

Oponentský posudek na bakalářskou práci:

LUKÁŠ PAVLÍK: TESTY NEZÁVISLOSTI V KONTINGENČNÍ TABULCE

Předložená práce velice přehledně a srozumitelně představuje problém testování nezávislosti v kontingenční tabulce. Uchazeč se kromě nejoblíbenějších testů nezávislosti ( $\chi^2$  test, G test a Fisherův přesný test) soustřeďuje na U-statistic permutation (USP) test s kritickými hodnotami vypočítanými pomocí permutačního přístupu. Ve třetí kapitole je stručně představená teorie U-statistik i s vlastním důkazem Lemma 1. Práci uzavírá simulační studie, která jasně popisuje situace, ve kterých USP test funguje lépe nebo hůře, než ostatní zkoumané testy (zdá se přitom, že pozorované rozdíly jsou způsobené hlavně odlišným vážením rozdílů mezi pozorovanými a očekávanými četnostmi).

Otázky k obsahu práce:

1. V úvodu práce se předpokládá pevný počet ( $n$ ) pozorování. Má smysl uvažovat hypotézu nezávislosti i v případě, že je počet pozorování náhodný nebo naopak v případě, že řádkové četnosti  $n_{i+}$  nejsou náhodné (jedná se o předem zvolené konstanty)?
2. Bylo by možné zvolit jako testovou statistiku jenom první člen výrazu pro  $\widehat{D}$  na str. 15?
3. Neměla by se předpokládat symetrie jádra v Lemma 1 a symetrie funkce  $S$  ve Větě 2?
4. Proč je na straně 20 nutné perturbovat nejméně čtyři prvky a jak se maximální počet perturbací liší pro sudé a liché  $I$  a  $J$ ? Je nezbytné používat v simulační studii pouze perturbace, které nemění marginální rozdělení?

Další otázky a připomínky:

**str. 3 dole** Co přesně zde znamená symbol  $\binom{X}{Y}$ ?

**str. 6** Jedná se o data simulovaná stejným způsobem jako v kapitole 4?

**str. 8, vzorec (2.2)** Může statistika  $G$  nabývat i záporných hodnot?

**str. 11, alg. 1, krok 5** Jak a proč se zde vybírají permutace bez opakování?

**str. 11** Nevadí, že kritická hodnota  $c_\alpha$  závisí na testové statistice  $T_0$ ?

Práce je zpracovaná velice přehledně a čitelně, všechny pojmy jsou dostatečně vysvětlené a i grafická úprava je vzorná (snad jen v obrázcích by u písmen 'ě' a 'ů' nemusely chybět háčky a kroužky).

**Závěr:** Práce přehledně a matematicky korektně zpracovává zadané téma. Kvalitně zpracovaná simulační studie dochází k jasným doporučením, kdy je výhodné použít zkoumaný USP test místo obvyklého  $\chi^2$  testu. Předloženou práci proto doporučuji uznat jako práci bakalářskou.

V Praze 5. června 2023

Doc. RNDr. Zdeněk Hlávka, Ph.D.  
KPMS MFF UK  
Sokolovská 83, 186 75 Praha 8