

## POSUDEK OPONENTA DIPLOMOVÉ PRÁCE

**Název:** Křivky závislosti odpovědi na dávce

**Autor:** Bc. Martin Hezoučský

### SHRNUTÍ OBSAHU PRÁCE

Diplomová práce bakalára Martina Hezoučského rieši jeden zo základných problémov farmaceutických firiem zapojených do výzkumu a vývoja nových liečiv—detekciu resp. rozpoznanie účinnej látky a stanovenie prípustnej resp. optimálnej dávky. Jedná sa o teoreticko-prakticky zameranú diplomovú prácu, ktorá na jednej strane predstavuje vhodný štatistický model pre daný problém a na strane druhej aplikuje navrhnutú metodológiu na konkrétne data a na základe získaných výsledkov poskytuje niektoré praktické doporučenia.

Z koncepčného pohľadu sa práca skladá z dvoch častí: V prvých dvoch kapitolách autor stručne predstavuje proces výroby nových liečiv a problém rozpoznania účinnej látky spolu s úlohou následného stanovenia vhodnej dávky. Predstavené sú niektoré základné a farmaceutickým priemyslom štandardne používané štatistické postupy. Autor sa na jednej strane snaží tieto postupy popísať formálnym matematicko-štatistickým jazykom, na druhej strane je ale prvú časť práce nutné považovať len za akýsi všeobecný “popularizačný” text, ktorému snaha o použitie formálneho matematického zápisu skôr značne uškodila, než prospela (často sa totiž opakuje nezavedené a dokonca aj nesprávne definované značenie, nezmyselná formulácia matematického textu, ale aj zásadné chyby v niektorých vzorcoch—viď nižšie niekoľko konkrétnych poznámok).

Druhá časť práce—kapitoly 3 a 4—predstavuje inovatívny a pomerne rozsiahly matematický aparát založený na detekcii a odhade change-pointu v tzv. modeloch postupnej zmeny, ktoré autor adaptoval na problematiku detekcie účinnej látky a stanovovania príslušnej dávky a poskytol tak metódu, ktorá nad rámec bežne používaných techník umožňuje aj využitie distribučných vlastností získaných odhadov (nutných napr. pri konštrukcii intervalov spoľahlivosti). V tejto časti práce, ktorá je po formálnej stránke napísaná výrazne lepšie ako predchádzajúca, je obsiahnutý aj hlavný autorov príspevok, spočívajúci v niektorých predstavených modifikáciách daných modelov a ich empirickom porovnaní. Z hľadiska praktickej využiteľnosti si ale myslím, že v tejto časti zase chýbajú niektoré podstatné informácie, bez ktorých praktik navrhnutú metodológiu nedokáže správne uplatniť (napr. nedostatočne vysvetlené bootstrap algoritmy—viď opäť niekoľko konkrétnych poznámok nižšie).

Z celkového hľadiska si ale myslím, že sa jedná o zaujímavú a celkom dobrú prácu. Možno by bolo vhodné z rozsahu navrhnutých a diskutovaných modelov mierne ubrať a namiesto toho niektoré zmiernené modely prediskutovať trochu podrobnejšie (práci trochu chýba akékoľvek odvodenie, dôkaz používaného tvrdenia, či vety). Na druhej strane, predložená práca je teoreticko-praktického charakteru a preto to nepovažujem za závažný nedostatok. Prácu jednoznačne doporučujem komisii uznať ako diplomovú prácu na KPMS.

Praha, 30.08.2020



RNDr. Matuš Maciak, Ph.D.  
maciak@karlin.mff.cuni.cz

## PRIPOMIENKY K PRVEJ ČASTI

- ❑ Str.7: Keď sa spomínajú pozorovania  $\mathbf{Y}$ , asi by bolo vhodné napísať, že sa jedná o náhodný vektor. A prípadne doplniť informáciu o tom, aké ma zložky. Bez toho je ťažké pozorovania  $\mathbf{Y}$  identifikovať s modelom (2.1).
- ❑ Str.7: Čo je  $\Theta$  a  $\Theta^0$  (číslo, vektor, matica)? Keďže sa jedná o diplomovú prácu na Matematicko-fyzikální fakulte, bol by vhodnejší presnejší zápis, napr.:  $\Theta \in \mathbb{R}^p$ , pre  $p \in \mathbb{N}$ . Analogická poznámka by sa dala aplikovať na väčšinu značenia použitého v prvých dvoch kapitolách.
- ❑ Str.9: Čo konkrétne sa myslí pod *silnou kontrolou FWER*? Ja to viem, a nepochybujem, že to vie aj autor. Ale bolo by dobré takúto informáciu pridať aj do práce.
- ❑ Str.10: Striktne vzaté, vôbec nie je zrejmé, vzhľadom k čomu sa minimalizuje výraz (2.8)? Taktiež, ako sa získa odhad  $\widehat{MED}_\Delta$  v (2.11), keďže hodnoty  $\Theta$  nie sú známe—nemali by byť vo výraze v (2.11) namiesto  $\Theta$  príslušné odhady  $\widehat{\Theta}$ ?
- ❑ Str.11: Hodnota  $E_0$  parametrizuje príslušnú krivku v Tabuľke 2.1, takže asi ťažko (resp. s pravdepodobnosťou nula) sa jedná o konkrétnu *hodnotu odezvy (efektu) pro placebo dávku*. Skôr sa jedná o *strednú hodnotu odezvy (efektu) pro placebo dávku*. Analogický pre niektoré ďalšie uvádzané parametre, ktoré dohromady parametrizujú strednú hodnotu, nie konkrétnu hodnotu nejakého daného subjektu.
- ❑ Str.13: Ako je definovaný vektor? Pri danom značení vektorové násobenie v hypotéze  $H_0^m$  vygeneruje maticu. Asi ma byť správne  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_{m_k})^\top$  na konci str.12.
- ❑ Str.15: Tvrdenie, že “*predpokládame, že kontrolní dávka  $d_0$  nemá žádný účinek, tj. pozorujeme pro ni  $Y_{0,j} = 0$  pro  $j = 1, 2, 3$* ” je nesprávne. Hodnota  $d_0$  predstavuje strednú hodnotu  $Y_{0,j}$  pro  $j = 1, 2, 3$ , a nie samotné sledované realizácie náhodnej veličiny  $Y_{0,j}$ . Mimochodom, prečo je zarovnanie okrajov na str.15 iné?
- ❑ Str.16: Naozaj sú dané testy uskutočnené na uvedenej hladine 0.95, ako je napísané v úvode strany 16? To by znamenalo, že pravdepodobnosť chyby prvého druhu je až 0.95. Nie je to príliš veľa? Zároveň nie je jasné, ako vyzerá nulová hypotéza pre Tabuľky 2.3 a 2.4. Zarovnanie na tejto strane je opäť neštandardné.

## PRIPOMIENKY K DRUHEJ ČASTI

- ❑ Str.22: V príklade č.3 je napísané, že počet pozorovaní je  $n = 15$ . V Obrázku 3.1 je ale 45 nameraných hodnôt. Ako teda vyzerajú podkladové data a prečo sa pre príslušné dávky uvažujú iba priemerné hodnoty? Nebolo by vhodnejšie pre účely odhadu rozptylu uvažovať pôvodné data, tj.  $n = 45$ ?
- ❑ Pri pohľade na data v Obrázku 3.2: Naozaj máte pocit, že predpoklad  $\mu = 0$  je z praktického hľadiska opodstatnený?
- ❑ Str.24: Prečo sa vo výraze (3.15) minimalizuje cez  $m \in [1, n]$ ? Predpokladám, že sa myslí množina  $\{1, \dots, n\}$ ? Taktiež nie je jasné, ako je definovaná hodnota  $\bar{x}_m$ .

- ❑ Str.25: Čo je to  $b$ ? Nemalo by sa jednať o postupnosť odhadov  $\{b_n\}_n$ , tak ako to je zavedené v značení na str.24?
- ❑ Str.26: Čo konkrétne máte na mysli, keď píšete, že “*vygenerujeme jeden soubor nových pozorování  $Y_i^*$* ”? Ako presne  $Y_i^*$  generujete? Analogická poznámka platí aj pri ďalších použitíach metódy bootstrap. Pre praktické využitie musí byť zrejmé, ako presne bootstrapove data získate. **Poprosím rozmyslieť a dostatočne podrobne pripraviť k obhajobe.**
- ❑ Str.29: Naozaj je na ľavom dolnom panely v Obrázku 3.5 hustota príslušného normálneho rozdelenia? Čo presne sa myslí slovom *príslušné* normálne rozdelenie? Prečo táto hustota chýba v pravom dolnom panely?
- ❑ Str.31: V tabuľke 3.4 sú empirické relatívne pokrytia neznámych parametrov. Jedná sa o asymptotické intervaly spoľahlivosti. Prečo je v simuláciach potom použitá iba jedna hodnota  $n = 30$ ? Analogická poznámka opäť platí aj pre niektoré ďalšie tabuľky.
- ❑ Str.34: Aké vlastnosti má odhad rozptylu na konci strany? Nebolo by vhodnejšie zohľadniť stupne volnosti? Resp. prečo je odhad definovaný práve takto?
- ❑ Str.41: Opäť, pri pohľade na Obrázok 3.11 mi pride zvláštno, že pred bodom zmeny sú všetky rezídua kladné (až na jedno jediné, ktoré je ale pomerne malé). Je v tomto prípade (z teoretického aj praktického pohľadu) predpoklad  $\mu = 0$  oprávnený?