

Posudek na disertační práci Mgr. Petra Kláštereckého: "Odhady parametrů v subkohortních studiích"

Práce je věnována problémům analýzy v regresních modelech pro životnost, a to v případech subkohortních studií. Znamená to, že z dané studované skupiny objektů – kohorty – jsou podle určitého schematu (nejčastěji náhodně, ale s různou vahou) vybírány pro analýzu „případy“ (objekty s pozorovanou událostí) a kontrolní objekty. V prvních částech práce se autor věnuje zejména modelu s proporcionálním rizikem (Proportional Hazard Model, PHM) a modelu s proporcionálními šancemi (Proportional Odds Model, POM). Rekapituluje metody pro odhadování parametrů a detailně i strukturu příslušných skórových procesů. Ukazuje, i pomocí simulovaných příkladů, jak různé váhy výběru subkohorty mohou ovlivnit vychýlení odhadu parametru, a to zejména v případech relativně nízkého výskytu sledovaných událostí.

Dále se autor zabývá i logistickým regresním modelem a jeho formální podobností (ač neobsahuje čas) s POM. To jej vede k myšlence využít posloupnosti odhadů logistického typu a jejich konvexních kombinací i v případě dat pocházejících z POM.

Konstrukce takovýchto odhadů a odvození jejich asymptotických vlastností (konzistence a asymptotické normality) je obsahem stěžejní 4. kapitoly práce. Nejprve je detailně popsána konstrukce a dokázány vlastnosti pro případ kombinace konečného počtu odhadů obdržенých v fixovaných časech. Pak je situace rozšířena na případ, kdy je nový odhad konstruován vždy v čase nového případu, teoretické vlastnosti jsou pak dokázány i v této obecnější situaci. Následuje 5. část práce, kde je chování navržených odhadů zkoumáno pomocí simulovaných příkladů. Data jsou generována jak z POM tak z PHM, a v případě subkohortních studií na předvedených příkladech je vychýlení kombinovaného logistického odhadu zjevně menší než u standardních odhadů (popsaných v částech 2 a 3).

Práce je psána se znalostí dané tematiky, jejích teoretických i praktických problémů, přináší nové výsledky v dané oblasti. Nenašel jsem závažnější nedostatky, mám jen několik poznámek a dotazů, spíš pro objasnění souvislostí:

1. V úvodních částech se pracuje s PHM a POM, ale hlavně z hlediska parametru β . Jistě by stálo za zmínku i odhadování (a vlastnosti odhadů) neparametrických komponent modelů, neb jednak charakterizují kontrolní skupinu, jednak jsou důležité pro případné testy dobré shody.
2. V části 3.4, simulační studie, není řečeno, jaká byla právě ona „základní“ charakteristika pro generovaná data (tj. základní riziková funkce v PHM například). Byla volena konstantní? Tatáž informace mi chybí v studii v kapitole 5.
3. Hledají se optimální váhy pro kombinaci odhadů (str. 45). Je nějak zaručeno, že jsou jednoznačné?
4. V numerické studii v kap. 5 by bylo zajímavé i sledovat vývoj odhadu $\alpha^*(t)$ reprezentující neznámou funkci. Ta by ale podle vzorce (4.1) mohla i záviset na kovariátě z (?). Zrovna tak by bylo zajímavé sledovat posloupnost logistických odhadů $\beta(t)$, případně z ní vyvozovat závěry o nekonstantnosti β .
5. Jak v případech s celou kohortou („full data“) a s cenzorováním zachází logistický model (case-control estimator) s cenzorovanými daty (v kapitole 5)?
6. V Závěru je stručná zmínka o rozšíření na případ časově proměnných kovariát. Jak by to ale bylo řešeno, zdá se mi, že se časově proměnné kovariáty obtížně včlení do odhadu odvozeného z logistického modelu. Používá se vůbec POM s časově proměnnými kovariátami?
7. Nejsou dobré výsledky (v simulační studii v kap. 5) odhadu pomocí kombinovaného logist. odhadu i v situaci PHM, a nakonec všeobecně dobré



výsledky prostého „case_control“ odhadu, jen důsledkem poměrně jednoduchých dat (exponenciálních?)?

Ještě maličkosti:

8. U Lemma 4.1, str. 48, se vyskytují skórové funkce U_α, U_β , které nebyly předtím přesně definovány (i když z kontextu je zřejmý jejich význam). Podobně na str. 49 se nejdříve vyskytnou U_i , a až později je řečeno, co symboly znamenají.
9. Symbol α je použit jak pro váhu výběru v kapitole 3, tak pro parametr logistického modelu.

Celkově lze shrnout, že práce je věnována problematice důležité pro rozvoj jak teorie tak metod statistické analýzy přežití, a přináší v této oblasti nové netriviální výsledky zajímavé jak teoreticky, tak prakticky. Soudím, že Mgr. Klášterecký v této disertaci jasně prokázal svou schopnost věnovat se vědeckému výzkumu a proto disertaci doporučuji k obhajobě.

31.10.2009



