

Barbora Zuzáková
Modelování velkých škod
posudek oponenta diplomové práce

Uchazečka se v práci zaměřuje na modelování a statistickou analýzu extrémů, což je zejména v pojišťovnictví časté téma. Hlavní část práce je věnována statistickým úlohám spojeným s nejvyššími hodnotami (extrémy) náhodného výběru. Je představeno mnoho různých postupů a metod a ty jsou nakonec ilustrovány na reálných datech. Toto téma rozhodně nepatří k jednoduchým, zejména i proto, že sice existují krásné teoretické výsledky, ty však často nemůžeme rozumně použít: i ve velkém výběru je jen málo opravdu vsokých hodnot, existují tři možná limitní rozdělení a pro každé z nich čelíme jiným nesnázím při použití, nemáme k dispozici konzistentní postupy pro všechny hodnoty parametrů. Své zhodnocení práce shrnu do několika otázek a připomínek:

- Práce často cituje různou klasickou literaturu o statistice extrémů s tím, že autoři něco doporučují či něco tvrdí, aniž by cokoliv z jejich argumentů bylo v práci uvedeno. Chápu, že do jedné diplomové práce vše zahrnout nejde, ale takto to vypadá, že autorka si ani nevytváří vlastní názor, natož aby bylo možné posoudit, nakolik autorka vlastně celé problematice porozuměla.
- Poznámka 2 na straně 5 je špatně formulovaná. Co je skutečně limitou výrazu na levé straně? A jak může být za šípkou výraz obsahující n ? Podobný problém vykazuje věta 4 na straně 8 nahoře.
- Jak byste ukázala, že jak metodou blokových maxim, tak metodou přesahů dostanete nezávislé stejně rozdělené náhodné veličiny?
- V části 3.3.3 se hovoří o penalizované metodě maximální věrohodnosti. Jak zdůvodníte penalizaci hodnot $\xi > 1$? Volila jste stejnou penalizaci i v páté kapitole v analýze dat (a proč jste volila tu, kterou jste volila)?
- V části 3.4.1 nerozumím několika řádkům pod vzorcem (3.25). Proč hodnota $\tilde{H}^{(u)}(z_{j:k})$ jde k jedné, když $z_{j:k}$ roste? Podle mého vzorec (3.24) jasně říká, že $\tilde{H}^{(u)}(z_{j:k}) = j/k$ s čímž souhlasím.
- Strana 21 dole: co znamená, že nějaká statistika „performs better than . . .“? A jak mám rozumět větě, že v praxi se doporučuje Andersonův-Darlingův test na prvním místě, zatímco Cramérův-von Misesův je druhou volbou? Kdy tedy volím C-vM test?
- Na straně 23: co se míní tím, že jsou požadavky cenzorovány? A dále, kde vidíme na logaritmických datech znatelnou kladnou šikmost?
- Prosil bych o vysvětlení a zdůvodnění postupu na straně 24 dole.
- Strana 29: celý ten postup mi přijde jako věštění z křišťálové koule. Jak mohu takovému postupu věřit? Obrázek mi radí 850, já se ad hoc rozhodnu pro 600, proč ne pro 650, 55é, nebo 700? K čemu tedy ten obrázek je?
- Proč od sebe oddělujete jednotlivé roky? Proč nemodelujete ty roky dohromady?
- Obrázek 5.4 by si zasloužil lepší popis i vysvětlení.
- Tabulka 5.2 na straně 31 ukazuje, jak odhady hrozně závisí na volbě prahu. Je neskutečný rozdíl mezi hodnotami parametru kladnými a zápornými, co tedy si s této tabulky mám vybrat? A jak se dají interpretovat velké rozdíly v parametrech mezi jednotlivými roky? Totéž se týká tabulky 5.6.
- K obrázku 6.2 na straně 42: mnohem víc by nás přeci zajímalo, jaká bude tato hodnota v budoucnu, abychom mohli pojišťovnu zajistit. Jak to udělat?

- Na začátku práce se mluví o tom, že je třeba modelovat několik největších škod v roce, ty že jsou rozhodující. Tomu se však práce, podle mě, nevěnuje vůbec. Celá práce je zaměřena na modelování extrému, tedy nejvyšší hodnoty z náhodného výběru. Problém popsáný v prvním odstavci úvodu ale vyžaduje společné modelování k nejvyšších hodnot, místo jedné extrémní. Zde pak narazíme na další potíže, protože k nejvyšších hodnot nepředstavuje náhodný výběr, jde o náhodný vektor se závislými složkami. Setkala se autorka někde s tímto problémem?

Přestože mám, jak vidno, několik kritických připomínek k práci, nedostatky nejsou nijak závažné. Hlavně bych rád z průběhu obhajoby pochopil, nakolik posluchačka porozuměla celkem náročné části statistiky. Celkově doporučuji **uznat práci jako diplomovou**.

Daniel Hlubinka
11.9.2013