

Oponentský posudek bakalářské práce

Mixed Poisson models for claim counts

Autor práce: Filip Hauptfleisch

Předložená bakalářská práce se zabývá vybranými aspekty regresního modelování počtů hlášených pojistných událostí prostřednictvím smíšených Poissonových pravděpodobnostních rozdělení. Obsahově je vlastní text rozčleněn do tří kapitol. První kapitola připomíná definici Poissonova rozdělení a jeho klíčové vlastnosti. Druhá kapitola se podrobněji věnuje problematice smíšených Poissonových distribucí. Přiměřeně je zde přítom reflektován širší kontext regresního modelování čítacích vysvětlovaných proměnných. Zatímco první dvě kapitoly předmětné bakalářské práce mají vesměs kompilační charakter, třetí lze považovat za vlastní příspěvek autora. Na základě reálných dat jsou zde porovnávány různé regresní modely vysvětlující počty hlášených pojistných událostí.

Vlastní vymezení i zpracování tématu bakalářské práce považuji za vyhovující. Autor se podle všeho dokázal orientovat v odborné literatuře a aplikovat zvolené statistické přístupy v prostředí výpočetního software **R**, a to včetně odpovídající diskuze a interpretace dosažených výsledků. Po formální stránce je práce bez problémů. Podmínky kladené na rozsah splňuje. Text je psán angličtinou na dobré úrovni. Obsahuje přijatelné množství stylistických, pravopisných či typografických pochybení. Úprava práce je rovněž adekvátní. Celkový dojem ovšem devaluje větší množství (převážně matematických) nepřesností.

V textu se například objevily tyto nedostatky:

- V sekci 1.1, strana 3 by ve vlastní definici Poissonova rozdělení mělo vystupovat $e^{-\lambda}$, nikoliv e^{λ} . Navíc se toto pravděpodobnostní rozdělení definuje pro všechna nezáporná celá k , nikoliv pouze pro přirozená k (viz například sekce 1.2.2).
- V tvrzení *Theorem 1* na straně 6 není specifikováno, o jaký typ konvergence se jedná. Nadto se normální rozdělení obvykle označuje prostřednictvím N , nikoliv n .
- V úvodu sekce 1.4, strana 6 by mělo být uvedeno: $P[X_i = x_i | Y_i = y_i]$.
- V sekci 1.4, strana 7 by ve formulaci věrohodnostní funkce mělo vystupovat $e^{-\lambda(\cdot)}$, nikoliv $e^{\lambda(\cdot)}$.
- V sekci 1.4, strana 7 očividně není korektně zapsána soustava rovnic, jejíž řešení vede k maximálně věrohodnému odhadu neznámého vektoru parametrů β .
- V sekci 1.4, strana 8: (i) rovnost zavádějící matice A zahrnuje nesprávnou dolní sumační mez, (ii) formule definující matice A , B obsahují součin $\mathbf{y}_i^T \mathbf{y}_i$ namísto adekvátní varianty $\mathbf{y}_i \mathbf{y}_i^T$, (iii) matice A , B by měly být značeny tučným řezem písma.
- V úvodu sekce 2.1.1, strana 9 není v rámci kalkulace střední hodnoty a rozptylu náhodné veličiny X plně objasněno použité značení.
- V sekci 2.1.2, strana 10 absentuje informace, pro jaká k , resp. y jsou definovány negativně binomické, resp. gama rozdělení a jakých možných hodnot mohou nabývat jim příslušející parametry.
- V sekci 2.2.1, strana 12 je pravděpodobně chybně zavedeno obecné *zero-inflated* rozdělení (pro $k > 0$).
- V sekci 2.2.1, strana 13 je pravděpodobně chybně zavedeno Poissonovo *zero-inflated* rozdělení. Nadto je pro něj zřejmě nesprávně spočten rozptyl.
- V sekci 2.2.2, strana 13 se má zjevně podmiňovat pomocí $[Y > 0]$, nikoliv $[X > 0]$. Platí dále, že $p_0 \in (0, 1)$.
- Na straně 14 jsou přítomny některé (typografické) nepřesnosti: (i) v první formulaci na této stránce chybí rovnítko, (ii) v rovnici (2.2) není jednotně zapisována indikátorová funkce a (iii) v předposlední větě na dané straně vystupuje $e - \lambda$ namísto $e^{-\lambda}$.

- V sekci 2.3.1, strana 15 (a dále) by měl být blíže specifikován počet stupňů volnosti LR statistiky.
- V sekci 3.1, strana 17, druhý odstavec by poslední část druhé věty „...*sample mean is 0.077.*“ měla být podle všeho nahrazena takto: „...*sample variance is 0.077.*“.
- Regresní modely diskutované v kapitole 3 mohly být pro lepší přehlednost explicitně formulovány.
- Není zřejmé, čemu odpovídá parametr θ z tabulky 3.3 (srovnejme kupříkladu se sekci 2.1.2).
- V úvodním odstavci přílohy B (stejně jako v prvním odstavci kapitoly 3) chybí nějaká informace (?).
- V bibliografickém přehledu se u knižních publikací obvykle uvádí rovněž místo vydání.

I přes výše uvedené nedostatky navrhuji bakalářskou práci přijmout k obhajobě.

V Pardubicích dne 13. června 2017

Radek Hendrych