

Kateřina Malíková: Shapirův-Wilkův test normality
(Posudek bakalářské práce)

Shapirův-Wilkův test je dnes asi nejpoužívanějším testem normality. Předložená práce obsahuje odvození testové statistiky W i jejích některých vlastností. Statistika je založena na prvních dvou momentech pořádkových statistik uspořádaného výběru rozsahu n z normovaného normálního rozdělení, jejichž vyčíslení je technicky velmi náročné. Proto se část práce zabývá aproximacemi příslušných koeficientů. Rozdělení statistiky W za platnosti testované hypotézy závisí jen na rozsahu výběru n , nicméně i zde je třeba používat aproximace, tentokrát založená na simulacích velkého rozsahu. Přehled aproximací je uveden ve 4. kapitole. Součástí předložené práce jsou také vlastní výsledky simulací.

Autorka práce se ve svém výkladu do jisté míry drží postupu původního článku Shapiro a Wilk (1965) včetně například poněkud nestandardního zápisu $\mathbf{m}^T \mathbf{V}^{-1} \mathbf{V}^{-1} \mathbf{m}$ místo $\mathbf{m}^T \mathbf{V}^{-2} \mathbf{m}$. Při přebírání argumentace článku je nedůsledná. Např. autoři píší o statistice W , že číselník a jmenovatel v její definici odhadují **až na jistou konstantu** stejnou kvantitu. Zmínku o konstantách autorka práce pominula. Zavádějící je také tvrzení o kvalitě statistiky $\hat{\sigma}$ jako odhadu směrodatné odchylky σ . Někdy nedůsledné je používání termínů: kolísání mezi aproximace a odhad, zřejmě záměna třetího a čtvrtého momentu s šikmostí a špičatostí. Není zřejmé, proč jsou v první části práce Tvrzení a dál pak Věty. Tvrzení 4 není tvrzení, pouze je opakován definiční vztah.

V porovnání s použitou literaturou autorka práce navíc dopočítala integrály na str. 13. Samostatně pomocí simulací ověřila momenty statistiky W . Bohužel, součástí práce jsou jen výsledky simulací, nikoliv podrobněji dokumentovaný postup.

Autorka práce zadání, které je přiměřené pro studenta bakalářského zaměřeného finanční matematikou, splnila, byť s několika nepřesnostmi. Po formální stránce je práce vyhotovena pěkně, tiskových chyb je málo.

Autorka práce projevila schopnost pochopit nový problém, popsat jej a místy doplnit. Proto přes zjištěné nedostatky navrhuji, aby práce byla jako bakalářská práce uznána.

V Praze 9. 6. 2014

Karel Zvára

Jednotlivé poznámky:

- str. 2** Označení $D()$ není v textu téměř používáno, zpravidla je vynecháno.
- str. 8 nahore** Jde o nejlepší nestranný **lineární** odhad založený na pořádkových statistikách. Bez tohoto doplnění není tvrzení pravdivé, transformace na uspořádaný výběr není lineární.
- str. 8 (2.3)** Ve jmenovateli má být C , nikoliv C^2 .
- str. 8 pod (2.4)** Statistika b je odhadem pevného **násobku** směrodatné odchylky, podobně pro S^2 jako odhadu rozptylu.
- str. 9 nahore** Tvrzení 4 nedává smysl, opakuje pouze zavedení vektoru \mathbf{a} v (2.4).
- str. 9 a jinde** Není vhodné používat termíny *aproximace* a *odhad* jako synonyma. *Odhad* má ve statistice zavedený význam, různý od významu slova *aproximace*.
- str. 9** Zmínka o matici \mathbf{V} do důkazu Tvrzení 6 nepatří, nijak s tímto důkazem nesouvisí.
- str.10 nahore** Až dosud byla v textu číslována tvrzení, nyní je zde s následujícím číslem věta. Proč ta změna?
- str. 10** Až dosud byl pomocí závorek kolem indexů rozlišován náhodný výběr a uspořádaný výběr. Ve 3. kapitole si má čtenář závorky domýšlet?
- str. 11** Důsledek o nezávislosti plyne snadno z Věty 7?
- str. 11** Co jsou y_i na konci důkazu Věty 8? Dosáhne svého maxima testová statistika i pro $\eta = 0$? Může svého maxima dosáhnout testová statistika i jindy?
- str. 14** Vektor \mathbf{a}^* je zaveden jako řádkový, pracuje se s ním jako se sloupcovým.
- str. 14, 10. ř. zdola** Je třeba vyloučit také $i = n$ nebo se omezit na $1 < i \leq [n/2]$.
- str. 14, 5. ř. zdola** Hodnota C závisí na \mathbf{V} nejen pro „dlouhé výběry“.
- str. 14, poslední dva řádky** Této poznámce nerozumím.
- str. 15, uprostřed** Konstanty a_1^*, a_n^* si nejsou rovny (absolutní hodnoty či druhé mocniny ano).
- str. 18** Co znamená označení μ_3 a μ_4 . Jde o centrální momenty nebo, jak se píše v textu o šikmost a špičatost?
- str.19** Nešťastné vyjádření, statistika W nemá pro konečné n normální rozdělení.
- str. 21** Kopírování způsobilo, že hodnoty y_i nejsou uspořádaný.
- str. 22, 2. ř. zdola** Píše se *na rozdíl*, nikoliv *narozdíl*.