použitá ako základ kapitoly 3 práce.

ŠPECIFICKÉ PRIPOMIENKY

1. Zdroje práce [LF] a [SV] uvádzajú celú radu metód konštrukcie predikčných intervalov. V sekcii 2.2 práce a príklade 3 sú predikčné intervaly definované pomocou *priamej plug-in metódy*, tj. kvantilov rozdelenia $F_W(\cdot, \hat{\theta}_n)$, kde W je typicky priamo predikovaná veličina Y (viz príklad 3), F_W je jej distribučná funkcia, a $\hat{\theta}_n$ je konzistentný odhad neznámeho parametru θ . Takýto prístup úplne ignoruje fakt, že θ je odhadované. Napriek tomu, že článok [LF] je citovaný v sekcii 2.2, priama plug-in metóda nie je hlavná metóda popisovaná v [LF]; v [LF] o tejto metóde autori skôr odrádzajú. Hlavným príspevkom [LF] je návrh metódy kalibrácie naivného priameho plug-in intervalu. Z tohto dôvodu sa mi nezdá byť korektné citovať [LF] ako (jediný) zdroj pre priamu plug-in metódu.
2. Autor by mal vysvetliť tvrdenie zo str. 11 práce:

“Náhodná veličina \widehat{W}_n nezávisí na parametri θ , čo znamená, že môže byť použitá ako pivotálna štatistika pre konštrukciu predikčného intervalu.”

Ďalej, nikde v článku [LF] nevidím dôkaz tvrdenia o konvergencii náhodných veličín \widehat{W}_n v distribúcii tak ako je tvrdené na str. 11 práce.

3. Čo presne znamená výraz zo str. 12

$$\widehat{W}_n \sim \text{Exp}\left(\frac{1}{\overline{X}_n}\right),$$

ak \overline{X}_n je náhodná veličina? Podľa čoho vieme, že “rozdelenie \widehat{W}_n nezávisí na parametri $\lambda > 0$ ”, ako je uvedené na str. 12?

4. Ako presne získavame dôležitú formulu (3.3) na str. 15?
5. Mohol by autor vysvetliť, prečo je pri konštrukcii konformného predikčného intervalu nutné položiť $y = \infty$, ako je uvedené na str. 15?

6. Ako je možné, že je veličina $\widehat{W}_n \sim N(\overline{X}_n, S_n^2)$ na str. 19 asymptoticky pivotálna, ak jej rozdelenie zrejme závisí na neznámych parametroch μ a σ^2 , ktoré \overline{X}_n a S_n^2 odhadujú? Podobne na str. 20 a inde.

GRAMATICKÉ A FORMÁLNE CHYBY

1. Slová ako *konštrukcia*, *funkcia*, alebo *štúdia* sa píše v piatom páde jednotného čísla (*o*) *konštrukcii*, *funkcii*, alebo *štúdiu*, a nie (*o*) *konštrukcií* atď. Túto chybu nájdeme v práci prakticky na každej strane.
2. Názov práce uvedený na titulnej strane je iný ako názov práce uvedený na str. iv (strana pred *Obsahom*).
3. *Náhodný výber* je anglicky *random sample*, nie *random selection* (abstrakt).
4. str. 2: *prevdepodobnostný*; str. 8: W po formuli (2.1) má byť \overline{W} , str. 9: v príklade 2 má byť $\Theta = \mathbb{R} \times (0, \infty)$; str. 15 a inde: po formuli (3.3) nemá byť text odsadený doprava; str. 15: *náhodný \mathcal{X}_n výber*; str. 17: *intervalu 3.3*; str. 19: *tvary predikčných intervaly*; str. 19: *parametrami \mathcal{X}_n* ; str. 20: píše sa *Raphsonovym*, nie *Rhapsonovým*; str. 21: *že v pre teoretické*; obr. 4.1: *frekventistický*; a množstvo ďalších.

ZÁVER

Podľa môjho názoru sa jedná o pomerne slabú prácu. V hlavnej časti sú, bez korektného odvodenia, uvedené dve možnosti konštrukcie predikčných intervalov. Tieto však prakticky nevychádzajú z citovanej literatúry. To, či je prácu možné uznať ako bakalársku prácu na MFF UK nechávam na zvážení komisie.



Stanislav Nagy
KPMS MFF UK
30. mája 2023