



P. KLÁŠTERECKÝ & M. KULICH

LOGISTICKÁ REGRESE V SUBKOHORTNÍCH STUDIÍCH

DNES V LISTU:

- Co jsou subkohortní studie, k čemu slouží a jak se pracuje s jejich daty
- Proč někdy nefungují tak dobře, jak by měly
- A nakonec jeden nápad, jak by se to dalo spravit

DÁLE SI NENECHTE UJÍT:

- Podrobný popis použitých metod a modelů
- Přiměřené množství důkazního materiálu a samozřejmě závěrečnou diskusi
- ! Vychází již v letošním sborníku !

Subkohortní studie

Klasické kohortní studie mohou být např. při zkoumání vzácných onemocnění velmi drahé – pro zajištění dostatečného počtu událostí ve výběru je často nutné sledovat i obrovskou kontrolní skupinu. Subkohortní (case-cohort) studie typicky pracuje pouze s náhodně vybranou malou částí původní kohorty, subkohortou, ke které se navíc přidají všechny pozorované události. Že vám to něco připomíná? Studie typu case-control?

Jak to funguje?

Celkem jednoduše. Data se obvykle analyzují regresním modelem, pro odhad jeho parametrů se vezmou skórové (odhadovací) rovnice známé z klasické analýzy přežívání a místo běžného součtu příspěvků od jednotlivých pozorování se použije součet vážený. Váhy pro pozorování mimo subkohortu se obvykle volí nulové (abychom nepočítali s neznámými hodnotami), pro ostatní pozorování potom $1/p_i$, kde p_i je pravděpodobnost výběru do subkohorty.

Asymptotické vlastnosti subkohortních odhadů však platí pouze za předpokladu, že $p_i \geq \varepsilon > 0$. Pro velmi malé pravděpodobnosti (stačí již $p_i \sim 10^{-3}$) začínají mít odhady vychýlení a jejich intervaly spolehlivosti slabší pokrytí – viz náš příspěvek na minulém Robustu. Jelikož je kontrolních pozorování často řádově tolik, kolik očekáváme událostí, jsou velmi nízké p_i při analýze vzácných jevů poměrně běžné. Problémy se pak objevují u všech typů regresních modelů a všech používaných subkohortních odhadů.

Studie typu case-control

podobnými problémy netrpí, protože poměr šancí odhadovaný logistickou regresí lze určit i z takto nevyváženého výběru bez použití vah pro jednotlivá pozorování. Absolutní člen z těchto dat odhadnout nelze, ale ve většině případů jej odhadovat nepotřebujeme.

Spojitost mezi oběma typy studií

Uvažujme pro doby do události model proporcionálních šancí, který lze zapsat ve tvaru

$$-\text{logit}(S(t|\mathbf{Z})) = G(t) + \beta' \mathbf{Z},$$

kde $S(t)$ značí pravděpodobnost přežití v čase t a $\text{logit}(x) = \log(x/(1-x))$. Neznámé parametry jsou zde $G(t) = -\text{logit}(S(t|\mathbf{Z} = \mathbf{0}))$, referenční logaritmická šance výskytu události v čase t , a vektor regresních koeficientů β . Model předpokládá, že poměr šancí na přežití

(survival odds ratio) mezi dvěma libovolnými subjekty, vyjádřený pomocí β , zůstává v čase konstantní.

V subkohortní studii máme k dispozici informaci o časech událostí nebo cenzorování. Kdybychom ale znali pouze indikátory výskytu sledované události v nějakém čase τ_0 , použili bychom logistickou regresí se stejným tvarem modelu, jen místo funkcí bychom dosadili konstanty $S(\tau_0|\mathbf{Z})$ a $G(\tau_0) = \alpha(\tau_0)$. To se bude hodit...

Využití logistické regrese

Předpoklad neměnnosti odhadovaného poměru šancí

- ⇒ model proporcionálních šancí odhaduje stejné regresní parametry β jako case-control logistická regrese provedená na konci studie
- ⇒ analýzu pomocí logistické regrese lze provést i v jiných časech než na konci studie a stále bude konzistentně odhadovat stejné parametry
- ⇒ získané odhady můžeme zkombinovat do jednoho a zachytit tak časovou informaci, která se při použití logistické regrese ztrácí

Kombinovaný odhad

- Výsledný odhad definujeme jako takovou konvexní lineární kombinaci dílčích logistických odhadů, která minimalizuje rozptyl
- Kombinovaný odhad je konzistentní a asymptoticky normální
- Dílčí logistické analýzy je nejvhodnější provádět v časech událostí, kde se mění hodnota skokové věrohodnostní funkce v modelu prop. šancí
- Při velmi malých pravděpodobnostech události odhadnutý poměr šancí dobře aproximuje relativní riziko ⇒ použitelnost i pro Coxův model

Sečteno a podtrženo

Nový odhad regresních parametrů pro model proporcionálních šancí funguje na principu zkombinování několika postupných odhadů stejných parametrů logistickou regresí. Jeho uplatnění předpokládáme zejména v subkohortních studiích s velmi řídkými jevy, kde v simulačních studiích vykazuje lepší vlastnosti než odhady dosud používané a navíc může posloužit jako

dobrá aproximace i pro známější a používanější model proporcionálních rizik.

Odhad byl prozatím odvozen a zkoumán při jednoduchém cenzorování konstantou (koncem studie). V dalším výzkumu je třeba věnovat pozornost především jeho chování při obecnějších typech cenzorování.