

*Oponentský posudek dizertační práce*

Název dizertační práce: **Normality test of the gene expression data**

Autor: Bobosharif K. Shokirov

Dizertační práce B.K. Shokirova se týká testů normality dat genové exprese. V podstatě jsou v práci navrženy a studovány dva testy normality pro mnohorozměrná data s dimenzí větší než počet pozorování a za speciálních předpokladů týkající se závislostní struktury. Opodstanění těchto předpokladů pro případ dat genových expresí je motivováno pracemi L. Klebanova, A. Yakovleva a dalších.

Práce je rozdělena do 4 kapitol. První slouží jako úvod do problematiky, vysvětluje pojem genová exprese a související biologické procesy. Pro statistický obsah práce jsou velice důležité podseky 1.4 a 1.5 týkající se závislosti mezi genovými expresemi. V kapitole 2 se autor nejprve zabývá problematikou rekonstrukce typu distribuční funkce a pak shrnuje známé charakterizační věty pro normální rozdělení potřebné v dalším. Autor zde mimo jiné ukazuje, že charakterizační věta dokázaná Sakatou (1977) neplatí pro  $k = 2$ .

V kapitole 3 je navržen test normality založený na tzv. testu rovnoměrnosti transformovaných dat na sféře. Zde jsou využity výsledky z předchozí kapitoly a rovněž výsledky L. Klebanova a dalších o testech založených na tzv.  $\mathfrak{N}$ -vzdálenostech. Kapitola obsahuje jak teoretické výsledky, tak výpočetní algoritmus a aplikaci na reálná data.

Čtvrtá kapitola se týká jiného testu normality. Jsou dokázány nové výsledky týkající se rekonstrukce kompletního typu rozdělení a nové charakterizační věty vhodné pro konstrukci testu. Nakonec je na vhodně transformovanou aplikovanou Kolmorovův-Smirnovův test, s tím že hypotetické rozdělení transformovaných dat je Meijerova G- typu. V kapitole je vypracován výpočetní algoritmus a provedena aplikace na reálná data.

Zásadní nové výsledky jsou v kapitolách 3 a 4, kde jsou dokázány nové věty týkající se tzv. rekonstrukce distribučních funkcí a charakterizační vět. To je využito ke konstrukci testů normality vhodných pro data genové exprese. Přínos je jak teoretický tak aplikační.

Velmi dobře je třeba hodnotit nalezení vhodných transformací dat a tím převedení původní úlohy testu o normalitě rozdělení dat na ekvivalentní úlohu testu o rozdělení dat transformovaných, na které už lze provést standardní test. Podstatné je, že jde o ekvivalentní úlohy, přičemž pro druhou je dostupný rozumný test.

Připomínky závažnějšího rázu. Vzhledem k tomu, že práce je obhajovaná v komisi pro pravděpodobnost a matematickou statistiku, je vhodné doplnit několik částí (vypracovat písemně před obhajobou a též zahrnout do prezentace):

- Přesně zformulovat model a uvažovanou hypotézu. Při podrobném čtení podseky 1.4 lze na základě citovaných prací Klebanova et al. předpokládat určitou závislostní strukturu. Z toho se v práci vychází, je to však ztraceno v úvodním textu a není již důkladněji zformulováno v kapitolách 3 a 4. Asi by někde měly být zformulovány základní předpoklady (nejen povídání o nich), formulace nulové hypotézy i aspoň některé alternativy.

- Vzhledem k tomu, že získat teoretické výsledky o chování za alternativy je velmi obtížné, měly by být provedeny aspoň simulace jak za nulové hypotézy tak za nějaké alternativy a teprve pak provést aplikaci na data.
- Tabulky 3.1, 3.2, 4.1, 4.2 by měly být pečlivěji a podrobněji popsány. V současné podobě je popis značně neúplný.
- Není uveden použitý software.

Méně závažné dotazy a komentáře:

- U některých vět nejsou uvedeny úplné předpoklady. Autor by měl projít všechna tvrzení a předpoklady doplnit. Asi by někde měly být zformulovány základní předpoklady (nejen povídaní o nich), formulace nulové hypotézy i aspoň přibližně možné alternativy.
- V matici  $\mathbf{X}$  na str. 17 a matici  $\mathbf{X}$  používané v kapitolách 3 a 4 jsou prohozené významy  $n$  a  $m$ .
- Otázka je, jak jsou získány numericky aproximace pro  $p$ -hodnoty, např. je otázka aproximace pro  $p$  na str. 36. Pokud je použito vyjádření (3.6), jde o nekonečnou řadu.
- Str. 38– Algoritmus 1, v bodu 4 jsou definovány veličiny  $Z_{ij}$ , ale v bodu 5 jsou jen  $Z_i$ .
- Str. 38 v bodu 6 je provedena transformace pomocí matice  $\Sigma$ , která je nahodná, to musí pokazit vlastnosti procedury teoreticky. Asi by byla vhodná nějaká poznámka. Rovněž v algoritmech je první krok-uspořádání dle velikosti roztylu, ty však neznáme-asi by to rovněž chtělo malou poznámku.
- Str. 28– (2.9) obsahuje závažnou tiskovou chybu.

Práce obsahuje nové výsledky ze dvou oblastí: jednak tzv. charakterizační a rekonstrukční věty, jednak testy normality pro speciální modely. Připomínky základního charakteru v podstatě znamenají podrobné doplnění a dovysvětlení některých partií, což je jistě možné. Je možné připravit text, který se vloží do dizertace a také vloží do prezentace.

Prezentace výsledků je standardní. Doktorand prokázal určitou schopnost samostatné a tvořivé vědecké práce.

Pokud B. Shokirov doplní části uvedené v připomínkách závažnějšího rázu a zodpoví též dotazy méně závažného charakteru, pak se domnívám, že práce (včetně doplnění) splňuje požadavky kladené na dizertační práci ve studijním oboru Pravděpodobnost a matematická statistika a doporučuji ji za ni uznat.

Prof. RNDr. Marie Hušková, DrSc., oponentka  
KPMS MFF UK

10.8.2015