

*Posudek oponenta na diplomovou práci*

Odhad momentů při intervalovém cenzorování typu I.

*Bc. Matej Ďurčík*

Práce se zabývá odhady střední hodnoty a rozptylu při intervalovém cenzorování typu I. Po úvodní kapitole se ve druhé a třetí kapitole autor zabývá metodou dekonvoluce za předpokladu, že známe rozdělení (cenzorovacích) časů  $T_1$ . V závěru třetí kapitoly je pak zmíněna možnost přístupu pomocí dekonvoluce v případě neznámého rozdělení  $T_1$ . Ve čtvrté kapitole představuje autor metodu neparametrické maximální věrohodnosti a tuto metodu porovnává s modelem dekonvoluce v páté kapitole.

Práce má logickou strukturu. Z jazykového hlediska práce neobsahuje žádné zásadní množství překlepů, přestože je psaná v angličtině. Na druhou stranu je však občas z práce až příliš cítit, že angličtina není mateřský jazyk autora a některá spojení jsou nepřirozená a značně šroubovaná.

Zatímco formální stránka práce je slušná, s obsahem práce to je, alespoň dle mého názoru, výrazně horší. Pro metodu dekonvoluce jsou prezentovány teoretické výsledky o asymptotické normalitě momentů, které téměř okamžitě plynou ze „správné verze“ CLV. Nicméně autorovi se nepodařilo toto použití CLV nijak vysvětlit. Není vůbec jasné, zda autor použití CLV rozumí a zda ví, kterou CLV vlastně používá. Není také zřejmé, co autor myslí tím, když říká, že nějaká veličina konverguje k normálnímu rozdělení.

Většinu **druhé** (str. 10–15) a **třetí kapitoly** (str. 20–26) pak zabírají simulace, které mají ilustrovat asymptotickou normalitu odhadů. Nicméně toto ověření asymptotické normality se omezí na vykreslení histogramu odhadů a konstatování, že s rostoucí rozsahem výběru se střední hodnota a rozptyl odhadů blíží svým limitním protějškům. Domnívám se, že např. vykreslení asymptotického rozdělení do histogramu odhadů by bylo daleko informativnější než vypisování výběrových průměrů a rozptylů na 5 až 6 platných číslic. Značně rušivě také působí to, že autor si zřejmě udělal šablonu na popis jednotlivých simulačních situací a pouze „doplňuje“ jiná čísla, zatímco věty se opakují. Např. ve všech situacích lze nalézt následující věty „*Simulations of the estimator  $M_{X,n}$  gave us values, which are close to the real values. The simulation results support the theorem . . . . Simulations of the estimator  $S_{X,n}$  gave us values, which confirm our Theorem 2.2.2. .*“ Závěrem bych rád zmínil, že tyto simulace vlastně jen dokumentují asymptotickou normalitu (vhodně znormovaného) výběrového průměru a výběrového rozptylu. Navíc scénář i) v kapitole 3.2 je zbytečný, protože se shoduje, jak ostatně sám autor poznamenává, se scénářem i) kapitoly 2.3. Daleko zajímavější mohla být simulační studie pro odhady (3.25) a (3.26). Bohužel na to asi nezbyl sám, jelikož autor píše „*Because of lack of time we have not been able to perform simulations of these estimators*“.

Zatímco teorie popsaná ve druhé a třetí kapitole je spíše jednoduchá, **čtvrtá kapitola** obsahuje již velmi hluboké a pokročilé teoretické výsledky a postupy, kterým není snadné porozumět. Bohužel autorovi se nepodařilo tyto výsledky nijak přiblížit a vysvětlit čtenáři. Po důkladnějším zkoumání lze pak zjistit, že autor většinu materiálu ze stran 29–31 opsal z knihy Groeneboom and Wellner (1992) a v tomto textu pouze zaměnil některá slova, aby se nejednalo o stoprocentní plagiát.

Čtvrtou kapitolu pak doplňují obrázky odhadů pro výběry z různých rozdělení. Bohužel však není jasné, jak autor tyto obrázky vytvořil, případně jaký software použil.

**Pátá kapitola** porovnává odhady pomocí metody dekonvoluce a odhady pomocí metody neparametrické maximální věrohodnosti. Nejdříve porovnává asymptotické rozptyly odhadu

střední hodnoty a potom pomocí simulací odhady střední hodnoty i rozptylu. Tato kapitola je nejzajímavější a obsahuje hlavní přínos autorovy práce. Bohužel prezentaci výsledků chybí zasvěcenější komentář. Simulační výsledky jsou shrnuty v jednom odstavci a na následujících devíti stranách (str. 39-47) jsou pouze obrázky a tabulky. Postrádám také alespoň nějakou diskusi podmínek, které oba přístupy vyžadují. V tabulkách a grafech (resp. jejich popisích) naprosto chybí uvedení teoretických hodnot parametrů  $\mu$  a  $\sigma^2$ , takže člověk neví, s jakými hodnotami má porovnávat průměr z odhadů.

#### Podrobnější připomínky:

- (1) Autor často nevysvětluje značení - např.  $F$  na str. 3,  $D$  na str. 8,  $\mathcal{AN}$  na str. 10,  $Un$  na str. 36.
- (2) Autor značně stěžuje čtenáři možnost dohledat si použitá tvrzení. Tak např. pro Fubiniho větu se autor odkazuje na knihu Rudin (1987) bez bližšího určení. Podobně se bez bližšího určení pro CLV autor odkazuje na knihu Serfling (1980). A kde hledat Theorem 4.2.1 mně již z okolního textu není jasné vůbec.
- (3) Při tvorbě bibliografie si autor nedělal hlavu s tím, aby se dostalo zejména časopiseckým článkům jednotného formátování.
- (4) Strana 31: Chybí vysvětlení, co to je „*isotonic regression*“.
- (5) Strana 31: Chybí vysvětlení, co to je „*standard two-sided Brownian motion*“.
- (6) Strana 34: Jistě by bylo užitečné více vysvětlit, proč (4.9) a (4.10) implikují (4.11).
- (7) Strana 35 1. odstav.: Autor označuje odhad metodou neparametrické maximální věrohodnosti jako asymptoticky eficientní. Chybí však vysvětlení, co tím míní.
- (8) Strana 38: Škoda, že autor blíže nevysvětluje proč „*Uniform deconvolution model ... work better for a small number of observations than the NPML method.*“
- (9) Domnívám se, že tabulky 5.1 až 5.6 by si zasloužily i výpočet střední čtvercové chyby. Bez ní se kvalita odhadů jen těžce porovnává.
- (10) Tabulky 5.1 až 5.6 by výrazně zpřehlednilo, pokud by čísla v ní byla zaokrouhlena na výrazné menší počet desetinných míst.
- (11) Prezentovaným histogramům by např. prospělo, pokud by na nich byla vyznačena teoretická hodnota odhadovaného parametru.
- (12) Strana 48: Domnívám se, že tvrzení: „*Simulations of the estimators in the uniform deconvolution method show results which are closer to the theoretical values of the mean  $\mu$  and variance  $\sigma^2$  than the results of the NPML based estimators for small samples.*“ je velmi zjednodušující a navíc neodpovídá výsledkům simulací z páté kapitoly.

#### LITERATURA

- Groeneboom, P. and Wellner, J. A. (1992). *Information bounds and nonparametric maximum likelihood estimation*. Basel: Birkhäuser Verlag.
- Rudin, W. (1987). *Real and complex analysis, 3rd ed.* McGraw-Hill, Inc., New York, NY, USA.
- Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

Závěrem bych řekl, že přestože autor odvedl nemalý kus práce, z prezentace výsledků není vůbec jasné, zda autor rozumí tomu, o čem píše. Doporučení, zda má být práce uznána jako práce diplomová, si tedy dovoluji odložit až na průběh obhajoby.

18. května 2012

Ing. Marek Omelka, Ph.D.  
KPMS MFF UK