

Posudek vedoucího diplomové práce

Název: Prostorová epidemiologie

Autor: Bc. Adéla Jalovcová

Mezi základní epidemiologické problémy patří studium dynamiky dané choroby v čase, často se však zapomíná na prostorovou povahu dat. Ta může přinést cenný vhled do problému například při šíření infekčních nemocí. Dnes už jsou k dispozici sofistikované hierarchické modely zahrnující prostorové autokorelace v datech a bayesovské metody odhadu jejich parametrů. Standardní metodou odhadu bylo donedávna použití Markov Chain Monte Carlo metod. Tento postup však vyžaduje vysoký výpočetní výkon, netriviální zkušenost uživatele, ladění řetězce a návrhových rozdělení, kontrolu konvergence ke stacionárnímu rozdělení apod.

V posledních několika letech se v bayesovské analýze prostorových dat výrazně prosazuje přístup založený na numerické aproximaci aposteriorního rozdělení namísto MCMC simulací. Tento postup umožňuje efektivně odhadovat parametry latentních gaussovských modelů zahrnujících prostorové závislosti a je označován jako Integrated Nested Laplace Approximation (INLA). Díky standardní implementaci v balíčku R-INLA se tento přístup stal standardem při odhadování parametrů daných modelů. Praktické použití tohoto balíčku je ale pořád netriviální a představuje jistou výzvu. Tomu nepomáhá ani kvalita dostupné nápovědy a zveřejněných ukázek a tutoriálů.

Studentka předkládá k obhajobě aplikovanou práci, kde na datech o výskytu nemoci COVID-19 v České republice ilustruje použití jak základních metod prostorové statistiky (testování významnosti prostorových autokorelací v kapitole 2), tak velmi komplexních latentních gaussovských modelů zahrnujících prostorovou (v kapitole 3), resp. časoprostorovou složku (v kapitole 4).

Metodologickým přínosem je použití obálkových testů k posouzení shody odhadnutého modelu s daty v sekci 3.5 – standardem v dostupné literatuře je použití skalárních testových statistik místo vektorových. Dalším přínosem je příprava zdrojových kódů v programu R, které lze využít jako šablonu při analýze jiných datových sad. Kódy jsou připojené k práci jako elektronická příloha. Vycházejí z kódů doprovázejících knížku Blangiardo a Cameletti (2015), které už vzhledem k aktualizacím balíčku R-INLA nejsou plně funkční. Aktualizace kódů stála studentku velké množství času a energie.

Původní zadání práce mířilo na využití statistických metod pro bodové procesy, kdy se předpokládá, že je zaznamenána poloha bydliště každého nakaženého, resp. místo jeho nakažení, pokud je možné takové místo identifikovat. V případě veřejně dostupných údajů o průběhu epidemie COVID-19 v České republice jsou však k dispozici pouze agregovaná data udávající počty nakažených na úrovni jednotlivých okresů. Proto jsem na začátku řešení práce navrhl zaměřit pozornost na statistické metody pro náhodná pole na mřížích, abychom mohli právě tato data využít. S úpravou tématu práce oproti původnímu zadání jsem tedy z pozice vedoucího práce nejen souhlasil, ale také jsem ji v první řadě inicioval.

Autorka pracovala pečlivě a samostatně, prokázala vlastní kreativitu, snahu proniknout do jádra problému a schopnost překonat všechny překážky, které analýza reálných dat přináší.

Použité zdroje jsou řádně citovány, jazyková i formální úroveň práce je velmi dobrá. Práci doporučuji přijmout jako diplomovou práci k obhajobě.

V Praze dne 24. 7. 2022

RNDr. Jiří Dvořák, Ph.D.