



**MATEMATICKO-FYZIKÁLNÍ
FAKULTA**
Univerzita Karlova

DIPLOMOVÁ PRÁCE

Tomáš Šlampaik

Chyba predikce pro smíšené modely

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: doc. RNDr. Arnošt Komárek, Ph.D.

Studijní program: Matematika

Studijní obor: Pravděpodobnost, matematická statistika
a ekonometrie

Praha 2018

Prohlašuji, že jsem tuto diplomovou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V Praze dne 10. července 2018

Podpis autora

Na tomto mieste ďakujem doc. RNDr. Arnoštovi Komárkovi, Ph. D., za obetovaný čas pri vedení práce a konštruktívne poznámky pri jej tvorbe.
Poďakovanie venujem aj svojej rodine a blízkym priateľom, ktorí mi boli oporou a podporovali ma počas celého štúdia.

Název práce: Chyba predikce pro smíšené modely

Autor: Tomáš Šlampaik

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: doc. RNDr. Arnošt Komárek, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Lineární smíšený model je jeden z možných nástrojů při hledání vhodných modelů pro longitudinální nebo skupinově závislá data. Práce se zabývá vyhodnocováním jeho predikční chyby. Uvádí nejprve výpočet střední čtvercové chyby predikce pomocí přímého výpočtu. Potom je v práci popsána metoda penalizace kovariancí a krosvalidace. Dále je ukázáno, jak se dá použít Akaikeovo informační kritérium v lineárním smíšeném modelu. Kvůli vlastnostem modelu jsou rozlišovány dva typy, marginální a podmíněné. Následně jsou popsány postupy jejich výpočtu a základní (asymptotické) vlastnosti. Nakonec práce obsahuje simulační studii, která sleduje chování marginálního a podmíněného kritéria při výběru správné varianční struktury náhodných efektů. Ukazuje se, že marginální kritérium má tendenci vybírat modely s nižším počtem náhodných efektů a podmíněné kritérium naopak upřednostňuje vyšší počet náhodných efektů.

Klíčová slova: smíšený model, chyba predikce, AIC

Title: Prediction error for mixed models

Author: Tomáš Šlampaik

Department: Department of Probability and Mathematical Statistics

Supervisor: doc. RNDr. Arnošt Komárek, Ph.D., Department of Probability and Mathematical Statistics

Abstract: A Linear mixed-effects model (LME) is one of the possible tools for longitudinal or group-dependent data. This thesis deals with evaluating of prediction error in LME. Firstly, it is derived the mean square error of prediction (MSEP) by direct calculation. Then the covariance penalty method and cross-validation is presented for evaluation of MSEP in LME. Further, it is shown how Akaike information criterion (AIC) can be used in mixed-effects models. Because of the model's properties two types of AIC are distinguished – marginal and conditional one. Subsequently, the procedures of AIC's calculation and its basic asymptotic properties are described. Finally, the thesis contains simulation study of behaviour of marginal and conditional AIC with the goal to choose the right variance structure of random effects. It turns out that the marginal criterion tends to select models with smaller number of random effects than conditional criterion.

Keywords: mixed-effects model, prediction error, AIC

Názov práce: Chyba predikcie pre zmiešané modely

Autor: Tomáš Šlampaik

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedúci diplomovej práce: doc. RNDr. Arnošt Komárek, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Lineárny zmiešaný model je jeden z možných nástrojov pri hľadaní vhodných modelov pre longitudinálne alebo skupinovo závislé dátové súbory. Práca sa zaoberá vyhodnocovaním predikčnej chyby lineárneho zmiešaného modelu. Uvádza najprv výpočet strednej štvorcovej chyby predikcie pomocou priameho výpočtu. Potom je v práci popísaná metóda penalizácie kovarianciou a krosvalidácia. Ďalej je ukázané, ako sa dá použiť Akaikovo informačné kritérium v lineárnom zmiešanom modeli. Kvôli vlastnostiam modelu sú rozlíšené dva typy, marginálne a podmienené. Následne sú popísané postupy ich výpočtu a základné (asymptotické) vlastnosti. Nakoniec práca obsahuje simulačnú štúdiu, ktorá sleduje správanie marginálneho a podmieneného kritéria pri výbere správnej variančnej štruktúry náhodných efektov. Ukazuje sa, že marginálne kritérium má tendenciu vybrať modely s nižším počtom náhodných efektov a podmienené kritérium naopak uprednostňuje vyšší počet náhodných efektov.

Kľúčové slová: zmiešaný model, chyba predikcie, AIC

Obsah

Úvod	2
1 Lineárny zmiešaný model a predikcia	4
1.1 Predikcia marginálne a podmienene	6
1.2 BLUE a BLUP	7
2 Chyba predikcie	9
2.1 Stredná štvorcová chyba predikcie	9
2.1.1 MSEP podmienenej predikcie	10
2.1.2 MSEP marginálnej predikcie	11
2.2 Penalizovanie kovarianciou	11
2.2.1 Parametrický bootstrap	14
2.2.2 Neparametrický bootstrap	14
2.3 Krosvalidácia	15
3 Akaikovo informačné kritérium	18
3.1 Vierohodnosť v LME	21
3.2 AIC pre lineárny zmiešaný model	23
3.2.1 Marginálne AIC	23
3.2.2 Podmienené AIC	24
3.2.3 Asymptotická nestrannosť AIC	34
4 Simulačná štúdia	42
4.1 Štúdia č. 1	43
4.2 Štúdia č. 2	45
4.3 Zhrnutie výsledkov	47
Záver	48
Zoznam použitej literatúry	49
A Prílohy	50
A.1 Prehľad teórie maximálnej vierohodnosti	50
A.2 Dôkaz Lemmy 12	52
A.3 Grafické výstupy simulačnej štúdie č. 1	54

Úvod

V rade štatistických aplikácií je jedným z hlavných cieľov analýzy zozbieraných dát výber vhodného modelu. Často skúmame závislosť pozorovanej premennej (hodnota fyzikálnej veličiny, prítomnosť choroby u pacienta, počet udalostí v časovom intervale a iné) na zvyšných pozorovaných veličinách, ktoré sme do štúdie zahrnuli. V tomto vzťahu závislú veličinu nazývame odozva a snažíme sa vhodným modelom popísať, ako závisí na nezávislých premenných, regresoroch. V praxi sa snažíme na základe zozbieraných dát vybrať model, ktorý im najlepšie odpovedá a je blízko realite. Reálne procesy sú však často komplikované, nepoznáme všetky ich zákonitosti, preto ich modelujeme v matematike stochasticky.

Prvým krokom je výber triedy modelov, resp. výber metódy, ako budeme dáta analyzovať. To úzko súvisí s ich štruktúrou, predovšetkým korelačnou. Záleží na tom, či pracujeme s pozorovaniami, o ktorých vieme, že sú nezávislé alebo s pozorovaniami, kde môžeme očakávať závislosti na rôznych úrovniach. V praxi je častejší prípad rozličných štruktúr závislosti, napríklad skupinovo závislé dáta (známe v rôznych oblastiach aplikácie pod názvami panelové dáta, zhlukové dáta, opakované merania, prípadne longitudinálne dáta). Výber metódy, resp. triedy modelov závisí predovšetkým na mechanizme vzniku dát, prípadne na ich zbere, ktorý môže určiť podobu možných korelačných štruktúr.

V našej práci budeme skúmať lineárne zmiešané modely, ktoré boli navrhnuté pre longitudinálne dáta alebo zhlukové dáta. Tie pozostávajú z nejakého počtu nezávislých subjektov, na ktorých buď sledujeme odozvu viackrát v rôznych časoch alebo sú tvorené skupinami subjektov a vrámci jednej skupiny očakávame, že sa vzájomne ovplyvňujú. Na výber možných modelov existuje viacero prístupov v literatúre a je úlohou štatistika vybrať jeho primárny cieľ. V našej práci rozoberieme prípad, keď sú hlavným cieľom dobré predikčné vlastnosti lineárneho zmiešaného modelu. S týmto cieľom vzniká potreba kvantifikovať chybu predikcie, ktorú sa následne snažíme samozrejme minimalizovať.

Cieľom práce je čitateľa najprv oboznámiť s možnými prístupmi odhadovania predikčnej chyby v lineárnom zmiešanom modeli, ktoré vychádzajú z prístupu penalizovania kovarianciou a z princípov krosvalidácie, ktorým sa venuje článok Efron (2004). Ďalej má práca pojednať o vlastnostiach Akaikeho kritéria zavedeného v článku Akaike (1974) a o jeho modifikáciách, ktoré sú potrebné v prípade závislých dát. Články Vaida a Blanchard (2005), Greven a Kneib (2010) alebo Kubokawa (2011) sa venujú problematike Akaikeho kritéria aj v lineárnom zmiešanom modeli a budeme ich v práci využívať. Nakoniec kladieme za cieľ zistiť pomocou simulačnej štúdie, ako sa tieto kritériá správajú pri výbere správneho modelu.

V prvej kapitole práce uvádzame definíciu lineárneho zmiešaného modelu, pripomenieme jeho základné vlastnosti a možnosti odhadovania parametrov. Druhá časť práce sa venuje strednej štvorcovej chybe predikcie a možnostiam jej výpočtu. Vysvetľuje spomínané princípy odhadovania predikčnej chyby krosvalidáciou alebo penalizovaním kovarianciou na príklade lineárneho zmiešaného modelu. Tretia kapitola je venovaná Akaikovmu informačnému kritériu. Vysvetľuje rozdielny princíp v závislosti na hierarchickej štruktúre modelu, podľa ktorého rozlišujeme marginálne a podmienené Akaikeho kritérium. Zahrňuje možnosti ich

výpočtu, ktoré sú použité v poslednej časti práce. Tá sa zaoberá porovnaním marginálneho a podmieneného kritéria pri výbere správneho modelu pomocou simulačnej štúdie. Konkrétne porovnáva výber správnej variančnej štruktúry náhodných efektov lineárneho zmiešaného modelu.

1. Lineárny zmiešaný model a predikcia

V tejto časti uvedieme definíciu lineárneho zmiešaného modelu a spomenieme vlastnosti, ktoré sú jednoduchými dôsledkami jeho predpokladov a sú uvádzané takmer v každej literatúre o zmiešaných modeloch. Definíciu môžeme nájsť v Laird a Ware (1982).

Pri práci s lineárnym zmiešaným modelom predpokladáme, že sledujeme N nezávislých subjektov. Označme $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^\top$ pre $i = 1, \dots, N$. Ďalej \mathbb{X}_i je matica regresorov typu $n_i \times p$ a \mathbb{Z}_i matica regresorov typu $n_i \times q$ taká, že lineárny priestor generovaný stĺpcami matice \mathbb{X}_i je nadpriestorom priestoru generovaného stĺpcami matice \mathbb{Z}_i . Ďalej budeme pozorovať realizácie náhodných objektov $\mathbf{Y}_i, \mathbb{X}_i, i = 1, \dots, N$.

Definícia 1. *Hovoríme, že vektory $\mathbf{Y}_1, \dots, \mathbf{Y}_N$ splňujú lineárny zmiešaný model (LME model), ak existuje p -rozmerný vektor konštánt $\boldsymbol{\beta}$ a q -rozmerné náhodné vektory \mathbf{b}_i také, že*

$$\begin{aligned} \mathbf{b}_i &\sim \mathcal{N}_q(\mathbf{0}, \mathbb{D}), \\ \mathbf{Y}_i | \mathbf{b}_i, \mathbb{X}_i &\sim \mathcal{N}_{n_i}(\mathbb{X}_i \boldsymbol{\beta} + \mathbb{Z}_i \mathbf{b}_i, \sigma_e^2 \mathbb{I}_{n_i}), \end{aligned}$$

kde $\sigma_e^2 > 0$ je reálna konštanta a \mathbb{D} konečná pozitívne definitná matica typu $q \times q$.

Poznámka. Matice \mathbb{Z}_i nemusíme uvažovať ako náhodné, pretože k nim pristupujeme tak, že sú odvodené zo stĺpcov matíc \mathbb{X}_i . Preto podmieňujeme len hodnotami matíc \mathbb{X}_i a z ich hodnôt vieme určiť \mathbb{Z}_i , teda aj podmienenú strednú hodnotu \mathbf{Y}_i pre $i = 1, \dots, N$.

Z predpokladu normality v Definícii 1 a vlastností podmieneného rozdelenia sa priamočiaro ukáže, že marginálne rozdelenia veličín Y_{ij} vektora \mathbf{Y}_i splňujúceho LME model sú normálne so strednou hodnotou $\mathbf{x}_{ij}^\top \boldsymbol{\beta}$, kde \mathbf{x}_{ij}^\top je riadok matice \mathbb{X}_i na mieste j pre $j = 1, \dots, n_i$ a $i = 1, \dots, N$. Variančná štruktúra je zase daná rozptylovou maticou vektora \mathbf{Y}_i . Tú vieme určiť jednoduchým výpočtom podmieneného rozptylu. Rozptylovú maticu vektora značíme symbolom Var .

$$\boldsymbol{\Sigma}_i \equiv \text{Var}[\mathbf{Y}_i | \mathbb{X}_i] = \text{Var}[\mathbf{E}[\mathbf{Y}_i | \mathbf{b}_i] | \mathbb{X}_i] + \mathbf{E}[\text{Var}[\mathbf{Y}_i | \mathbf{b}_i] | \mathbb{X}_i].$$

Keďže $\boldsymbol{\beta}$ je konštanta a \mathbf{b}_i sú náhodné, platí

$$\text{Var}[\mathbf{E}[\mathbf{Y}_i | \mathbf{b}_i] | \mathbb{X}_i] = \text{Var}[\mathbb{X}_i \boldsymbol{\beta} + \mathbb{Z}_i \mathbf{b}_i | \mathbb{X}_i] = \mathbb{Z}_i \mathbb{D} \mathbb{Z}_i^\top$$

a σ_e^2 je taktiež konštantný parameter, preto

$$\mathbf{E}[\text{Var}[\mathbf{Y}_i | \mathbf{b}_i] | \mathbb{X}_i] = \mathbf{E}[\sigma_e^2 \mathbb{I}_{n_i} | \mathbb{X}_i] = \sigma_e^2 \mathbb{I}_{n_i}.$$

Takže pre rozptylovú maticu platí $\boldsymbol{\Sigma}_i = \sigma_e^2 \mathbb{I}_{n_i} + \mathbb{Z}_i \mathbb{D} \mathbb{Z}_i^\top$ a na LME sa môžeme pozeráť aj marginálne, a síce

$$\mathbf{Y} | \mathbb{X} \sim \mathcal{N}_n(\mathbb{X} \boldsymbol{\beta}, \boldsymbol{\Sigma}), \quad (1.1)$$

kde $\mathbf{Y} = (\mathbf{Y}_1^\top, \dots, \mathbf{Y}_N^\top)^\top, n = n_1 + \dots + n_N$, matica \mathbb{X} je typu $n \times p$ tvorená maticami \mathbb{X}_i usporiadanými pod sebou a matica $\boldsymbol{\Sigma}$ je blokovo diagonálna s blokmi $\boldsymbol{\Sigma}_i, i = 1, \dots, N$.

Poznámka. Požadovanie toho, aby stĺpce matice \mathbb{Z}_i boli obsiahnuté v stĺpcovom priestore matice \mathbb{X}_i je technický predpoklad. Dal by sa obísť, ak by sme pripustili nenulovú strednú hodnotu v rozdelení vektorov \mathbf{b}_i . Rozdelenie regresných matíc pre nás nebude kľúčové, väčšinou budeme pracovať s podmieneným rozdelením pri daných hodnotách týchto matíc. Dôležitý je len predpoklad, aby ich rozdelenie nebolo závislé na parametroch modelu.

V našom značení sa môžeme na situáciu pozeráť tak, že odhadujeme regresný model \mathbf{Y}_i v závislosti na regresoroch \mathbb{U}_i . Predpokladáme, že $\mathbf{E}[\mathbf{Y}_i|\mathbb{U}_i] = \mathbb{U}_i\boldsymbol{\gamma}_i$, pričom niektoré zložky parametra $\boldsymbol{\gamma}_i$ sú rovnaké pre všetky $i = 1, \dots, N$. Označme ich $\boldsymbol{\gamma}^{(1)}$. Ostatné zložky parametrov $\boldsymbol{\gamma}_i$ sa líšia pre rôzne i , označíme $\boldsymbol{\gamma}_i^{(2)}$. Môžeme si predstaviť, že pochádzajú z nejakého rozdelenia so strednou hodnotou $\boldsymbol{\gamma}^{(2)}$ a istou variabilitou. Pre každé i potom ale existuje $\boldsymbol{\gamma}_i^*$ tak, že $\boldsymbol{\gamma}_i^{(2)} = \boldsymbol{\gamma}^{(2)} + \boldsymbol{\gamma}_i^*$. V kontexte LME modelu následne máme $\boldsymbol{\beta} = (\boldsymbol{\gamma}^{(1)}, \boldsymbol{\gamma}^{(2)})$ a nazývame ich *pevné efekty*, tie predstavujú priemerný efekt v populácii im odpovedajúceho regresora. Vektory $\mathbf{b}_i = \boldsymbol{\gamma}_i^*$, ktoré označujeme ako *náhodné efekty*, sa snažia zachytiť variabilitu efektu okolo jeho priemernej hodnoty.

Oproti klasickému lineárnemu modelu uvažuje LME model odlišnú variančnú štruktúru, konkrétne skupinovo závislé dáta namiesto homoskedasticity klasického lineárneho modelu. Podľa LME modelu sú ľubovoľné dve zložky vektorov \mathbf{Y}_i a \mathbf{Y}_j pre $i \neq j$ nekorelované, kým dve zložky toho istého vektora \mathbf{Y}_i korelované byť môžu.

Tvrdenie 1. *V LME modeli platia vzťahy*

$$\begin{aligned} \text{cov}(Y_{ij}, Y_{kl}) &= 0, & i \neq k, \\ \text{cov}(Y_{ij}, Y_{ik}) &= \mathbf{z}_{ij}^\top \mathbb{D} \mathbf{z}_{ik}, & j \neq k, \\ \text{var } Y_{ij} &= \sigma_e^2 + \mathbf{z}_{ij}^\top \mathbb{D} \mathbf{z}_{ij}, \end{aligned}$$

kde \mathbf{z}_{ij}^\top je riadok matice \mathbb{Z}_i na mieste j .

Dôkaz. Vzťahy vyplývajú jednoducho zo štruktúr matice $\boldsymbol{\Sigma}$, resp. matíc $\boldsymbol{\Sigma}_i$ pre $i = 1, \dots, N$. Stačí zistiť, čo sa nachádza na konkrétnych miestach matíc $\boldsymbol{\Sigma}_i$. \square

Parametre, ktoré popisujú LME model a potrebujeme ich odhadnúť sú jednak pevné efekty (vektor $\boldsymbol{\beta}$) a taktiež parametre, ktoré popisujú variančnú štruktúru (parameter σ_e^2 a matica \mathbb{D}). Predpokladáme, že matica \mathbb{D} je pozitívne definitná, teda existuje jej Choleského rozklad. Stačí uvažovať parametre matice z rozkladu, ktorých je najviac $[q(q+1)]/2$ a budeme ich niekedy značiť $\boldsymbol{\psi}$.

K výberu vhodných regresorov do modelu môžeme podobne ako v klasickom lineárnom modeli pristúpiť štatistickými testami na podmodel a model tak, že vynecháme niektoré pevné efekty vo väčšom modeli. V literatúre, napríklad v publikácii Henderson (1984), sú tieto testy odvodené. Sú však citlivé na porušenie zvolenej variančnej štruktúry, ktorá je daná parametrami $\boldsymbol{\psi}$. Preto je dôležitá voľba regresnej matice \mathbb{Z}_i pre náhodné efekty a schopnosť odhadovať variančné parametre.

Metódou maximálnej vierohodnosti sa dajú odvodiť testy na porovnanie podmodelu a modelu s rozdielnou štruktúrou náhodných efektov a rovnakou

štruktúrou pevných efektov. Tie však testujú parametre na kraji parametrického priestoru, čím porušujú predpoklady asymptotických vlastností teórie maximálnej vierohodnosti a musia sa upravovať.

Možným prístupom, ako sa pri výbere modelu vyhnúť vyššie uvedeným testom, môže byť porovnávanie predikčných schopností LME modelu. Predikciou rozumieme odhadovanie hodnôt náhodných veličín, resp. vektorov, ktoré vznikajú predpokladaným pravdepodobnostným mechanizmom. Chybou predikcie budeme rozumieť kvantifikáciu toho, ako ďaleko sme s predikciou od skutočných hodnôt, ktoré mechanizmus generuje. Výber medzi dvomi alebo viacerými modelmi potom môžeme previesť na výber modelu s najnižšou chybou predikcie.

1.1 Predikcia marginálne a podmienene

Na ohodnotenie predikčných schopností LME modelu potrebujeme zaviesť pojem predikcie. Lineárny zmiešaný model je štrukturovaný hierarchicky, preto budeme rozlišovať typy predikcie podľa úrovne, na ktorej sa snažíme predikovať. Teraz vysvetlíme, čo budeme označovať ako replikované dáta, ktoré môžeme následne predikovať.

Predpokladajme, že súbor $\mathbf{Y}_i, \mathbb{X}_i, \mathbb{Z}_i, i = 1, \dots, N$ splňuje LME model a dáta sú konkrétnou realizáciou náhodného mechanizmu, ktorý je popísaný štruktúrou modelu. Označme realizované hodnoty regresných matíc $\mathbf{x}_1, \dots, \mathbf{x}_N$ pre matice pevných efektov a z nich odvodené matice náhodných efektov $\mathbf{z}_1, \dots, \mathbf{z}_N$. Ďalej budeme predpokladať, že môžeme znova pozorovať proces generovaný rovnakým náhodným mechanizmom a obdržíme nové hodnoty odozvy \mathbf{Y}_i^* , regresných matíc \mathbb{X}_i^* (resp. odvodené \mathbb{Z}_i^*) a náhodných efektov \mathbf{b}_i^* pre $i = 1, \dots, N$. Všetky tieto náhodné objekty sú nezávislé na predošlých pozorovaných vektoroch a maticiach.

Definícia 2.

(1) *Náhodný vektor*

$$\mathbf{Y}^* = \begin{pmatrix} \mathbf{Y}_1^* \\ \vdots \\ \mathbf{Y}_N^* \end{pmatrix},$$

kde $\mathbf{Y}_i^* = (Y_{i1}^*, \dots, Y_{in_i}^*)^\top$ nazveme replikované dáta, ak ich hodnoty sú konkrétnou realizáciou podmieneného rozdelenia $\mathbf{Y}_i^* | \mathbb{X}_i^* = \mathbf{x}_i$.

(2) *Vektor $\widehat{\mathbf{Y}} = (\widehat{\mathbf{Y}}_1^\top, \dots, \widehat{\mathbf{Y}}_N^\top)^\top$, kde $\widehat{\mathbf{Y}}_i = (\widehat{Y}_{i1}, \dots, \widehat{Y}_{in_i})^\top$, nazveme predikciou replikovaných dát \mathbf{Y}^* , ak je nejakou funkciou pozorovaných hodnôt $\mathbf{Y}_1, \dots, \mathbf{Y}_N$.*

Poznámka. Replikované dáta $\mathbf{Y}_1^*, \dots, \mathbf{Y}_N^*$ sa oproti dátovému súboru $\mathbf{Y}_i, \mathbb{X}_i, \mathbb{Z}_i$ líšia v realizácii náhodných efektov a realizácii náhodných členov pochádzajúcich z podmieneného rozdelenia \mathbf{Y}^* pri známych hodnotách náhodných efektov. Môžeme sa na to pozerať tak, že pozorujeme nezávislú realizáciu vektorov náhodných efektov $\mathbf{b}_i^* \sim \mathcal{N}_q(\mathbf{0}, \mathbb{D})$ a následne nezávisle na \mathbf{Y}_i a \mathbf{b}_i vektory \mathbf{Y}_i^* s podmieneným rozdelením

$$\mathbf{Y}_i^* | \mathbf{b}_i^*, \mathbb{X}_i^* = \mathbf{x}_i \sim \mathcal{N}_{n_i}(\mathbf{x}_i \boldsymbol{\beta} + \mathbf{z}_i \mathbf{b}_i^*, \sigma_e^2 \mathbb{I}_{n_i}).$$

Jedným možným prístupom predikcie hodnôt replikovaných dát, je predikovanie na marginálnej úrovni. Pri tomto type predikcie vychádzame z marginálneho rozdelenia vektora \mathbf{Y} , zo vzťahu (1.1). Vhodnou predikciou môže byť odhad strednej hodnoty

$$\widehat{\mathbf{Y}}^M = \mathbb{X}\tilde{\boldsymbol{\beta}}, \quad (1.2)$$

kde $\tilde{\boldsymbol{\beta}}$ je ľubovoľný odhad parametra $\boldsymbol{\beta}$, ktorý závisí len na dátach. Najčastejšie budeme v tomto prípade pracovať s najlepším lineárnym nestranným odhadom $\hat{\boldsymbol{\beta}}$, ktorý je explicitne popísaný v časti 1.2. Pripomíname naviac, že matica \mathbb{X} je tvorená maticami \mathbb{X}_i usporiadanými pod seba.

Druhý typ predikcie je podmienený a vychádza z podmienenej strednej hodnoty pri znalosti náhodných efektov. V tomto prípade sa však potrebujeme vysporiadať s hodnotami vektorov $\mathbf{b}_1, \dots, \mathbf{b}_N$, ktoré nepozorujeme. Často ich nazývame aj skryté (latentné) dáta. Hodnoty $\mathbf{b}_1, \dots, \mathbf{b}_N$ musíme preto taktiež predikovať. Ľubovoľnú funkciu s hodnotami $\hat{\mathbf{b}} = (\hat{\mathbf{b}}_1^\top, \dots, \hat{\mathbf{b}}_N^\top)^\top$, ktorá závisí na hodnotách \mathbf{Y}_i , \mathbb{X}_i a \mathbb{Z}_i nazveme predikciou (skrytých) náhodných efektov. Za predikciu odozvy potom berieme

$$\widehat{\mathbf{Y}}^P = \begin{pmatrix} \widehat{\mathbf{Y}}_1^P \\ \vdots \\ \widehat{\mathbf{Y}}_N^P \end{pmatrix} = \begin{pmatrix} \mathbb{X}_1\tilde{\boldsymbol{\beta}} + \mathbb{Z}_1\hat{\mathbf{b}}_1 \\ \vdots \\ \mathbb{X}_N\tilde{\boldsymbol{\beta}} + \mathbb{Z}_N\hat{\mathbf{b}}_N \end{pmatrix} = \mathbb{X}\tilde{\boldsymbol{\beta}} + \mathbb{Z}\hat{\mathbf{b}}, \quad (1.3)$$

kde matica \mathbb{Z} je blokovo diagonálna s blokmi \mathbb{Z}_i a na mieste odhadu $\tilde{\boldsymbol{\beta}}$ budeme najčastejšie uvažovať $\hat{\boldsymbol{\beta}}$ z časti 1.2 rovnako ako v marginálnom prípade.

1.2 BLUE a BLUP

Ďalej uvedieme príklad predikcie, ktorá sa pri práci s LME modelom používa najčastejšie, pretože je v istom zmysle optimálna. Na odvodení výpočtu a vlastností sá najviac podieľal americký matematik C. R. Henderson od 50. rokov 20. storočia, čo následne zhrnul v publikácii Henderson (1984), z ktorej prevezmeme niektoré výsledky.

V tretej kapitole jeho publikácie uvádza postup odvodenia tzv. Hendersonových odhadovacích rovníc, ktoré v prípade modelu z Definície 1 sú v tvare

$$\begin{pmatrix} \mathbb{X}^\top\mathbb{X} & \mathbb{X}^\top\mathbb{Z} \\ \mathbb{Z}^\top\mathbb{X} & \mathbb{Z}^\top\mathbb{Z} + \sigma_e^2\mathbb{D}_*^{-1} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{b}} \end{pmatrix} = \begin{pmatrix} \mathbb{X}^\top\mathbf{Y} \\ \mathbb{Z}^\top\mathbf{Y} \end{pmatrix}, \quad (1.4)$$

kde \mathbb{D}_* je diagonálne bloková matica s blokmi \mathbb{D} , čo sú rozptylové matice náhodných efektov a matice \mathbb{X} a \mathbb{Z} sú rovnaké ako vo výrazoch (1.1), resp. (1.3). Vektor \mathbf{Y} je vektor všetkých pozorovaní odozvy. Riešením Hendersonových rovníc získame odhady pevných efektov a predikcie náhodných efektov. V kapitolách 3 a 5 svojej publikácie ukazuje Henderson tvar riešenia

$$\hat{\boldsymbol{\beta}} = (\mathbb{X}^\top\boldsymbol{\Sigma}^{-1}\mathbb{X})^{-1}(\mathbb{X}^\top\boldsymbol{\Sigma}^{-1}\mathbf{Y}), \quad (1.5)$$

$$\hat{\mathbf{b}} = \mathbb{D}_*\mathbb{Z}^\top\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbb{X}\hat{\boldsymbol{\beta}}), \quad (1.6)$$

kde $\boldsymbol{\Sigma}$ je variančná matica marginálneho rozdelenia vektora \mathbf{Y} , viď vzťah (1.1).

Tvar odhadu (1.5) odpovedá teórii lineárnych modelov so všeobecnou rozptylovou maticou. Odhad sa odvodí pomocou metódy takzvaných vážených najmenších štvorcov. Rovnosť (1.6) ukazuje jeden možný príklad predikcie náhodných efektov v LME modeli. S jej využitím taktiež dostávame jeden príklad predikcie chápanej podmienene, a to dosadením odhadu $\hat{\boldsymbol{\beta}}$ a predikcie $\hat{\mathbf{b}}$ do výrazu (1.3).

Ďalej vysvetlíme, v akom zmysle je predikcia náhodných efektov optimálna. Jej odvodenie v Henderson (1984) vychádza z hľadania najlepšej lineárnej nestrannej predikcie (BLUP) náhodných efektov a súčasne hľadania najlepšieho lineárneho nestranného odhadu (BLUE) pevných efektov. Podstatné vlastnosti odtiaľ prevezmeme a zhrnieme v nasledujúcej vete. Označme najskôr inverz matice, ktorá sa vyskytuje v Hendersonových odhadovacích rovniciach (1.4).

$$\mathbb{V} = \begin{pmatrix} \mathbb{V}_{11} & \mathbb{V}_{12} \\ \mathbb{V}_{21} & \mathbb{V}_{22} \end{pmatrix} = \sigma_e^2 \begin{pmatrix} \mathbb{X}^\top \mathbb{X} & \mathbb{X}^\top \mathbb{Z} \\ \mathbb{Z}^\top \mathbb{X} & \mathbb{Z}^\top \mathbb{Z} + \sigma_e^2 \mathbb{D}_*^{-1} \end{pmatrix}^{-1}.$$

Veta 2. *Za predpokladu, že matica \mathbb{D}_* a parameter σ_e^2 sú známe, platí*

(i) $\hat{\mathbf{b}}$ je najlepšia lineárna nestranná predikcia vektora \mathbf{b} , t.j.

$$\forall \mathbf{c} \in \mathbb{R}^{q^*} : \hat{\mathbf{b}} = \arg \min_{\tilde{\mathbf{b}}} \mathbb{E}(\mathbf{c}^\top \tilde{\mathbf{b}} - \mathbf{c}^\top \mathbf{b})^2,$$

kde $q^* = Nq$ a $\tilde{\mathbf{b}}$ je ľubovoľná lineárna funkcia \mathbf{Y} s nulovou strednou hodnotou.

(ii) $\hat{\mathbf{b}}$ maximalizuje $\text{cor}(\tilde{\mathbf{b}}, \mathbf{b})$ na množine všetkých lineárnych nestranných predikcií $\tilde{\mathbf{b}}$.

(iii) $\mathbb{E}[\mathbf{b} | \hat{\mathbf{b}}] = \hat{\mathbf{b}}$.

(iv) $\text{var } \mathbb{L}\hat{\boldsymbol{\beta}} = \mathbb{L} \left(\mathbb{X}^\top \boldsymbol{\Sigma}^{-1} \mathbb{X} \right)^{-1} \mathbb{L}^\top$.

(v) $\text{cov}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}) = \mathbf{0}$, $\text{cov}(\hat{\boldsymbol{\beta}}, \mathbf{b}) = -\mathbb{V}_{21}$.

(vi) $\text{var } \hat{\mathbf{b}} = \text{cov}(\hat{\mathbf{b}}, \mathbf{b}) = \mathbb{D}_* - \mathbb{V}_{22}$.

(vii) $\text{var}(\hat{\mathbf{b}} - \mathbf{b}) = \mathbb{V}_{22}$.

Dôkaz. Vid Henderson (1984), kapitola 5. □

Niektoré z výsledkov Vety 2 následne použijeme pri odvodzovaní chyby predikcie voči kvadratickej stratovej funkcii. Chybu predikcie sme ešte nedefinovali, avšak v bode (i) predchádzajúcej vety vidíme jednu z možností, s ktorou budeme pracovať. Tvrdí sa tam, že $\hat{\mathbf{b}}$ má najnižšiu chybu predikcie voči kvadratickej stratovej funkcii zo všetkých lineárnych funkcií \mathbf{Y} s nulovou strednou hodnotou.

2. Chyba predikcie

V nasledujúcej časti ukážeme vybrané možnosti, ktorými môžeme určovať chybu predikcie, aby sme následne mohli vybrať model s najnižšou chybou. Ukáže sa, že záleží aj na presnosti odhadovaných parametrov modelu, pretože chyba predikcie bude závisieť predovšetkým na správnom vysvetlení variability navrhnutým modelom.

2.1 Stredná štvorcová chyba predikcie

Predstava o chybe predikcie je schopnosť číselne vyjadriť, ako ďaleko sme s predikciou od skutočných hodnôt odozvy. Naše dáta predstavujú iba jednu realizáciu, preto budeme pracovať so strednou hodnotou. Zvolíme kvadratickú stratovú funkciu a celkovú chybu predikcie budeme počítat ako súčet chýb jednotlivých zložiek. Pripomeňme z predchádzajúcej kapitoly, že \mathbf{x}_i a \mathbf{z}_i pre $i = 1, \dots, N$ označujú konkrétnu realizáciu regresných matíc pozorovaných v dátovom súbore.

Definícia 3. Pre predikciu $\widehat{\mathbf{Y}} = (\widehat{\mathbf{Y}}_1^\top, \dots, \widehat{\mathbf{Y}}_N^\top)^\top$, kde $\widehat{\mathbf{Y}}_i = (\widehat{Y}_{i1}, \dots, \widehat{Y}_{in_i})^\top$ pre $i = 1, \dots, N$, definujeme strednú štvorcovú chybu predikcie (MSEP) vzťahom

$$\text{MSEP}(\widehat{\mathbf{Y}}) = \sum_{i=1}^N \sum_{j=1}^{n_i} \mathbb{E} \left[(Y_{ij}^* - \widehat{Y}_{ij})^2 \mid \mathbb{X}_i = \mathbb{X}_i^* = \mathbf{x}_i \right], \quad (2.1)$$

kde strednú hodnotu uvažujeme voči rozdeleniu vektora $(\mathbf{Y}^\top, (\mathbf{Y}^*)^\top)^\top$ za podmienky, že hodnoty regresných matíc replikovaných dát sú rovnaké ako pri pozorovaných dátach v súbore.

Poznámka. Stĺpce matice \mathbb{Z}_i sú odvodené zo stĺpcov matice \mathbb{X}_i . Jav $[\mathbb{X}_i = \mathbb{X}_i^* = \mathbf{x}_i]$ je ekvivalentný javu $[\mathbb{X}_i = \mathbb{X}_i^* = \mathbf{x}_i, \mathbb{Z}_i = \mathbb{Z}_i^* = \mathbf{z}_i]$. Ak budeme pracovať s podmienenou strednou hodnotou, kde podmieňujeme týmto javom, budeme skrátene používať \mathbb{E}^* .

Súčet druhých mocnín zložiek vektora je rovný druhej mocnine Euklidovskej normy, preto môžeme rovnosť (2.1) pomocou zavedeného značenia \mathbb{E}^* prepísať do tvaru

$$\text{MSEP}(\widehat{\mathbf{Y}}) = \sum_{i=1}^N \mathbb{E}^* \left[(\mathbf{Y}_i^* - \widehat{\mathbf{Y}}_i)^\top (\mathbf{Y}_i^* - \widehat{\mathbf{Y}}_i) \right]. \quad (2.2)$$

Vychádzajme zo vzťahu (1.1) o marginálnom rozdelení, podľa ktorého je $\mathbf{x}_i \boldsymbol{\beta}$ podmienená stredná hodnota \mathbb{E}^* vektorov \mathbf{Y}_i a \mathbf{Y}_i^* . Ďalej berieme do úvahy predikcie, pre ktoré je $\mathbf{x}_i \boldsymbol{\beta}$ ich stredná hodnota. Oba špeciálne prípady $\widehat{\mathbf{Y}}^M$ a $\widehat{\mathbf{Y}}^P$ to splňujú, ak pracujeme s BLUE a BLUP parametra $\boldsymbol{\beta}$ a vektorov \mathbf{b}_i z časti 1.2. Strednú hodnotu preto pričítame a odčítame v zátvorkách vzťahu (2.2) a máme

$$\text{MSEP}(\widehat{\mathbf{Y}}) = \sum_{i=1}^N \mathbb{E}^* \left[(\mathbf{Y}_i^* \mp \mathbf{x}_i \boldsymbol{\beta} - \widehat{\mathbf{Y}}_i)^\top (\mathbf{Y}_i^* \mp \mathbf{x}_i \boldsymbol{\beta} - \widehat{\mathbf{Y}}_i) \right].$$

Upravíme obe zátvorky do tvaru

$$\mathbf{Y}_i^* \mp \mathbf{x}_i \boldsymbol{\beta} - \widehat{\mathbf{Y}}_i = (\mathbf{Y}_i^* - \mathbf{x}_i \boldsymbol{\beta}) - (\widehat{\mathbf{Y}}_i - \mathbf{x}_i \boldsymbol{\beta}). \quad (2.3)$$

Ďalším krokom je výpočet súčinu zátvoriek po úprave a určenie podmienenej strednej hodnoty. Vektory $\widehat{\mathbf{Y}}_i$ sú predikcie, závisia len na hodnotách vektora \mathbf{Y} , ktorý je podľa definície replikovaných dát nezávislý s vektorom \mathbf{Y}^* . Z toho dôvodu je $\mathbf{E}^* (\mathbf{Y}_i^* - \mathbf{x}_i \boldsymbol{\beta})^\top (\widehat{\mathbf{Y}}_i - \mathbf{x}_i \boldsymbol{\beta}) = 0$ a platí

$$\begin{aligned} \text{MSEP}(\widehat{\mathbf{Y}}) &= \sum_{i=1}^N \mathbf{E}^* \left[(\mathbf{Y}_i^* - \mathbf{x}_i \boldsymbol{\beta})^\top (\mathbf{Y}_i^* - \mathbf{x}_i \boldsymbol{\beta}) \right] + \\ &\quad + \sum_{i=1}^N \mathbf{E}^* \left[(\widehat{\mathbf{Y}}_i - \mathbf{x}_i \boldsymbol{\beta})^\top (\widehat{\mathbf{Y}}_i - \mathbf{x}_i \boldsymbol{\beta}) \right]. \end{aligned}$$

Oba členy v predošlej rovnosti majú skalárny súčin rovnakého vektora s nulovou strednou hodnotou, preto výsledkom strednej hodnoty je súčet podmienených rozptylov, čo môžeme napísať aj ako stopu (značíme tr) variančnej matice. Použijeme rovnaké značenie Var^* pri podmieňovaní javom ako v \mathbf{E}^* . Pre strednú štvorcovú chybu teda platí

$$\text{MSEP}(\widehat{\mathbf{Y}}) = \sum_{i=1}^N \text{tr}(\text{Var}^* \mathbf{Y}_i^*) + \sum_{i=1}^N \text{tr}(\text{Var}^* \widehat{\mathbf{Y}}_i). \quad (2.4)$$

Naviac pre prvý člen na pravej strane rovnosti (2.4) platí

$$\sum_{i=1}^N \text{tr}(\text{Var}^* \mathbf{Y}_i^*) = \sum_{i=1}^N \text{tr}(\boldsymbol{\Sigma}_i) = n\sigma_e^2 + \text{tr}(\mathbb{Z}\mathbb{D}_*\mathbb{Z}^\top), \quad (2.5)$$

kde druhá rovnosť vychádza z tvaru rozptylovej matice vektora \mathbf{Y}_i^* , a síce máme $\boldsymbol{\Sigma}_i = \sigma_e^2 \mathbb{I}_{n_i} + \mathbb{Z}_i \mathbb{D} \mathbb{Z}_i^\top$. Stačí pripomenúť, že $n = n_1 + \dots + n_N$ a matica \mathbb{Z} je blokovo diagonálna s blokmi $\mathbb{Z}_i, i = 1, \dots, N$, rovnako ako matica \mathbb{D}_* s blokmi \mathbb{D} , ktorých je celkom N .

2.1.1 MSEP podmienenej predikcie

Ďalej použitím Vety 2 určíme, ako vyzerá stredná štvorcová chyba pre podmienenú predikciu $\widehat{\mathbf{Y}}^P$. Pripomíname, že používame predikcie s dosadeným BLUE odhadom parametru $\boldsymbol{\beta}$ a BLUP predikciou náhodných efektov z predchádzajúcej kapitoly.

Tvrdenie 3. *Za predpokladu, že matica \mathbb{D}_* a parameter σ_e^2 sú známe, platí*

$$\text{MSEP}(\widehat{\mathbf{Y}}^P) = n\sigma_e^2 + 2 \text{tr}(\mathbb{Z}\mathbb{D}_*\mathbb{Z}^\top) + \text{tr}\left(\mathbb{X}(\mathbb{X}^\top \boldsymbol{\Sigma}^{-1} \mathbb{X})^{-1} \mathbb{X}^\top\right) - \text{tr}(\mathbb{Z}\mathbb{V}_{22}\mathbb{Z}^\top),$$

kde n predstavuje rozmer vektora \mathbf{Y} v LME modeli a zvyšné matice sú zavedené v prvej kapitole.

Dôkaz. Vychádzame zo vzťahu (2.4), kde chceme vypočítať druhý člen na pravej strane pre podmienenú predikciu. Použijeme rovnosť, ktorá definuje predikciu, $\widehat{\mathbf{Y}}_i^P = \mathbb{X}_i \widehat{\boldsymbol{\beta}} + \mathbb{Z}_i \widehat{\mathbf{b}}_i$ a bod (v) Vety 2 o nulovej kovariancii odhadu a predikcie pevných a náhodných efektov. Preto platí

$$\begin{aligned} \sum_{i=1}^N \text{tr}(\text{Var}^* \widehat{\mathbf{Y}}_i^P) &= \sum_{i=1}^N \text{tr}(\text{Var}^*(\mathbb{X}_i \widehat{\boldsymbol{\beta}} + \mathbb{Z}_i \widehat{\mathbf{b}}_i)) = \\ &= \sum_{i=1}^N \text{tr}(\text{Var}^* \mathbb{X}_i \widehat{\boldsymbol{\beta}}) + \sum_{i=1}^N \text{tr}(\text{Var}^* \mathbb{Z}_i \widehat{\mathbf{b}}_i). \end{aligned}$$

Obe sumy sčítame podobným spôsobom ako pri odvodení vzťahu (2.5) tak, že súčet stôp vyjadríme ako stopu matice s väčším rozmerom, ktorá ma na diagonále jednotlivé sčítance. V značení LME modelov z prvej kapitoly sa predošlá rovnosť preto upraví nasledovne

$$\sum_{i=1}^N \text{tr} \left(\text{Var}^* \widehat{\mathbf{Y}}_i^P \right) = \text{tr} \left(\mathbb{X} \left(\text{Var}^* \widehat{\boldsymbol{\beta}} \right) \mathbb{X}^\top \right) + \text{tr} \left(\mathbb{Z} \left(\text{Var}^* \widehat{\mathbf{b}} \right) \mathbb{Z}^\top \right). \quad (2.6)$$

Ďalej stačí dosadiť tvar rozptylovej matice odhadov pevných efektov, resp. predikcie náhodných efektov (body (iv) a (vi) vo Vete 2) a využiť linearitu operátora stopy. Rovnosť (2.6) sa upraví do tvaru

$$\sum_{i=1}^N \text{tr} \left(\text{Var}^* \widehat{\mathbf{Y}}_i^P \right) = \text{tr} \left(\mathbb{X} \left(\mathbb{X}^\top \boldsymbol{\Sigma}^{-1} \mathbb{X} \right)^{-1} \mathbb{X}^\top \right) + \text{tr} \left(\mathbb{Z} \mathbb{D}_* \mathbb{Z}^\top \right) - \text{tr} \left(\mathbb{Z} \mathbb{V}_{22} \mathbb{Z}^\top \right). \quad (2.7)$$

Sčítaním vzťahov (2.7) a (2.5) sa dokáže znenie tvrdenia. □

2.1.2 MSEP marginálnej predikcie

Pre marginálnu predikciu odvodíme obdobný vzťah na výpočet strednej štvorcovej chyby predikcie.

Tvrdenie 4. *Za predpokladu, že matica \mathbb{D}_* a parameter σ_e^2 sú známe, platí*

$$\text{MSEP} \left(\widehat{\mathbf{Y}}^M \right) = n\sigma_e^2 + \text{tr} \left(\mathbb{Z} \mathbb{D}_* \mathbb{Z}^\top \right) + \text{tr} \left(\mathbb{X} \left(\mathbb{X}^\top \boldsymbol{\Sigma}^{-1} \mathbb{X} \right)^{-1} \mathbb{X}^\top \right),$$

kde n predstavuje rozmer vektora \mathbf{Y} v LME modeli a zvyšné matice sú zavedené v prvej kapitole.

Dôkaz. Dôkaz tohto tvrdenia je obdobný ako dôkaz Tvrdenia 3. Rozdiel je len v definícii predikcie. Teraz máme $\widehat{\mathbf{Y}}_i^M = \mathbb{X}_i \widehat{\boldsymbol{\beta}}$, preto

$$\sum_{i=1}^N \text{tr} \left(\text{Var}^* \widehat{\mathbf{Y}}_i^M \right) = \text{tr} \left(\mathbb{X} \left(\text{Var}^* \widehat{\boldsymbol{\beta}} \right) \mathbb{X}^\top \right) = \text{tr} \left(\mathbb{X} \left(\mathbb{X}^\top \boldsymbol{\Sigma}^{-1} \mathbb{X} \right)^{-1} \mathbb{X}^\top \right). \quad (2.8)$$

Dosadením (2.8) a (2.5) z predošlého dôkazu do rovnosti (2.4) dostaneme znenie tohto tvrdenia. □

2.2 Penalizovanie kovarianciou

V oboch predchádzajúcich výpočtoch sme určili MSEP predikcie pri znalosti variančných parametrov. V praxi ich však nepoznáme, preto by sme neznáme hodnoty v zneniach tvrdení 3 a 4 museli odhadnúť. K tomu by sme mohli použiť napríklad postup metódou maximálnej vierohodnosti alebo jeho modifikáciu,

ktorá pracuje s tzv. restringovanou vierohodnosťou (REML). Túto modifikáciu môžeme napríklad nájsť v Patterson a Thompson (1971).

Ďalšia možnosť je postupovať pri výpočte MSEP iným spôsobom než v predošlej časti. Budeme vychádzať z článku Efron (2004), kde sa pracuje so všeobecnejšou triedou stratových funkcií a použijeme vetu o optimizme.

Definícia 4. Pre reálnu konkávnou funkciu q definujme príslušnú stratovú funkciu predikcie (resp. odhadu) $\hat{\mu}$ náhodnej veličiny Y vzťahom

$$Q(Y, \hat{\mu}) = q(\hat{\mu}) + q'(\hat{\mu})(Y - \hat{\mu}) - q(Y),$$

kde $q'(\hat{\mu})$ je derivácia funkcie q vyčíslená v odhade $\hat{\mu}$.

Poznámka. Takto definovaná stratová funkcia je dotyčnicou ku q v bode odhadu $\hat{\mu}$. Penalizuje odhad vzdialenosťou dotyčnice od hodnoty funkcie $q(Y)$ v skutočnej hodnote Y . Inak sa na to môžeme pozeráť, že za stratu berieme iba lineárnu aproximáciu skutočného rozdielu funkčných hodnôt q v bode odhadu a v bode skutočnej hodnoty náhodnej veličiny.

Celkovú chybu predikcie (resp. odhadu) vektora $\mathbf{Y} \in \mathbb{R}^n$ budeme počítat podobne ako pri MSEP, a síce určíme súčet chýb jednotlivých zložiek

$$Q(\mathbf{Y}, \hat{\boldsymbol{\mu}}) = \sum_{i=1}^n Q(Y_i, \hat{\mu}_i).$$

Použijeme značenie Efrona z článku pri všeobecnej stratovej funkcii

$$\text{err}_i = Q(Y_i, \hat{\mu}_i), \quad \text{Err}_i = \mathbf{E}_0 (Q(Y_i^*, \hat{\mu}_i)), \quad \hat{\lambda}_i = -q'(\hat{\mu}_i)/2,$$

kde Y_i^* predstavuje zložky replikovaného vektora dát, strednú hodnotu \mathbf{E}_0 rozumieme voči rozdeleniu replikovaných dát a navyše podmieňujeme hodnotou $\hat{\mu}_i$. S týmto značením môžeme sformulovať vetu o optimizme.

Veta 5. Pre stratovú funkciu $Q(\mathbf{Y}, \hat{\boldsymbol{\mu}})$ platí

$$\mathbf{E} \text{Err}_i = \mathbf{E} (\text{err}_i + \Omega_i),$$

kde

$$\Omega_i = 2 \text{cov}(\hat{\lambda}_i, Y_i).$$

Stredné hodnoty a kovarianciu chápeme vzhľadom k rozdeleniu vektora Y .

Dôkaz. Viď článok Efron (2004), Optimism Theorem 1. □

Poznámka. Označenie, veta o optimizme, vychádza z toho, že rozdiel $\text{Err}_i - \text{err}_i$ sa nazýva optimizmus. V znení vety máme vzťah na výpočet jeho strednej hodnoty.

Prepíšme značenie z článku do nášho kontextu LME modelu so štvorcovou chybou Q .

$$\text{err}_{ij} = (Y_{ij} - \hat{Y}_{ij})^2, \quad \text{Err}_{ij} = \mathbf{E}_0 (Y_{ij}^* - \hat{Y}_{ij})^2, \quad (2.9)$$

kde E_0 je podmienená stredná hodnota voči rovnakému rozdeleniu ako uvažujeme pri E^* s rozdielom, že podmieňujeme navyše hodnotou \widehat{Y} . Preto v rámci počítania E_0 môžeme považovať \widehat{Y}_{ij} za pevné. S takto zavedeným značením platí

$$\text{MSEP}(\widehat{Y}) = \sum_{i=1}^N \sum_{j=1}^{n_i} E^* \text{Err}_{ij}, \quad (2.10)$$

Na výpočet člena $E^* \text{Err}_{ij}$ použijeme vetu o optimizme zo spomínaného článku.

Tvrdenie 6. *Platí*

$$E^* \text{Err}_{ij} = E^* \left(\text{err}_{ij} + 2 \text{cov}(\widehat{Y}_{ij}, Y_{ij}) \right). \quad (2.11)$$

Dôkaz. Ide o špeciálny prípad Vety 2.2 o optimizme. Pracuje sa tam však so všeobecnejšou stratovou funkciou, preto ukážeme krátko, ako sa dá dokázať tvrdenie v tomto špeciálnom prípade. Pre prehľadnosť výpočtu označme strednú hodnotu marginálnej veličiny Y_{ij} ako μ_{ij} a ďalej upravíme Err_{ij} pričítaním a odčítaním strednej hodnoty

$$\text{Err}_{ij} = E_0 (Y_{ij}^* - \mu_{ij})^2 + E_0 (\widehat{Y}_{ij} - \mu_{ij})^2 - 2 E_0 (Y_{ij}^* - \mu_{ij})(\widehat{Y}_{ij} - \mu_{ij}).$$

Keďže μ_{ij} a \widehat{Y}_{ij} sú voči E_0 konštanty a $E_0 (Y_{ij}^* - \mu_{ij})^2 = E^* (Y_{ij} - \mu_{ij})^2$, môžeme upraviť na tvar

$$\text{Err}_{ij} = E^* (Y_{ij} - \mu_{ij})^2 + (\widehat{Y}_{ij} - \mu_{ij})^2 - 2 (\widehat{Y}_{ij} - \mu_{ij}) E^* (Y_{ij} - \mu_{ij}). \quad (2.12)$$

Na rovnosť (2.12) aplikujeme E^* a dostaneme

$$E^* \text{Err}_{ij} = E^* (Y_{ij} - \mu_{ij})^2 + E^* (\widehat{Y}_{ij} - \mu_{ij})^2 - 2 E^* (\widehat{Y}_{ij} - \mu_{ij}) E^* (Y_{ij} - \mu_{ij}). \quad (2.13)$$

Ďalej vypočítame strednú hodnotu err_{ij} , najprv pričítame a odčítame strednú hodnotu μ_{ij} .

$$E^* \text{err}_{ij} = E^* (Y_{ij} - \mu_{ij})^2 + E^* (\widehat{Y}_{ij} - \mu_{ij})^2 - 2 E^* (\widehat{Y}_{ij} - \mu_{ij})(Y_{ij} - \mu_{ij}). \quad (2.14)$$

Stačí použiť vzorec na výpočet kovariancie $\text{cov}(X, Y) = EXY - EX EY$ a odčítaním (2.14) od (2.13) dostaneme

$$E^* \text{Err}_{ij} - E^* \text{err}_{ij} = 2 \text{cov}(\widehat{Y}_{ij} - \mu_{ij}, Y_{ij} - \mu_{ij}),$$

kde kovarianciu chápeme vzhľadom k rozdeleniu pôvodných a replikovaných dát pri podmienke rovnakých regresorov \mathbb{X} podobne ako E^* . Posun o konštantu nemá na kovarianciu vplyv, preto je tvrdenie dokázané. \square

Výhoda tohto prístupu je v tom, že hodnoty err_{ij} závisia len na hodnotách predikcií a pozorovaných dát, ktoré máme k dispozícii. Preto môžeme jednoducho určiť

$$\text{err} = \sum_{i=1}^N \sum_{j=1}^{n_i} \text{err}_{ij} = \sum_{i=1}^N \sum_{j=1}^{n_i} (Y_{ij} - \widehat{Y}_{ij})^2. \quad (2.15)$$

Z podoby druhého člena v rovnosti (2.11) vychádza názov tohto prístupu, a síce penalizácia kovarianciou. Vidíme, že celková chyba predikcie sa skladá z dvoch zložiek. Prvá z nich, err , vyjadruje variabilitu pôvodných dát okolo predikcie. Druhú z nich, $\text{cov}(\widehat{Y}_{ij}, Y_{ij})$, musíme odhadovať. V článku Efron (2004) je uvedených niekoľko možností vo všeobecnom prípade. Teraz priblížime, ako sa dajú použiť pri práci s LME modelom.

2.2.1 Parametrický bootstrap

Prvý príklad odhadovania $\text{cov}(\hat{Y}_{ij}, Y_{ij})$ je parametrický bootstrap. Po odhadnutí parametrov LME modelu dostaneme parametrický odhad hustoty vektora \mathbf{Y} . Je to hustota normálneho $\mathcal{N}_n(\hat{\mathbf{Y}}, \hat{\mathbf{\Sigma}})$ rozdelenia. Vektor $\hat{\mathbf{Y}}$ je jedna z uvažovaných predikcií a $\hat{\mathbf{\Sigma}}$ je odhad rozptylovej matice, získaný dosadením odhadov variančných parametrov, $\hat{\sigma}_e^2$ a $\hat{\mathbb{D}}$, do vzťahu pre výpočet $\mathbf{\Sigma}_i$. Máme teda

$$\hat{\mathbf{\Sigma}} = \text{diag}(\hat{\mathbf{\Sigma}}_1, \dots, \hat{\mathbf{\Sigma}}_N), \quad \hat{\mathbf{\Sigma}}_i = \hat{\sigma}_e^2 \mathbb{I}_{n_i} + \mathbb{Z}_i \hat{\mathbb{D}} \mathbb{Z}_i^\top.$$

Potom prevedieme náhodný výber (nazývame bootstrapový výber) $\mathbf{Y}^{(1)}, \dots, \mathbf{Y}^{(B)}$ vektorov z rozdelenia $\mathcal{N}_n(\hat{\mathbf{Y}}, \hat{\mathbf{\Sigma}})$ a rovnakým spôsobom ako pri počítaní predikcie odhadneme ich strednú hodnotu za predpokladu, že splňujú LME model predpokladaný pre dátový súbor. Stačí dosadiť do vzorcov (1.2) alebo (1.3) vektor $\mathbf{Y}^{(b)}$ pre každé $b = 1, \dots, B$ v závislosti na tom, či sme na odhad hustoty použili marginálnu predikciu $\hat{\mathbf{Y}}^M$ alebo podmienenú $\hat{\mathbf{Y}}^P$.

Členy $\text{cov}(\hat{Y}_{ij}, Y_{ij}) \equiv \text{cov}_{ij}$ následne odhadneme pomocou výberovej kovariancie tak, že predikciu nahradíme bootstrapovým odhadom a skutočnú hodnotu zložkou vektora v bootstrapovom výbere.

$$\widehat{\text{cov}}_{ij} = \frac{1}{B-1} \sum_{b=1}^B \hat{Y}_{ij}^{(b)} (Y_{ij}^{(b)} - \bar{Y}_{ij}), \quad \bar{Y}_{ij} = \frac{1}{B} \sum_{b=1}^B Y_{ij}^{(b)}.$$

Celkovo tak dostaneme odhad MSEP predikcie $\hat{\mathbf{Y}}$ v tvare

$$\widehat{\text{MSEP}}(\hat{\mathbf{Y}}) = \text{err} + 2 \sum_{i=1}^N \sum_{j=1}^{n_i} \widehat{\text{cov}}_{ij}. \quad (2.16)$$

2.2.2 Neparametrický bootstrap

K problému sa dá pristúpiť aj bez použitia odhadov parametrov modelu. Napríklad pomocou neparametrického bootstrapu. Prevedieme náhodný výber s opakovaním z dátového súboru $\mathbf{Y}_i, \mathbb{X}_i, i = 1, \dots, N$. To znamená, že vyberieme množinu indexov $(i_1^{(b)}, \dots, i_N^{(b)})$, $b = 1, \dots, B$ takú, že prvky $i_k^{(b)}$ sú z množiny $\{1, \dots, N\}$ a môžu sa vo výbere opakovať. Za bootstrapový výber označíme dátový súbor $\mathbf{Y}_k^{(b)}, \mathbb{X}_k^{(b)}, k = i_1^{(b)}, \dots, i_N^{(b)}$. Na základe tohto výberu môžeme odhadnúť parameter β , prípadne pri práci s $\hat{\mathbf{Y}}^P$ aj predikcie náhodných efektov. Použijeme riešenie Hendersonových rovníc (1.4), kam dosadíme maticu $\mathbb{X}_{(b)}$ s p stĺpcami tvorenú pod sebou zoradenými maticami $\mathbb{X}_k^{(b)}$ pre $k = i_1^{(b)}, \dots, i_N^{(b)}$. Ďalej dosadíme maticu $\mathbf{\Sigma}_{(b)}$, ktorá je blokovo diagonálna s blokmi $\mathbf{\Sigma}_k^{(b)}$ obdobnými maticiam $\mathbf{\Sigma}_i$ z prvej kapitoly. Ich rozmery odpovedajú dátovému súboru, ktorý sme obdržali bootstrapovým výberom a nakoniec dosadíme aj bootstrapový výber $\mathbf{Y}_{(b)}$ zložený z vektorov $\mathbf{Y}_k^{(b)}$ zoradenými pod sebou.

Pomocou odhadu, prípadne predikcie náhodných efektov, následne určíme predikciu na základe bootstrapového výberu $\hat{\mathbf{Y}}^{(b)}$ výpočtom rovnosti (1.2) alebo (1.3) v závislosti na type predikcie. Pripomíname, že maticu \mathbb{X} v týchto rovnostiach použijeme z pôvodného dátového súboru, nie z bootstrapového výberu.

Pri tomto postupe sú bootstrapové výbery rôzne dlhé a niektoré pozorovania Y_{ij} sa v nich viackrát opakujú. Preto nemôžeme priamočiaro odhadnúť potrebný

člen výberovou kovarianciou. Vhodné je zohľadniť fakt, že v prípade, kde máme viackrát zopakované pozorovanie Y_{ij} , je predikcia $\hat{Y}_{ij}^{(b)}$ presnejšia a viac korelovaná s hodnotou $Y_{ij}^{(b)}$ než by sme očakávali od skutočnej závislosti medzi predikciou \hat{Y}_{ij} a pozorovaniami Y_{ij} . Počet opakovaní v bootstrapovom výbere zohľadníme tak, že odhadneme cov_{ij} váženým súčtom, ako je ukázané ďalej.

Označme $N_i^{(b)}$ počet, kolkokrát sa vyskytol index i v bootstrapovom výbere $(i_1^{(b)}, \dots, i_N^{(b)})$. Ďalej pre $h = 1, \dots, N$ buď $I_i^{(b)}(h) = \mathbb{1}\{N_i^{(b)} = h\}$ indikátor toho, že sa vo výbere vyskytuje práve h opakovaní $\mathbf{Y}_i, \mathbb{X}_i$. Na základe toho definujeme odhady

$$\widetilde{\text{cov}}_{ij}(h) = \frac{\sum_{b=1}^B I_i^{(b)}(h) \hat{Y}_{ij}^{(b)} Y_{ij}}{\sum_{b=1}^B I_i^{(b)}(h)},$$

ktoré odhadujú kovarianciu pre výbery, kde sa vyskytuje h opakovaní Y_{ij} . Konečný odhad definujeme ako vážený súčet týchto odhadov

$$\widetilde{\text{cov}}_{ij} = \sum_{h=1}^N w(h) \widetilde{\text{cov}}_{ij}(h),$$

kde $w(h) \geq 0$ môže byť ľubovoľne zvolená váhová funkcia, ktorá spĺňa

$$\sum_{h=1}^N w(h) = 1$$

a navyše očakávame, že bude klesajúca vzhľadom k vyššie uvedenému postrehu.

Poznámka. Používame iba nenulové h , pretože pri $h = 0$ sa v bootstrapovom výbere nevyskytuje Y_{ij} ani raz. Preto nepovažujeme tieto výskyty za vhodné na odhadovanie $\text{cov}(\hat{Y}_{ij}, Y_{ij})$.

Najjednoduchšou váhovou funkciou je pravdepodobnosť toho, že sa v opakovanom výbere z N prvkov vyskytne práve h opakovaní zvoleného prvku a ostatné budú rozdielne za predpokladu, že sa tam vyskytne aspoň raz.

$$w(h) = \left[\binom{N}{h} \frac{(N-1)^{N-h}}{N^N} \right] / \left[1 - \left(\frac{N-1}{N} \right)^N \right], \quad h = 1, \dots, N.$$

2.3 Krosvalidácia

Ďalšou možnosťou ako sa vysporiadať s odhadom chyby predikcie je pomocou krosvalidácie, ktorá sa v súvislosti s predikčnými chybami často používa. Dobré je v tomto prípade pracovať s výberom nezávislých vektorov. Preto sa zameriame na nasledovné vyjadrenie chyby predikcie

$$\text{MSEP}(\hat{\mathbf{Y}}) = \sum_{i=1}^N \mathbb{E}^* \left[\left(\mathbf{Y}_i^* - \hat{\mathbf{Y}}_i \right)^\top \left(\mathbf{Y}_i^* - \hat{\mathbf{Y}}_i \right) \right]. \quad (2.17)$$

Ďalej označme $\mathbf{Y}_{(i)}$ vektor pozorovaní, kde vynecháme \mathbf{Y}_i ,

$$\mathbf{Y}_{(i)} = (\mathbf{Y}_1^\top, \dots, \mathbf{Y}_{i-1}^\top, \mathbf{Y}_{i+1}^\top, \dots, \mathbf{Y}_N^\top)^\top.$$

Pri podmienovaní javom $[\mathbb{X}_i = \mathbb{X}_i^* = \mathbf{x}_i, \mathbf{Y}_{(i)} = \mathbf{y}_{(i)}]$ budeme značiť podmienenú strednú hodnotu symbolom $\mathbf{E}_{(i)}$. Vektor $\mathbf{y}_{(i)}$ predstavuje realizáciu vektora v našom dátovom súbore. Oproti \mathbf{E}^* sa líši tým, že navyše podmieňujeme hodnotou vektora $\mathbf{Y}_{(i)}$. Označme podmienenú kovarianciu medzi predikciou a pozorovanou hodnotou

$$\text{cov}_{(i)}(\widehat{Y}_{ij}, Y_{ij}) = \mathbf{E}_{(i)}\left(\widehat{Y}_{ij} - \mathbf{E}_{(i)}\widehat{Y}_{ij}\right)\left(Y_{ij} - \mathbf{E}_{(i)}Y_{ij}\right).$$

Tento postup je taktiež motivovaný článkom Efron (2004) a nasledujúce tvrdenie je tam uvedené bez dôkazu s odôvodnením, že ide iba o rafinovanejšiu verziu vety o optimizme.

Tvrdenie 7. *Platí*

$$\mathbf{E}_{(i)} \text{Err}_i = \mathbf{E}_{(i)}\left(\text{err}_i + 2 \sum_{j=1}^{n_i} \text{cov}_{(i)}(\widehat{Y}_{ij}, Y_{ij})\right),$$

kde

$$\text{Err}_i = \mathbf{E}_{(0)}\left[\left(\mathbf{Y}_i^* - \widehat{\mathbf{Y}}_i\right)^\top \left(\mathbf{Y}_i^* - \widehat{\mathbf{Y}}_i\right)\right], \quad \text{err}_i = \left(\mathbf{Y}_i - \widehat{\mathbf{Y}}_i\right)^\top \left(\mathbf{Y}_i - \widehat{\mathbf{Y}}_i\right) = \sum_{j=1}^{n_i} \text{err}_{ij}.$$

Stredná hodnota $\mathbf{E}_{(0)}$ je podmienená stredná hodnota rovnakým javom ako \mathbf{E}_ a navyše hodnotou predikcie $\widehat{\mathbf{Y}}_i$.*

Dôkaz. V špeciálnom prípade štvorcovej stratovej funkcie môže celý dôkaz prebiehať rovnako ako dôkaz Tvrdenia 6. Najprv pričítame a odčítame strednú hodnotu $\boldsymbol{\mu}_i = \mathbb{X}_i \boldsymbol{\beta}$ a potom postupne aplikujeme $\mathbf{E}_{(0)}$ na Err_i , pričom využijeme rovnosť stredných hodnôt

$$\mathbf{E}_{(0)}\left[\left(\mathbf{Y}_i^* - \boldsymbol{\mu}_i\right)^\top \left(\mathbf{Y}_i^* - \boldsymbol{\mu}_i\right)\right] = \mathbf{E}_{(i)}\left[\left(\mathbf{Y}_i - \boldsymbol{\mu}_i\right)^\top \left(\mathbf{Y}_i - \boldsymbol{\mu}_i\right)\right].$$

Máme preto

$$\begin{aligned} \text{Err}_i = \mathbf{E}_{(i)}\left[\left(\mathbf{Y}_i - \boldsymbol{\mu}_i\right)^\top \left(\mathbf{Y}_i - \boldsymbol{\mu}_i\right)\right] &+ \left(\widehat{\mathbf{Y}}_i - \boldsymbol{\mu}_i\right)^\top \left(\widehat{\mathbf{Y}}_i - \boldsymbol{\mu}_i\right) + \\ &+ 2 \left(\widehat{\mathbf{Y}}_i - \boldsymbol{\mu}_i\right)^\top \mathbf{E}_{(i)}\left(\mathbf{Y}_i - \boldsymbol{\mu}_i\right). \end{aligned}$$

Ďalej môžeme počítat $\mathbf{E}_{(i)}$

$$\begin{aligned} \mathbf{E}_{(i)} \text{Err}_i = \mathbf{E}_{(i)}\left[\left(\mathbf{Y}_i - \boldsymbol{\mu}_i\right)^\top \left(\mathbf{Y}_i - \boldsymbol{\mu}_i\right)\right] &+ \mathbf{E}_{(i)}\left[\left(\widehat{\mathbf{Y}}_i - \boldsymbol{\mu}_i\right)^\top \left(\widehat{\mathbf{Y}}_i - \boldsymbol{\mu}_i\right)\right] + \\ &+ 2 \mathbf{E}_{(i)}\left[\left(\widehat{\mathbf{Y}}_i - \boldsymbol{\mu}_i\right)^\top \mathbf{E}_{(i)}\left(\mathbf{Y}_i - \boldsymbol{\mu}_i\right)\right]. \end{aligned} \quad (2.18)$$

Podobne vypočítame strednú hodnotu err_i

$$\begin{aligned} \mathbf{E}_{(i)} \text{err}_i = \mathbf{E}_{(i)}\left[\left(\mathbf{Y}_i - \boldsymbol{\mu}_i\right)^\top \left(\mathbf{Y}_i - \boldsymbol{\mu}_i\right)\right] &+ \mathbf{E}_{(i)}\left[\left(\widehat{\mathbf{Y}}_i - \boldsymbol{\mu}_i\right)^\top \left(\widehat{\mathbf{Y}}_i - \boldsymbol{\mu}_i\right)\right] + \\ &+ 2 \mathbf{E}_{(i)}\left[\left(\widehat{\mathbf{Y}}_i - \boldsymbol{\mu}_i\right)^\top \left(\mathbf{Y}_i - \boldsymbol{\mu}_i\right)\right]. \end{aligned} \quad (2.19)$$

K dokončeniu dôkazu stačí odčítať rovnosti (2.19) a (2.18).

□

Ďalej by sme MSEP odhadovali obdobne ako pri metódach penalizácie kovarianciou. Takýto proces by preto potreboval pre každé $i = 1, \dots, N$ odhadnúť parametre LME modelu na základe hodnoty $\mathbf{Y}_{(i)}$ a príslušných regresných matíc \mathbb{X}_j , $j \in \{1, \dots, N\} \setminus \{i\}$. Pomocou odhadu regresného koeficientu, označme $\hat{\boldsymbol{\beta}}^{(i)}$, prípadne aj predikcie $\hat{\mathbf{b}}^{(i)}$ náhodných efektov určíme predikciu $\hat{\mathbf{Y}}_{(i)}$ odozvy pre subjekt i . V závislosti na type použitej predikcie v pôvodnom súbore vyberieme marginálnu alebo podmienenú predikciu aj pri krosvalidácii. Výpočet prevedieme použitím matíc dátového súboru \mathbb{X}_i , prípadne \mathbb{Z}_i a získaného odhadu, prípadne predikcie v rovnostiach (1.2) alebo (1.3). Potom môžeme odhadnúť

$$\widehat{\mathbf{E}}_{(i)\text{err}_i} = \sum_{j=1}^{n_i} (Y_{ij} - \hat{Y}_{(i)j})^2.$$

Na odhadovanie $\text{cov}_{(i)}(\hat{Y}_{ij}, Y_{ij})$ sa dajú použiť obe bootstrapové techniky podobným spôsobom ako pri penalizovaní kovarianciou. Pre každé $i = 1, \dots, N$ však potrebujeme náhodný výber z iného rozdelenia pri parametrickom bootstrape, čo zvyšuje náročnosť výpočtu.

3. Akaikovo informačné kritérium

Táto kapitola sa bude venovať ďalšiemu možnému prístupu, ako ohodnocovať predikčné vlastnosti modelu. Uvedieme Akaikovo informačné kritérium (AIC). Ide o jedno z viacerých informačných kritérií, ktoré boli navrhnuté na vyjadrenie predikčných vlastností modelu tak, aby modely s lepšími predikčnými schopnosťami mali nižšie hodnoty kritérií. Teda hľadanie vhodných modelov potom môžeme previesť na minimalizáciu týchto číselných hodnôt podobne ako pri predikčnej chybe. Hlavným zdrojom pre túto časť práce je článok Greven a Kneib (2010), ktorý prípadne doplníme čiastkovými výsledkami z ďalších zdrojov a vlastným odvodzovaním.

V článku Akaike (1974) je ukázané, ako môžeme využiť vierohodnosť pri porovnávaní predikčných schopností modelov a je odvodené AIC za predpokladu splnenia štandardných podmienok regularity (viď Príloha A.1), ktoré sú zavedené v teórii maximálnej vierohodnosti. Pripomeňme, že hľadanie maximálne vierohodných odhadov úzko súvisí s minimalizovaním Kullbackovej-Leiblerovej divergencie.

Definícia 5. *Predpokladajme, že náhodný vektor $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$ má hustotou g vzhľadom ku σ -konečnej miere μ . Označme ďalej f_θ súbor hustôt, ktoré predstavujú predpokladaný model o rozdelení vektora \mathbf{Y} , kde $\theta \in \Theta \subset \mathbb{R}^k$ je neznámy parameter. Kullbackova-Leiblerova divergencia (KL divergencia) je definovaná vzťahom*

$$\text{KL}(f_\theta, g) = \int \log \left(\frac{g(\mathbf{z})}{f_\theta(\mathbf{z})} \right) g(\mathbf{z}) d\mu(\mathbf{z}) = \mathbf{E}_z \left[\log(g(\mathbf{z})) - \log(f_\theta(\mathbf{z})) \right],$$

kde \mathbf{E}_z vyjadruje strednú hodnotu vzhľadom k novej realizácii \mathbf{z} nezávislej na pozorovanej realizácii \mathbf{y} s predpokladaným rovnakým rozdelením.

Poznámka. Prepojenie s teóriou maximálnej vierohodnosti je cez hustotu f_θ . Model, ktorý predpokladáme o vektore \mathbf{Y} , môžeme chápať ako vierohodnosť parametra θ reprezentovanú hustotou $f_\theta(\mathbf{y})$. Hodnota \mathbf{y} je pozorovaná realizácia vektora \mathbf{Y} .

Tvrdenie 8. *Platí*

$$\begin{aligned} \text{KL}(g, g) &= 0, \\ \text{KL}(f_\theta, g) &\geq 0, \quad \theta \in \Theta. \end{aligned}$$

Dôkaz. Rovnosť vyplýva priamo z definície dosadením skutočnej hustoty, pretože $\log(1) = 0$. Nerovnosť sa dá dokázať využitím Jensenovej nerovnosti nasledovne. Z definície máme

$$\text{KL}(f_\theta, g) = \mathbf{E}_z \log \left(\frac{g(\mathbf{z})}{f_\theta(\mathbf{z})} \right) = - \mathbf{E}_z \log \left(\frac{f_\theta(\mathbf{z})}{g(\mathbf{z})} \right). \quad (3.1)$$

Keďže logaritmus je konkávna funkcia a $f_\theta(\mathbf{z}) \leq g(\mathbf{z})$, platia nasledujúce nerovnosti

$$\mathbf{E}_z \log \left(\frac{f_\theta(\mathbf{z})}{g(\mathbf{z})} \right) \leq \log \left(\mathbf{E}_z \frac{f_\theta(\mathbf{z})}{g(\mathbf{z})} \right) \leq \log(1) = 0.$$

Táto nerovnosť je ekvivalentná s nasledujúcou

$$-\mathbf{E}_z \log \left(\frac{f_{\boldsymbol{\theta}}(\mathbf{z})}{g(\mathbf{z})} \right) \geq 0. \quad (3.2)$$

Zo vzťahov (3.1) a (3.2) dostaneme odhad zo znenia tvrdenia. \square

Na KL divergenciu sa teda môžeme pozeráť ako na mieru vzdialenosti predpokladaného modelu $f_{\boldsymbol{\theta}}$ od skutočného rozdelenia g . Keďže $\mathbf{E}_z \log(g(\mathbf{z}))$ je konštanta, minimalizovanie $\text{KL}(f_{\boldsymbol{\theta}}, g)$ odpovedá minimalizácii $\mathbf{E}_z (-\log(f_{\boldsymbol{\theta}}(\mathbf{z})))$, resp. maximalizácii $\mathbf{E}_z \log(f_{\boldsymbol{\theta}}(\mathbf{z}))$, vzhľadom k parametru $\boldsymbol{\theta}$. Maximálne vierohodný odhad zo vzťahu (A.1) na základe pozorovaní \mathbf{y} , o ktorých predpokladáme, že sú nezávislé na \mathbf{z} , by sme v tomto prípade hľadali ako

$$\hat{\boldsymbol{\theta}}(\mathbf{y}) = \arg \max_{\boldsymbol{\theta} \in \Theta} f_{\boldsymbol{\theta}}(\mathbf{y}) = \arg \max_{\boldsymbol{\theta} \in \Theta} \log(f_{\boldsymbol{\theta}}(\mathbf{y})).$$

Možným bodovým odhadom minima KL divergencie je teda hodnota nadobudnutá v maximálne vierohodnom odhade, a síce

$$\widehat{\text{KL}}(f_{\boldsymbol{\theta}}, g) = \text{KL}(f_{\hat{\boldsymbol{\theta}}(\mathbf{y})}, g) = \mathbf{E}_z \log \left(\frac{g(\mathbf{z})}{f_{\hat{\boldsymbol{\theta}}(\mathbf{y})}(\mathbf{z})} \right).$$

Museli by sme však poznať skutočné rozdelenie dané hustotou g . Avšak minimalizovanie odhadnutej priemernej hodnoty vzhľadom k rozdeleniu \mathbf{y} , teda $\min \mathbf{E}_y \widehat{\text{KL}}(f_{\boldsymbol{\theta}}, g)$, je opodstatnené pri hľadaní vhodného modelu. Táto minimalizácia je ekvivalentná nasledovnej

$$\min -2 \mathbf{E}_y \mathbf{E}_z \log(f_{\hat{\boldsymbol{\theta}}(\mathbf{y})}(\mathbf{z})), \quad (3.3)$$

pretože $\mathbf{E}_z \log(g(\mathbf{z}))$ je konštanta.

Definícia 6. Výraz z minimalizácie (3.3) nazývame Akaikova informácia (AI)

$$\text{AI} = -2 \mathbf{E}_y \mathbf{E}_z \log(f_{\hat{\boldsymbol{\theta}}(\mathbf{y})}(\mathbf{z})).$$

Poznámka. Násobenie dvojkou nemá na minimalizáciu žiaden vplyv, ale je vhodné pri ďalšom počítaní, ktoré vychádza zo znalosti rozdelení štatistík testov, ktoré sú založené na pomere vierohodností.

Hodnotu vektora \mathbf{z} nepozorujeme, možným odhadom AI môže byť preto hodnota vierohodnosti pre maximálne vierohodný odhad v pozorovanej realizácii \mathbf{y} , teda $\widehat{\text{AI}} = -2 \log(f_{\hat{\boldsymbol{\theta}}(\mathbf{y})}(\mathbf{y}))$. Tento odhad ale nie je nestranný, ako ukazuje Tvrdenie 9.

Tvrdenie 9. Označme $\boldsymbol{\theta}_0$ skutočnú hodnotu parametra, ktorá minimalizuje KL divergenciu. Potom vychýlenie odhadu $\widehat{\text{AI}}$ je

$$\begin{aligned} \text{bias}(\widehat{\text{AI}}) &= \mathbf{E}_y \mathbf{E}_z (\widehat{\text{AI}} - \text{AI}) = 2 \mathbf{E}_y (\log f_{\boldsymbol{\theta}_0}(\mathbf{y}) - \log f_{\hat{\boldsymbol{\theta}}(\mathbf{y})}(\mathbf{y})) + \\ &\quad + \mathbf{E}_y 2 \mathbf{E}_z (\log f_{\hat{\boldsymbol{\theta}}(\mathbf{y})}(\mathbf{z}) - \log f_{\boldsymbol{\theta}_0}(\mathbf{z})). \end{aligned}$$

Dôkaz. Dosadením definície a odhadu máme

$$\text{bias}(\widehat{\text{AI}}) = 2 \mathbb{E}_{\mathbf{y}} \mathbb{E}_{\mathbf{z}} \left(\log f_{\widehat{\boldsymbol{\theta}}(\mathbf{y})}(\mathbf{z}) - \log f_{\widehat{\boldsymbol{\theta}}(\mathbf{y})}(\mathbf{y}) \right).$$

Pričítaním a odčítaním členov $\log f_{\boldsymbol{\theta}_0}(\mathbf{z})$ a $\log f_{\boldsymbol{\theta}_0}(\mathbf{y})$ v predchádzajúcej rovnosti môžeme pravú stranu upraviť do tvaru

$$2 \mathbb{E}_{\mathbf{y}} \left(\log f_{\boldsymbol{\theta}_0}(\mathbf{y}) - \log f_{\widehat{\boldsymbol{\theta}}(\mathbf{y})}(\mathbf{y}) \right) + \mathbb{E}_{\mathbf{y}} 2 \mathbb{E}_{\mathbf{z}} \left(\log f_{\widehat{\boldsymbol{\theta}}(\mathbf{y})}(\mathbf{z}) - \log f_{\boldsymbol{\theta}_0}(\mathbf{z}) \right) + 2 \mathbb{E}_{\mathbf{y}} \mathbb{E}_{\mathbf{z}} \left(\log f_{\boldsymbol{\theta}_0}(\mathbf{z}) - \log f_{\boldsymbol{\theta}_0}(\mathbf{y}) \right). \quad (3.4)$$

Vyplyva to z toho, že pri počítaní $\mathbb{E}_{\mathbf{z}}$ môžeme chápať všetky hodnoty závislé na \mathbf{y} ako konštanty. Prvé dva členy výrazu (3.4) sú zhodné s tvrdením, preto k dokončeniu stačí ukázať

$$\mathbb{E}_{\mathbf{y}} \mathbb{E}_{\mathbf{z}} \left(\log f_{\boldsymbol{\theta}_0}(\mathbf{z}) - \log f_{\boldsymbol{\theta}_0}(\mathbf{y}) \right) = \mathbb{E}_{\mathbf{y}} \mathbb{E}_{\mathbf{z}} \log \left(\frac{f_{\boldsymbol{\theta}_0}(\mathbf{z})}{f_{\boldsymbol{\theta}_0}(\mathbf{y})} \right) = 0. \quad (3.5)$$

Z predpokladu $\boldsymbol{\theta}_0$ minimalizuje KL divergenciu, preto platí

$$\mathbb{E}_{\mathbf{z}} \log \left(\frac{g(\mathbf{z})}{f_{\boldsymbol{\theta}_0}(\mathbf{z})} \right) = \mathbb{E}_{\mathbf{y}} \log \left(\frac{g(\mathbf{y})}{f_{\boldsymbol{\theta}_0}(\mathbf{y})} \right).$$

Výrazy závislé na \mathbf{y} sú vzhľadom ku $\mathbb{E}_{\mathbf{z}}$ konštanty a naopak, preto po ich pridaní a úprave odčítaním strán rovností je táto rovnosť ekvivalentná

$$\mathbb{E}_{\mathbf{y}} \mathbb{E}_{\mathbf{z}} \log \left(\frac{f_{\boldsymbol{\theta}_0}(\mathbf{z})}{f_{\boldsymbol{\theta}_0}(\mathbf{y})} \right) = \mathbb{E}_{\mathbf{y}} \mathbb{E}_{\mathbf{z}} \log \left(\frac{g(\mathbf{z})}{g(\mathbf{y})} \right).$$

Pravá strana rovnosti je rovná

$$\mathbb{E}_{\mathbf{y}} \mathbb{E}_{\mathbf{z}} \log \left(\frac{g(\mathbf{z})}{g(\mathbf{y})} \right) = \mathbb{E}_{\mathbf{z}} \log g(\mathbf{z}) - \mathbb{E}_{\mathbf{y}} \log g(\mathbf{y}) = 0,$$

kde posledná rovnosť vychádza z nezávislosti realizácií a toho, že g je hustota skutočného rozdelenia oboch vektorov. Rovnosť (3.5) je týmto dokázaná, a preto aj celé tvrdenie. □

Ďalším krokom je odvodenie informačného kritéria. Predpokladajme, že sú splnené všetky podmienky regularity z teórie maximálnej vierohodnosti, ktoré uvádzame v prílohe A.1. Potom podľa Vety 19 platí konvergencia v distribúcii odvodená pri teste pomerom vierohodností, a síce

$$2 \left(\log f_{\widehat{\boldsymbol{\theta}}(\mathbf{y})}(\mathbf{y}) - \log f_{\boldsymbol{\theta}_0}(\mathbf{y}) \right) \xrightarrow[n \rightarrow \infty]{d} \chi_k^2, \quad (3.6)$$

$$2 \mathbb{E}_{\mathbf{z}} \left(\log f_{\boldsymbol{\theta}_0}(\mathbf{z}) - \log f_{\widehat{\boldsymbol{\theta}}(\mathbf{y})}(\mathbf{z}) \right) \xrightarrow[n \rightarrow \infty]{d} \chi_k^2. \quad (3.7)$$

Konvergencia (3.6) je uvedená v bode (iii) spomínanej vety a (3.7) je uvedená v článku Greven a Kneib (2010).

Nestranný odhad AI dostaneme odčítaním vychýlenia od známeho odhadu, $\widehat{\text{AI}} - \text{bias}(\widehat{\text{AI}})$. Stredné hodnoty $\mathbb{E}_{\mathbf{y}}$, ktoré sa vyskytujú vo vyjadrení vychýlenia, nahradíme strednými hodnotami ich asymptotických rozdelení za predpokladu splnenia podmienok regularity. Obe veličiny sú aproximované χ_k^2 rozdelením, čo sa tvrdí vo vzťahoch (3.6), (3.7) a porovnaním vychýlenia po otočení znamienok. Preto súčet oboch stredných hodnôt je $2k$. Takto bolo odvodené Akaikovo informačné kritérium.

Definícia 7. Pre parameter $\boldsymbol{\theta} \in \Theta$, kde $\Theta \subset \mathbb{R}^k$ a pozorovanú realizáciu \mathbf{y} náhodného vektora \mathbf{Y} definujeme Akaikovo informačné kritérium (AIC) rovnosťou

$$\text{AIC} = -2 \log f_{\hat{\boldsymbol{\theta}}(\mathbf{y})}(\mathbf{y}) + 2k,$$

kde $\hat{\boldsymbol{\theta}}(\mathbf{y})$ je maximálne vierohodný odhad parametra $\boldsymbol{\theta}$ na základe vierohodnosti $f_{\boldsymbol{\theta}}$.

Z predošlých výpočtov a úvah vieme, že v prípade splnených podmienok regularity je AIC asymptoticky nestranným odhadom AI. Preto minimalizovanie AIC pre viacero možných modelov môžeme chápať ako minimalizovanie priemernej vzdialenosti modelu od skutočného rozdelenia.

3.1 Vierohodnosť v LME

Pre odvodenie AIC v lineárnom zmiešanom modeli je vhodné pripomenúť, ako vyzerá vierohodnosť modelu. Pripomeňme, že parameter $\boldsymbol{\psi}$ označuje variančné parametre, ktoré popisujú maticu \mathbb{D} . Ďalšími parametrami LME modelu sú pevné efekty $\boldsymbol{\beta}$ a reziduálny rozptyl σ_e^2 . Vierohodnosť lineárneho zmiešaného modelu je hustota vektora \mathbf{Y} pri daných parametroch. Navyiac vieme, že vektory $\mathbf{Y}_1, \dots, \mathbf{Y}_N$ sú nezávislé. Preto

$$L(\boldsymbol{\beta}, \boldsymbol{\psi}, \sigma_e^2) = \prod_{i=1}^N f(\mathbf{Y}_i | \boldsymbol{\beta}, \boldsymbol{\psi}, \sigma_e^2).$$

Využijeme vzťah pre počítanie marginálnej hustoty integráciou a dostaneme

$$L(\boldsymbol{\beta}, \boldsymbol{\psi}, \sigma_e^2) = \prod_{i=1}^N \int f(\mathbf{Y}_i | \mathbf{b}_i, \boldsymbol{\beta}, \boldsymbol{\psi}, \sigma_e^2) f(\mathbf{b}_i) d\mathbf{b}_i. \quad (3.8)$$

Vierohodnosť má tvar typický pre hierarchické modely. Ide o integrál z hustoty združeného rozdelenia pozorovaných a latentných dát. V každom prípade vieme hustotu združeného rozdelenia vždy písať ako súčin hustoty dát podmienenej skrytými náhodnými efektami a hustoty náhodných efektov. Z definície LME modelu vieme, že uvažujeme hustoty normálnych rozdelení, teda pôjde o hustotu normálneho rozdelenia z marginálneho vyjadrenia (1.1).

$$L^M(\boldsymbol{\beta}, \boldsymbol{\psi}, \sigma_e^2) = (2\pi)^{-n/2} |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{Y} - \mathbb{X}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \mathbb{X}\boldsymbol{\beta}) \right\}, \quad (3.9)$$

kde pripomíname tvar rozptylovej matice $\boldsymbol{\Sigma} = \sigma_e^2 \mathbb{I}_n + \mathbb{Z}\mathbb{D}_*\mathbb{Z}^\top$. Hustotu pozorovaných dát pri znalosti skrytých budeme niekedy označovať termínom podmienená vierohodnosť.

$$\begin{aligned} L^P(\boldsymbol{\beta}, \boldsymbol{\psi}, \sigma_e^2 | \mathbf{b}) &= \prod_{i=1}^N f(\mathbf{Y}_i | \mathbf{b}_i, \boldsymbol{\beta}, \boldsymbol{\psi}, \sigma_e^2) = \\ &= (2\pi\sigma_e^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma_e^2} (\mathbf{Y} - \mathbb{X}\boldsymbol{\beta} - \mathbb{Z}\mathbf{b})^\top (\mathbf{Y} - \mathbb{X}\boldsymbol{\beta} - \mathbb{Z}\mathbf{b}) \right\}. \end{aligned} \quad (3.10)$$

Odhadovanie parametrov pomocou maximálnej vierohodnosti prebieha iteračne. Postupne odhadujeme jednotlivé parametre a prechádzame k profilovej

vierohodnosti na odhadnutie ďalších. Numericky potom opakujeme jednotlivé kroky pre aktuálne hodnoty odhadov. Teraz uvidíme jedno možné vyjadrenie profilových vierohodností pre jednotlivé parametre. Vyjadrenie neskôr využijeme v ďalšej časti práce pri teoretickom odvodení korekčného členu v podmienenom Akaikovom kritériu. Označme maticu

$$\mathbb{V} = \frac{1}{\sigma_e^2} \mathbb{\Sigma} = \mathbb{I}_n + \mathbb{Z} \mathbb{R}_* \mathbb{Z}^\top, \quad \mathbb{R}_* = \frac{1}{\sigma_e^2} \mathbb{D}_*. \quad (3.11)$$

Budeme predpokladať, že matica \mathbb{R}_* je parametrizovaná parametrami $\boldsymbol{\psi}^*$, ktoré sú odvodené z $\boldsymbol{\psi}$ po predelení σ_e^2 . Maximalizácia marginálnej vierohodnosti pri známych variančných parametroch (3.9) odpovedá minimalizácii kvadratickej formy, ktorá sa nachádza v exponenciále tohto vzťahu. Ide o zhodný výsledok ako riešenie Hendersonových rovníc

$$\hat{\boldsymbol{\beta}} \equiv \hat{\boldsymbol{\beta}}(\boldsymbol{\psi}^*, \sigma_e^2) = (\mathbb{X}^\top \mathbb{\Sigma}^{-1} \mathbb{X})^{-1} \mathbb{X}^\top \mathbb{\Sigma}^{-1} \mathbf{Y} = (\mathbb{X}^\top \mathbb{V}^{-1} \mathbb{X})^{-1} \mathbb{X}^\top \mathbb{V}^{-1} \mathbf{Y}, \quad (3.12)$$

kde posledná rovnosť sa overí jednoduchým dosadením prvého vzťahu z (3.11) do vyjadrenia odhadu. Dosadením tohto vyjadrenia a nahradením matice $\mathbb{\Sigma}$ výrazom $\sigma_e^2 \mathbb{V}$ vo vzťahu (3.9) odvodíme profilovú vierohodnosť pre variančné parametre

$$L^M(\boldsymbol{\psi}^*, \sigma_e^2 | \hat{\boldsymbol{\beta}}) = (2\pi\sigma_e^2)^{-n/2} |\mathbb{V}|^{-1/2} \exp \left\{ -\frac{1}{2\sigma_e^2} (\mathbf{Y} - \mathbb{X}\hat{\boldsymbol{\beta}})^\top \mathbb{V}^{-1} (\mathbf{Y} - \mathbb{X}\hat{\boldsymbol{\beta}}) \right\}.$$

Následne pri znalosti $\boldsymbol{\psi}^*$ vieme klasickým postupom uvažovania logaritmu tejto profilovej vierohodnosti a deriváciou odvodiť, že maximum je nadobudnuté pre

$$\hat{\sigma}_e^2 \equiv \hat{\sigma}_e^2(\boldsymbol{\psi}^*) = \frac{1}{n} (\mathbf{Y} - \mathbb{X}\hat{\boldsymbol{\beta}})^\top \mathbb{V}^{-1} (\mathbf{Y} - \mathbb{X}\hat{\boldsymbol{\beta}}).$$

Nakoniec dosadením tohto odhadu získame profilovú vierohodnosť pre variančné parametre parametrizujúce \mathbb{R}_* v tvare

$$L^M(\boldsymbol{\psi}^* | \hat{\boldsymbol{\beta}}, \hat{\sigma}_e^2) = (2\pi)^{-n/2} |\mathbb{V}|^{-1/2} \left[\frac{1}{n} (\mathbf{Y} - \mathbb{X}\hat{\boldsymbol{\beta}})^\top \mathbb{V}^{-1} (\mathbf{Y} - \mathbb{X}\hat{\boldsymbol{\beta}}) \right]^{-n/2} e^{-n/2}. \quad (3.13)$$

Maximalizácia tohto výrazu následne vedie na odhadnutie $\hat{\boldsymbol{\psi}}^*$. Avšak v tomto prípade už neexistuje explicitné vyjadrenie. Ako sme spomínali, tvar profilovej vierohodnosti (3.13) použijeme neskôr v práci.

Pre úplnosť ukážeme, ako sa častejšie postupuje pri hľadaní maximálne vierohodných odhadov. Vyššie popísaný postup nedáva návod, ako nájsť predikcie náhodných efektov. Na to slúžia Hendersonove rovnice a úprava vierohodnosti (3.8) po dosadení predikcií náhodných efektov. Túto úpravu stručne popíšeme. Pre výpočet integrálu označme rozklad matice $\sigma_e^2 \mathbb{D}^{-1} = \mathbb{L}^\top \mathbb{L}$ (tento rozklad matice existuje, pretože \mathbb{D} je pozitívne definitná). V tomto značení sa dá vierohodnosť písať v tvare

$$L^M(\boldsymbol{\beta}, \boldsymbol{\psi}, \sigma_e^2) = \prod_{i=1}^N \frac{|\mathbb{L}|}{(2\pi\sigma_e^2)^{n_i/2}} \int \frac{1}{(2\pi\sigma_e^2)^{q/2}} \exp \left\{ -\frac{1}{2\sigma_e^2} \left\| \tilde{\mathbf{Y}}_i - \tilde{\mathbb{X}}_i \boldsymbol{\beta} - \tilde{\mathbb{Z}}_i \mathbf{b}_i \right\|^2 \right\} d\mathbf{b}_i, \quad (3.14)$$

kde

$$\tilde{\mathbf{Y}}_i = \begin{pmatrix} \mathbf{Y}_i \\ \mathbf{0} \end{pmatrix}, \quad \tilde{\mathbb{X}}_i = \begin{pmatrix} \mathbb{X}_i \\ \mathbf{0} \end{pmatrix}, \quad \tilde{\mathbb{Z}}_i = \begin{pmatrix} \mathbb{Z}_i \\ \mathbb{L} \end{pmatrix}$$

a $\|\cdot\|$ označuje Euklidovskú normu ($\|\mathbf{a}\|^2 = \mathbf{a}^\top \mathbf{a}$). Pri fixnom parametri $\boldsymbol{\beta}$ je Euklidovská norma vo vyjadrení (3.14) minimalizovaná pre

$$\hat{\mathbf{b}}_i = (\tilde{\mathbf{Z}}_i^\top \tilde{\mathbf{Z}}_i)^{-1} \tilde{\mathbf{Z}}_i^\top (\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_i \boldsymbol{\beta}).$$

Ďalej zo znalosti normality náhodných efektov máme

$$\int \frac{1}{(2\pi\sigma_e^2)^{q/2}} \exp \left\{ -\frac{1}{2\sigma_e^2} (\mathbf{b}_i - \hat{\mathbf{b}}_i)^\top \tilde{\mathbf{Z}}_i^\top \tilde{\mathbf{Z}}_i (\mathbf{b}_i - \hat{\mathbf{b}}_i) \right\} d\mathbf{b}_i = \frac{1}{\sqrt{|\tilde{\mathbf{Z}}_i^\top \tilde{\mathbf{Z}}_i|}}.$$

Rozklad Euklidovskej normy pomocou Pytagorovej vety na

$$\|\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_i \boldsymbol{\beta} - \tilde{\mathbf{Z}}_i \mathbf{b}_i\|^2 = \|\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_i \boldsymbol{\beta} - \tilde{\mathbf{Z}}_i \hat{\mathbf{b}}_i\|^2 + (\mathbf{b}_i - \hat{\mathbf{b}}_i)^\top \tilde{\mathbf{Z}}_i^\top \tilde{\mathbf{Z}}_i (\mathbf{b}_i - \hat{\mathbf{b}}_i)$$

nám umožní použitím predošlých vzťahov upraviť vierohodnosť do tvaru

$$L^M(\boldsymbol{\beta}, \boldsymbol{\psi}, \sigma_e^2) = \frac{1}{(2\pi\sigma_e^2)^{\frac{n}{2}}} \exp \left\{ \frac{-1}{2\sigma_e^2} \sum_{i=1}^N \|\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_i \boldsymbol{\beta} - \tilde{\mathbf{Z}}_i \hat{\mathbf{b}}_i\|^2 \right\} \prod_{i=1}^N \frac{|\mathbb{L}|}{\sqrt{|\mathbf{Z}_i^\top \mathbf{Z}_i + \mathbb{L}^\top \mathbb{L}|}}. \quad (3.15)$$

S týmto vyjadrením vierohodnosti sa následne rovnakým postupom nájdú odhady $\boldsymbol{\beta}$, σ_e^2 , ktoré majú podobné explicitné vyjadrenia. Nakoniec aj odhad parametra $\boldsymbol{\psi}$, ktorý parametrizuje maticu \mathbb{L} . Algoritmus potom prebieha iteračne.

3.2 AIC pre lineárny zmiešaný model

V kontexte LME modelu máme podobne ako pri hľadaní predikcie dve základne možnosti, ako uplatniť myšlienku informačného kritéria. Budeme ich tiež rozlišovať ako marginálna a podmienená verzia, čo bude súvisieť s hierarchickou úrovňou modelu. Každá z nich má svoje výhody a nevýhody. Vlastnosti oboch postupne uvedieme v nadväzujúcich častiach práce.

3.2.1 Marginálne AIC

Marginálny prístup vychádza z vierohodnosti marginálneho rozdelenia vektora a vierohodnosti (3.9). Počet parametrov tohto modelu je $p + Q + 1$, kde p je počet pevných parametrov $\boldsymbol{\beta}$, $1 \leq Q \leq q(q+1)/2$ predstavuje počet variančných parametrov $\boldsymbol{\psi}$ a posledným parametrom je σ_e^2 . Marginálnu verziu Akaikovo informačného kritéria potom definujeme výrazom

$$\text{mAIC} = -2 \log L^M(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\psi}}, \hat{\sigma}_e^2) + 2(p + Q + 1), \quad (3.16)$$

kde $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\psi}}$ a $\hat{\sigma}_e^2$ sú maximálne vierohodné odhady parametrov.

Lineárny zmiešaný model nespĺňa všetky podmienky regularity teórie maximálnej vierohodnosti, ak pracujeme s marginálnou AI. V marginálnom modeli (1.1) nie sú zložky Y_{ij} vektora \mathbf{Y} na sebe nezávislé. Taktiež nastávajú problémy s variančnými parametrami, pretože parametrický priestor Θ nie je otvorená množina a môže sa stať, že sa snažíme identifikovať parameter na hranici množiny Θ .

Z toho dôvodu marginálne Akaikovo kritérium nemusí byť nestranným odhadom Akikovej informácie. Problém môže nastať pri porovnávaní modelov, ktoré

sa líšia iba variančnou štruktúrou náhodných efektov. To úzko súvisí s testami na podmodely, kde je menší model určený odstránením náhodného efektu z väčšieho modelu, čo je ekvivalentné nulovému rozptylu náhodného efektu vo väčšom modeli. Menší model je tak identifikovaný parametrom ležiacim na hranici parametrického priestoru a testy pomerom vierohodnosti majú problémy. V článku Greven a Kneib (2010) je príklad pre jeden neznámy rozptyl náhodného efektu, podľa ktorého mAIC nie je nestranným odhadom.

Neskôr v práci ukážeme, za akých predpokladov je mAIC asymptoticky nestranným odhadom AI, čo preberieme z článku Kubokawa (2011). Bližším podrobnostiam sa venujeme v časti 3.2.3 tejto práce.

3.2.2 Podmienené AIC

Podmienený prístup vychádza z rozdelenia vektorov \mathbf{Y}_i pri znalosti náhodných efektov. Podľa článku Vaida a Blanchard (2005) je tento prístup vhodnejší, ak predpokladáme, že náhodné vektory \mathbf{Y} a \mathbf{Z} splňujú LME model s rovnakými hodnotami náhodných efektov. To znamená, že našim hlavným cieľom je vybrať najvhodnejší model s lepšimi vlastnosťami predikcie náhodných efektov, resp. predikcie na úrovni tzv. zhlukov. Na hodnoty náhodných efektov \mathbf{b}_i sa môžeme v tomto prístupe pozeráť ako na parametre.

Podmienená Akaikova informácia (cAI) splňuje rovnakú úlohu ako AI v marginálnom prístupe. Predpokladáme, že náhodný vektor \mathbf{Y} má hustotu $g(\cdot|\mathbf{b})$, kde \mathbf{b} je vektor skutočných hodnôt náhodných efektov s prípadne všeobecnou hustotou h . Ďalej označíme \mathbf{z} novú realizáciu tohto náhodného vektora, ktorá je podmienene na hodnotách \mathbf{b} nezávislá s realizáciou \mathbf{y} . Navyše nech existuje skutočná hodnota $\boldsymbol{\theta}_0$ parametra $\boldsymbol{\theta} \in \Theta$ taká, že hustota $g(\cdot|\mathbf{b}) = f_{\boldsymbol{\theta}_0}(\cdot|\mathbf{b})$, ktorá môže byť reprezentovaná napríklad podmienenou vierohodnosťou.

Definícia 8. Podmienenú Akaikovu informáciu *definujeme vzťahom*

$$\text{cAI} = -2 \mathbf{E}_{\mathbf{y}, \mathbf{b}} \mathbf{E}_{\mathbf{z}|\mathbf{b}} \log f_{\hat{\boldsymbol{\theta}}(\mathbf{y})}(\mathbf{z}|\hat{\mathbf{b}}(\mathbf{y})),$$

kde $\hat{\boldsymbol{\theta}}(\mathbf{y})$ a $\hat{\mathbf{b}}(\mathbf{y})$ sú ľubovoľné odhady, resp. predikcie na základe realizácie \mathbf{y} .

Poznámka. Oproti AI sa definícia líši použitím podmienenej vierohodnosti a taktiež tým, že prvá stredná hodnota je vzhľadom ku združenému rozdeleniu \mathbf{y} a \mathbf{b} .

Pri práci s LME pracujeme s parametrami $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\psi}^\top, \sigma_e^2)^\top$. Za hustotu $f_{\boldsymbol{\theta}}(\cdot|\mathbf{b})$ berieme podmienenú vierohodnosť LME modelu, výraz (3.10). Za odhad parametra $\hat{\boldsymbol{\theta}}(\mathbf{y})$ dosadzujeme maximálne vierohodný, ktorý sa získa maximalizáciou vierohodnosti (3.15) a predikcie náhodných efektov $\hat{\mathbf{b}}(\mathbf{y})$ sú z Hendersonových rovníc (1.4). Taktiež pre novú nezávislú realizáciu vektora sme zaviedli pojem replikovaných dát \mathbf{Y}^* v Definícii 2. Pre LME model je teda podmienená Akaikova informácia rovná

$$\text{cAI} = -2 \mathbf{E}_{\mathbf{Y}, \mathbf{b}} \mathbf{E}_{\mathbf{Y}^*|\mathbf{b}} \log \left[\left(2\pi\hat{\sigma}_e^2 \right)^{-n/2} \exp \left\{ -\frac{1}{2\hat{\sigma}_e^2} \left\| \mathbf{Y}^* - \mathbb{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\mathbf{b}} \right\|^2 \right\} \right]. \quad (3.17)$$

Poznámka. Maximálne vierohodný odhad $\hat{\boldsymbol{\beta}}$ je totožný s riešením Hendersonových rovníc, ak pokladáme variančné parametre za známe. V opačnom prípade

ho získame dosadením odhadov $\hat{\boldsymbol{\psi}}$ a $\hat{\sigma}_e^2$ do Hendersonových rovníc, a tak neexistuje uzavreté vyjadrenie odhadu $\hat{\boldsymbol{\beta}}$. Riešenie sa hľadá iteračným numerickým algoritmom.

Podmienené Akaikovo kritérium sa odvodí hľadaním nestranného odhadu cAI. Zaoberali sa tým v článkoch Vaida a Blanchard (2005) a Liang a kol. (2008). Všetky prístupy sa dajú zhrnúť na nasledovný tvar podmieneného AIC

$$\text{cAIC} = -2 \log L^P(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\psi}}, \hat{\sigma}_e^2 | \hat{\mathbf{b}}) + 2\rho, \quad (3.18)$$

kde člen ρ sa líši podľa toho, ktoré z variančných parametrov $\boldsymbol{\psi}$ a σ_e^2 považujeme za známe, resp. neznáme. Na ρ sa taktiež môžeme pozeráť ako na korekčný člen, ktorý upravuje hodnotu logaritmu vierohodnosti v maximálne vierohodnom odhade tak, aby bolo cAIC nestranným odhadom cAI. Plní teda podobnú úlohu ako rozmer parametrického priestoru, keď sme pracovali s marginálnym AI vo všeobecnom prípade.

Tvrdenie 10. *Za predpokladu, že poznáme hodnotu σ_e^2 , platí*

$$\mathbf{E}_{\mathbf{Y}, \mathbf{b}} \left(-2 \log L^P(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\psi}}, \sigma_e^2 | \hat{\mathbf{b}}) - \text{cAI} \right) = -2 \mathbf{E}_{\mathbf{b}} [\rho], \quad (3.19)$$

kde

$$\rho = \mathbf{E}_{\mathbf{Y} | \mathbf{b}} \left[\frac{1}{\sigma_e^2} (\mathbf{Y} - \boldsymbol{\mu})^\top \widehat{\mathbf{Y}}^P \right]. \quad (3.20)$$

Podmienené Akaikovo kritérium je v tom prípade nestranný odhad podmienenej Akaikovej informácie.

Dôkaz. Pripomeňme definíciu podmienenej predikcie $\widehat{\mathbf{Y}}^P = \mathbb{X}\hat{\boldsymbol{\beta}} + \mathbb{Z}\hat{\mathbf{b}}$ a dosadme vierohodnosť (3.10) a výraz (3.17) do ľavej strany rovnosti (3.19) zo znenia. Vidíme, že sa potom rovná

$$\begin{aligned} -2 \mathbf{E}_{\mathbf{Y}, \mathbf{b}} \left(-\frac{n}{2} \log(2\pi\sigma_e^2) - \frac{\|\mathbf{Y} - \widehat{\mathbf{Y}}^P\|^2}{2\sigma_e^2} \right) + \\ + 2 \mathbf{E}_{\mathbf{Y}, \mathbf{b}} \mathbf{E}_{\mathbf{Y}^* | \mathbf{b}} \left(-\frac{n}{2} \log(2\pi\sigma_e^2) - \frac{\|\mathbf{Y}^* - \widehat{\mathbf{Y}}^P\|^2}{2\sigma_e^2} \right). \end{aligned}$$

Prvý a tretí člen sú konštanty, ktoré sa od seba odčítajú. Navyiac stredná hodnota sa dá počítať postupným podmieňovaním (schematicky $\mathbf{E}_{\mathbf{Y}, \mathbf{b}} = \mathbf{E}_{\mathbf{b}} \mathbf{E}_{\mathbf{Y} | \mathbf{b}}$). Preto je ľavá strana rovnosti (3.19) rovná

$$-2 \mathbf{E}_{\mathbf{b}} \left[\frac{1}{2\sigma_e^2} \left(\mathbf{E}_{\mathbf{Y} | \mathbf{b}} \mathbf{E}_{\mathbf{Y}^* | \mathbf{b}} \|\mathbf{Y}^* - \widehat{\mathbf{Y}}^P\|^2 - \mathbf{E}_{\mathbf{Y} | \mathbf{b}} \|\mathbf{Y} - \widehat{\mathbf{Y}}^P\|^2 \right) \right]. \quad (3.21)$$

Označme $\boldsymbol{\mu} = \mathbb{X}\boldsymbol{\beta} + \mathbb{Z}\mathbf{b}$ podmienenú strednú hodnotu vektorov \mathbf{Y} , \mathbf{Y}^* a $\widehat{\mathbf{Y}}^P$, podmienenú hodnotami \mathbf{b} . Normu vektorov z predošlej rovnosti upravíme pomocou nasledujúceho vzťahu

$$\|\mathbf{Y}^\perp - \widehat{\mathbf{Y}}^P\|^2 = \|\mathbf{Y}^\perp - \boldsymbol{\mu}\|^2 + \|\widehat{\mathbf{Y}}^P - \boldsymbol{\mu}\|^2 - 2(\mathbf{Y}^\perp - \boldsymbol{\mu})^\top (\widehat{\mathbf{Y}}^P - \boldsymbol{\mu}), \quad (3.22)$$

kde za \mathbf{Y}^\perp môžeme dosadiť \mathbf{Y} alebo \mathbf{Y}^* . Z definície replikovaných dát a ich nezávislosti s predikciou $\widehat{\mathbf{Y}}^P$ odvodíme nasledujúce rovnosti

$$\begin{aligned} \mathbb{E}_{\mathbf{Y}^*|b} \|\mathbf{Y}^* - \boldsymbol{\mu}\|^2 &= \mathbb{E}_{\mathbf{Y}|b} \|\mathbf{Y} - \boldsymbol{\mu}\|^2, \\ \mathbb{E}_{\mathbf{Y}|b} \mathbb{E}_{\mathbf{Y}^*|b} \|\widehat{\mathbf{Y}}^P - \boldsymbol{\mu}\|^2 &= \mathbb{E}_{\mathbf{Y}|b} \|\widehat{\mathbf{Y}}^P - \boldsymbol{\mu}\|^2. \end{aligned}$$

S ich využitím a dosadením vzťahu (3.22) zjednodušíme výraz (3.21) na nasledujúci

$$-2 \mathbb{E}_b \left\{ \frac{1}{\sigma_e^2} \mathbb{E}_{\mathbf{Y}|b} \left[(\mathbf{Y} - \boldsymbol{\mu})^\top (\widehat{\mathbf{Y}}^P - \boldsymbol{\mu}) - \mathbb{E}_{\mathbf{Y}^*|b} (\mathbf{Y}^* - \boldsymbol{\mu})^\top (\widehat{\mathbf{Y}}^P - \boldsymbol{\mu}) \right] \right\}. \quad (3.23)$$

Ďalej kvôli nezávislosti replikovaných dát a predikcie $\widehat{\mathbf{Y}}^P$ platí

$$\mathbb{E}_{\mathbf{Y}^*|b} (\mathbf{Y}^* - \boldsymbol{\mu})^\top (\widehat{\mathbf{Y}}^P - \boldsymbol{\mu}) = \left[\mathbb{E}_{\mathbf{Y}^*|b} (\mathbf{Y}^* - \boldsymbol{\mu}) \right]^\top (\widehat{\mathbf{Y}}^P - \boldsymbol{\mu}) = 0.$$

Taktiež platí

$$\mathbb{E}_{\mathbf{Y}|b} (\mathbf{Y} - \boldsymbol{\mu})^\top \boldsymbol{\mu} = \left[\mathbb{E}_{\mathbf{Y}|b} (\mathbf{Y} - \boldsymbol{\mu}) \right]^\top \boldsymbol{\mu} = 0. \quad (3.24)$$

Preto zjednodušíme výraz (3.23) do tvaru

$$-2 \mathbb{E}_b \frac{1}{\sigma_e^2} \mathbb{E}_{\mathbf{Y}|b} (\mathbf{Y} - \boldsymbol{\mu})^\top \widehat{\mathbf{Y}}^P. \quad (3.25)$$

Výraz (3.25) je rovný $-2 \mathbb{E}_b (\rho)$. Celkovo dostávame nestrannosť

$$\mathbb{E}_{\mathbf{Y},b} \left(-2 \log L^P(\widehat{\boldsymbol{\beta}}, \boldsymbol{\psi}, \sigma_e^2 | \widehat{\mathbf{b}}) + 2\rho \right) = c\text{AI}.$$

□

Pozrime sa, ako vyzerá korekčný člen v prípade známej rozptylovej matice náhodných efektov.

Veta 11. *Za predpokladu, že matica \mathbb{D}_* a parameter σ_e^2 sú známe, je nestranný odhad $c\text{AI}$ v tvare*

$$-2 \log L^P(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\psi}}, \widehat{\sigma}_e^2 | \widehat{\mathbf{b}}) + 2 \text{tr}(\mathbb{H}),$$

kde \mathbb{H} je projekčná matica vektora \mathbf{Y} na podmienenú projekciu, t.j. matica splňujúca $\widehat{\mathbf{Y}}^P = \mathbb{H}\mathbf{Y}$.

Poznámka. V prípade známych variančných parametrov je teda ρ z rovnosti (3.18) rovné $\text{tr}(\mathbb{H})$.

Poznámka. Tvar projekčnej matice \mathbb{H} vychádza z Hendersonových odhadovacích rovníc. Pre podmienenú projekciu $\widehat{\mathbf{Y}}^P = \mathbb{X}\widehat{\boldsymbol{\beta}} + \mathbb{Z}\widehat{\mathbf{b}}$ máme túto projekčnú maticu

$$\mathbb{H} = \begin{pmatrix} \mathbb{X} & \mathbb{Z} \end{pmatrix} \begin{pmatrix} \mathbb{X}^\top \mathbb{X} & \mathbb{X}^\top \mathbb{Z} \\ \mathbb{Z}^\top \mathbb{X} & \mathbb{Z}^\top \mathbb{Z} + \sigma_e^2 \mathbb{D}_*^{-1} \end{pmatrix}^{-1} \begin{pmatrix} \mathbb{X}^\top \\ \mathbb{Z}^\top \end{pmatrix}. \quad (3.26)$$

Z vlastností operátora stopy platí

$$\text{tr}(\mathbb{H}) = \text{tr} \left\{ \begin{pmatrix} \mathbb{X}^\top \mathbb{X} & \mathbb{X}^\top \mathbb{Z} \\ \mathbb{Z}^\top \mathbb{X} & \mathbb{Z}^\top \mathbb{Z} + \sigma_e^2 \mathbb{D}_*^{-1} \end{pmatrix}^{-1} \begin{pmatrix} \mathbb{X}^\top \mathbb{X} & \mathbb{X}^\top \mathbb{Z} \\ \mathbb{Z}^\top \mathbb{X} & \mathbb{Z}^\top \mathbb{Z} \end{pmatrix} \right\}.$$

Dôkaz. Dôkaz vychádza z článku Vaida a Blanchard (2005). Pri známych variančných parametroch potrebujeme len maximálne vierohodný odhad $\hat{\beta}$ a predikciu náhodných efektov $\hat{\mathbf{b}}$. Ostatné parametre poznáme. Použijeme rovnosť (3.20) z predošlého tvrdenia, ktoré sme oproti dôkazu v spomínanom článku podrobnejšie ukázali a vyčlenili, pretože sa použije aj v ďalšej časti.

$$\rho = \mathbb{E}_{\mathbf{Y}|\mathbf{b}} \frac{1}{\sigma_e^2} (\mathbf{Y} - \boldsymbol{\mu})^\top \widehat{\mathbf{Y}}^P,$$

kam dosadíme vyjadrenie $\widehat{\mathbf{Y}}^P = \mathbb{H}\mathbf{Y}$ a upravíme (3.25) na

$$\rho = \frac{1}{\sigma_e^2} \mathbb{E}_{\mathbf{Y}|\mathbf{b}} (\mathbf{Y} - \boldsymbol{\mu})^\top \mathbb{H}(\mathbf{Y} - \boldsymbol{\mu}), \quad (3.27)$$

kde $\boldsymbol{\mu}$ môžeme pridať do zátvorky, pretože platí

$$\mathbb{E}_{\mathbf{Y}|\mathbf{b}} (\mathbf{Y} - \boldsymbol{\mu})^\top (\mathbb{H}\boldsymbol{\mu}) = \left[\mathbb{E}_{\mathbf{Y}|\mathbf{b}} (\mathbf{Y} - \boldsymbol{\mu}) \right]^\top (\mathbb{H}\boldsymbol{\mu}) = 0.$$

Vo výraze (3.27) máme kvadratickú formu, čo je skalár, preto môžeme aplikovať operátor stopy, a vnútri následne upraviť poradie násobenia matic. Dostaneme

$$\begin{aligned} \rho &= \frac{1}{\sigma_e^2} \mathbb{E}_{\mathbf{Y}|\mathbf{b}} \operatorname{tr} \left((\mathbf{Y} - \boldsymbol{\mu})^\top \mathbb{H}(\mathbf{Y} - \boldsymbol{\mu}) \right) = \\ &= \frac{1}{\sigma_e^2} \operatorname{tr} \left(\mathbb{H} \mathbb{E}_{\mathbf{Y}|\mathbf{b}} (\mathbf{Y} - \boldsymbol{\mu}) (\mathbf{Y} - \boldsymbol{\mu})^\top \right) = \operatorname{tr}(\mathbb{H}), \end{aligned} \quad (3.28)$$

kde posledná rovnosť vyplýva z toho, že rozptylová matica \mathbf{Y} pri podmienení hodnotami \mathbf{b} je $\sigma_e^2 \mathbb{I}_n$. Posledná rovnosť je znením vety, ako sme uvideli v poznámke pod vetou. □

Poznámka. V článku Vaida a Blanchard (2005) je navrhnuté odhadnúť korekčný člen dosadením maximálne vierohodných odhadov, teda výrazom $\hat{\rho} = \mathbb{H}(\hat{\boldsymbol{\psi}}, \hat{\sigma}_e^2)$. Avšak je potrebné uvažovať aj korekciu za odhadovanie variančných parametrov, s ktorou pracuje asymptoticky neustranný odhad cAI, o ktorom píšeme na konci časti 3.2.3 tejto práce.

Ďalej chceme odvodiť, ako má vyzeráť korekčný člen, ak maticu \mathbb{D} odhadujeme, teda používame maximálne vierohodný odhad parametra $\boldsymbol{\psi}$. K tomu najprv sformulujeme pomocnú lemmu, ktorá je z článku Stein (1981).

Lemma 12. *Nech Y je náhodná veličina s normálnym rozdelením $\mathcal{N}(\mu, \sigma^2)$ a g je reálna diferencovateľná funkcia taká, že $\mathbb{E}|g'(Y)| < \infty$. Označme*

$$h(y) = g\left(\frac{y - \mu}{\sigma}\right).$$

Potom platí

$$\mathbb{E} h'(Y) = \mathbb{E} \left[\frac{Y - \mu}{\sigma^2} h(Y) \right].$$

Dôkaz. Viď článok Stein (1981), lemma 1 alebo príloha A.2 tejto práce. □

Poznámka. Predpoklad, že h definovaná v znení lemy je diferencovateľná so strednou hodnotou $\mathbb{E} |h'(Y)| < \infty$, je ekvivalentný rovnakému predpokladu o funkcii g .

V ďalšej vete a tvrdení budeme pracovať s deriváciami jedného vektora podľa zložiek iného. Pripomenieme teda, čo myslíme ďalej používaným značením. Pre vektory $\mathbf{X} \in \mathbb{R}^n$ a $\mathbf{Y} \in \mathbb{R}^m$ rozumieme

$$\frac{\partial \mathbf{X}}{\partial \mathbf{Y}^\top} = \mathbb{M},$$

kde \mathbb{M} je matica typu $n \times m$ so zložkami $\partial X_i / \partial Y_j$ pre $i = 1, \dots, n$ a $j = 1, \dots, m$. Budeme pracovať s prípadom $n = m$, kde je výsledkom štvorcová matica a prípadom $n = 1$, čo dáva ako výsledok riadkový vektor.

Veta 13. *Za predpokladu, že parameter σ_e^2 je známy, nestranný odhad cAI je v tvare*

$$-2 \log L^P(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\psi}}, \hat{\sigma}_e^2 | \hat{\mathbf{b}}) + 2\Phi_0(\mathbf{Y}),$$

kde

$$\Phi_0(\mathbf{Y}) = \sum_{i=1}^N \text{tr} \left(\frac{\partial \hat{\mathbf{Y}}_i^P}{\partial \mathbf{Y}_i^\top} \right) = \text{tr} \left(\frac{\partial \hat{\mathbf{Y}}^P}{\partial \mathbf{Y}^\top} \right).$$

Poznámka. V prípade známeho reziduálneho rozptylu σ_e^2 a neznámej variančnej matice \mathbb{D} náhodných efektov je ρ z rovnosti (3.18) rovné $\text{tr} \left(\frac{\partial \hat{\mathbf{Y}}^P}{\partial \mathbf{Y}^\top} \right)$.

Dôkaz. Kostra dôkazu je v článku Liang a kol. (2008) a využíva predošlú lemmu. Využijeme najprv vzťah (3.20) z vyčleneného Tvrdenia 3.19 a bližšie popíšeme odvodenie znenia.

$$\rho = \mathbb{E}_{\mathbf{Y}|\mathbf{b}} \frac{1}{\sigma_e^2} (\mathbf{Y} - \boldsymbol{\mu})^\top \hat{\mathbf{Y}}^P = \sum_{i=1}^N \sum_{j=1}^{n_i} \mathbb{E}_{\mathbf{Y}|\mathbf{b}} \frac{1}{\sigma_e^2} (Y_{ij} - \mu_{ij}) \hat{Y}_{ij}^P, \quad (3.29)$$

kde druhá rovnosť vychádza z počítania skalárneho súčinu ako súčet súčinov jednotlivých zložiek vektorov a $\mu_{ij} = \mathbf{x}_{ij}^\top \boldsymbol{\beta} + \mathbf{z}_{ij}^\top \mathbf{b}_i$ je stredná hodnota veličiny Y_{ij} pri podmienení hodnotou \mathbf{b}_i . Následne vyjadríme podmienenú predikciu vzťahom $\hat{\mathbf{Y}}^P = \mathbb{H}(\hat{\boldsymbol{\psi}}) \mathbf{Y}$, kde projekčnú maticu $\mathbb{H}(\hat{\boldsymbol{\psi}})$ tentokrát chápeme po dosadení maximálne vierohodného odhadu $\hat{\boldsymbol{\psi}}$ do vzťahu (3.26). Môžeme sa na ňu teda pozerať ako na nejakú funkciu

$$\hat{\mathbf{Y}}^P = \mathbf{h}(\mathbf{Y}) = \begin{pmatrix} \mathbf{h}_1(\mathbf{Y}_1) \\ \vdots \\ \mathbf{h}_N(\mathbf{Y}_N) \end{pmatrix}, \quad \mathbf{h}_i(\mathbf{Y}_i) = \begin{pmatrix} h_{i1}(Y_{i1}) \\ \vdots \\ h_{in_i}(Y_{in_i}) \end{pmatrix}.$$

Dosadením vyjadrenia predikcie pomocou funkcií h_{ij} do vzťahu (3.29) dostávame

$$\rho = \sum_{i=1}^N \sum_{j=1}^{n_i} \mathbb{E}_{\mathbf{Y}|\mathbf{b}} \frac{Y_{ij} - \mu_{ij}}{\sigma_e^2} h_{ij}(Y_{ij}).$$

Ďalej použijeme znenie Lemmy 12, podľa ktorej

$$\rho = \sum_{i=1}^N \sum_{j=1}^{n_i} \mathbb{E}_{\mathbf{Y}|\mathbf{b}} h'_{ij}(Y_{ij}) = \sum_{i=1}^N \sum_{j=1}^{n_i} \mathbb{E}_{\mathbf{Y}|\mathbf{b}} \frac{\partial \hat{Y}_{ij}^P}{\partial Y_{ij}} = \mathbb{E}_{\mathbf{Y}|\mathbf{b}} \sum_{i=1}^N \text{tr} \left(\frac{\partial \hat{\mathbf{Y}}_i^P}{\partial \mathbf{Y}_i^\top} \right).$$

Stredná hodnota sa dá počítať postupným podmieňovaním, preto platí

$$\mathbf{E}_b(\rho) = \mathbf{E}_b \mathbf{E}_{\mathbf{Y}|b} \sum_{i=1}^N \text{tr} \left(\frac{\partial \widehat{\mathbf{Y}}_i^P}{\partial \mathbf{Y}_i^\top} \right) = \mathbf{E}_{\mathbf{Y},b} \sum_{i=1}^N \text{tr} \left(\frac{\partial \widehat{\mathbf{Y}}_i^P}{\partial \mathbf{Y}_i^\top} \right).$$

Z toho dôvodu stačí brať $\rho = \sum_{i=1}^N \text{tr} \left(\frac{\partial \widehat{\mathbf{Y}}_i^P}{\partial \mathbf{Y}_i^\top} \right)$, ako je v znení vety. □

Hodnotu $\Psi_0(\mathbf{Y})$ môžeme buď explicitne počítať alebo hodnoty parciálnych derivácií $\partial \widehat{Y}_{ij}^P / \partial Y_{ij}$ nahradíme numerickou aproximáciou

$$\frac{\widehat{Y}_{ij}^P(\mathbf{Y} + t \mathbf{e}_{ij}) - \widehat{Y}_{ij}^P}{t},$$

kde $\widehat{Y}_{ij}^P(\mathbf{Y} + t \mathbf{e}_{ij})$ predstavuje predikciu Y_{ij} z dátového súboru, ktorý vznikne z pôvodného pričítaním t k pozorovaniu Y_{ij} . Hodnotu t volíme malú, pretože pri $t \rightarrow 0$ konverguje aproximačný člen k požadovanej parciálnej derivácii. Nevýhodou je, že potrebujeme n opakovaní procesu predikcie v LME modeli.

Explicitný výpočet korekčného členu $\Psi_0(\mathbf{Y})$ môžeme nájsť v článku Greven a Kneib (2010). Pred jeho uvedením ešte zavedieme značenie, ktoré prevezmeme z článku Kubokawa (2011) a je častejšie používané v časti 3.2.3 tejto práce.

Uvažujeme variančné parametre $\boldsymbol{\psi}^*$ z parametrického priestoru $\Psi \subseteq \mathbb{R}^Q$, ktoré pokladáme teraz za neznáme a pre $i, j = 1, \dots, Q$ označíme deriváciu ľubovoľnej matice \mathbb{S} závislej na zložkách parametra $\boldsymbol{\psi}^*$ nasledovne

$$\mathbb{S}_{(i)} = \frac{\partial \mathbb{S}}{\partial \psi_i^*}, \quad \mathbb{S}_{(ij)} = \frac{\partial^2 \mathbb{S}}{\partial \psi_i^* \partial \psi_j^*}.$$

Tvrdenie 14. *Predpokladajme, že σ_e^2 je známe a zvyšné variančné parametre $\boldsymbol{\psi}^*$ z parametrického priestoru $\Psi \subseteq \mathbb{R}^Q$ sú neznáme. Potom sa korekčný člen $\Psi_0(\mathbf{Y})$ z Vety 13 rovná*

$$\Psi_0(\mathbf{Y}) = n - \text{tr}(\widehat{\mathbb{A}}) + \sum_{j=1}^s \mathbf{e}_j^\top \widehat{\mathbb{B}}^{-1} \widehat{\mathbb{G}} \widehat{\mathbb{A}} \widehat{\mathbb{V}}_{(j)} \widehat{\mathbb{A}} \mathbf{Y},$$

kde predpokladáme, že po nejakej permutácii zložiek parametra $\boldsymbol{\psi}^*$ ho usporiadame na $\boldsymbol{\psi}^* = (\boldsymbol{\psi}_s^\top, \boldsymbol{\psi}_t^\top, \boldsymbol{\psi}_{Q-s-t}^\top)^\top$ pre nejaké $0 \leq s \leq Q$ a $0 \leq t \leq Q - s$ také, že

$$\boldsymbol{\psi}_s = \begin{pmatrix} \psi_s^1 \\ \psi_s^2 \end{pmatrix}, \quad \boldsymbol{\psi}_t = \begin{pmatrix} \psi_t^1 \\ \psi_t^2 \end{pmatrix}.$$

Pri tomto usporiadaní pre parametrický priestor platí

$$\Psi = \left\{ \boldsymbol{\psi}^*; \boldsymbol{\psi}_s^1 \in (0, \infty)^{S_1}, \boldsymbol{\psi}_s^2 \in \Psi_s^2 \subseteq \mathbb{R}^{S_2}, \boldsymbol{\psi}_t^1 \in [0, \infty)^{T_1}, \right. \\ \left. \boldsymbol{\psi}_t^2 \in F_t(\psi_{t_i}, \psi_{t_j}) \subset \mathbb{R}^{T_2}, \boldsymbol{\psi}_{Q-s-t} \in F(\boldsymbol{\psi}_s, \boldsymbol{\psi}_t) \subset \mathbb{R}^{Q-s-t} \right\},$$

kde $S_1 + S_2 = s$, $T_1 + T_2 = t$ a $\boldsymbol{\psi}_s^2$ leží vnútri Ψ_s^2 . Pre hodnoty funkcií F_t a F platí $F_t(\psi_{t_i}, 0) = F_t(0, \psi_{t_i}) = 0$ pre všetky zložky t_i parametra $\boldsymbol{\psi}_t^1$ a $F(\boldsymbol{\psi}_s, \mathbf{0}) = \mathbf{0}$ pre všetky $\boldsymbol{\psi}_s$. Maximálne virohodné odhady $\widehat{\boldsymbol{\psi}}_t$ a $\widehat{\boldsymbol{\psi}}_{Q-s-t}$ sú nulové. Ďalej \mathbf{e}_j značí

jednotkový vektor s jednotkou na mieste j , matica \mathbb{V} je zavedená vzťahom (3.11) a matica \mathbb{A} má tvar

$$\mathbb{A} = \widehat{\mathbb{V}}^{-1} - \widehat{\mathbb{V}}^{-1} \mathbb{X} \left(\mathbb{X}^\top \widehat{\mathbb{V}}^{-1} \mathbb{X} \right)^{-1} \mathbb{X}^\top \widehat{\mathbb{V}}^{-1}.$$

Riadok matice \mathbb{G} typu $s \times n$ na mieste j je rovný

$$2 \left[(\mathbf{Y}^\top \mathbb{A} \mathbf{Y}) \mathbf{Y}^\top \mathbb{A} \mathbb{V}_{(j)} \mathbb{A} - (\mathbf{Y}^\top \mathbb{A} \mathbb{V}_{(j)} \mathbb{A} \mathbf{Y}) \mathbf{Y}^\top \mathbb{A} \right]$$

a \mathbb{B} je odvodená z Hessovej matice pre $\boldsymbol{\psi}_s^*$ typu $s \times s$ a má prvky pre $j, k = 1, \dots, s$ v tvare $r_{jk} - s_{jk} - t_{jk}$, kde

$$\begin{aligned} r_{jk} &= \frac{(\mathbf{Y}^\top \mathbb{A} \mathbf{Y})^2}{n} \text{tr} \left(\mathbb{V}_{(jk)} \mathbb{V}^{-1} - \mathbb{V}_{(j)} \mathbb{V}^{-1} \mathbb{V}_{(k)} \mathbb{V}^{-1} \right), \\ s_{jk} &= (\mathbf{Y}^\top \mathbb{A} \mathbb{V}_{(j)} \mathbb{A} \mathbf{Y}) (\mathbf{Y}^\top \mathbb{A} \mathbb{V}_{(k)} \mathbb{A} \mathbf{Y}) \\ t_{jk} &= (\mathbf{Y}^\top \mathbb{A} \mathbf{Y}) \mathbf{Y}^\top (\mathbb{A}^\top \mathbb{V}_{(jk)} \mathbb{A} - 2 \mathbb{A}^\top \mathbb{V}_{(j)} \mathbb{A} \mathbb{V}_{(k)} \mathbb{A}) \mathbf{Y}. \end{aligned}$$

Poznámka. Parametre ψ_s^1 a ψ_t^1 predstavujú diagonálu matice \mathbb{R}_* zo vzťahu (3.11), ktorá je odvodená z rozptylovej matice náhodných efektov. Rozlišujeme tým rozptyly, ktorých maximálne vierohodné odhady sú nenulové a rozptyly, ktoré buď nulové predpokladáme alebo ich odhady ležia na hranici parametrického priestoru. Množina Ψ_s^1 predstavuje všetky možné hodnoty kovariancie medzi zložkami parametra $\boldsymbol{\psi}_s^1$. Rovnako funkcia F_t určuje možné hodnoty kovariancie medzi zložkami parametra $\boldsymbol{\psi}_t^1$ a F zase krížové kovariancie medzi nejakou zložkou $\boldsymbol{\psi}_s^1$ a $\boldsymbol{\psi}_t^1$. Preto požadujeme, aby boli nulové, ak sa zaujímate o kovarianciu s efektom, ktorého rozptyl odhadneme nulou.

Dôkaz. Postup dôkazu môžeme nájsť v článku Greven a Kneib (2010), Theorem 3. Dá sa zhrnúť do štyroch krokov, ktoré v práci podrobnejšie odvodíme, kým v článku sú len stručne okomentované.

- KROK 1: Ukážeme, že platí vyjadrenie $\widehat{\mathbf{Y}}^P = (\mathbb{I}_n - \widehat{\mathbb{V}}^{-1} \widehat{\mathbb{P}}) \mathbf{Y}$, z čoho derivovaním vyplýva

$$\Psi_0(\mathbf{Y}) = \text{tr} \left(\frac{\partial \widehat{\mathbf{Y}}^P}{\partial \mathbf{Y}^\top} \right) = \text{tr} \left(\mathbb{I}_n - \widehat{\mathbb{V}}^{-1} \widehat{\mathbb{P}} - \sum_{j=1}^Q \frac{\partial}{\partial \psi_j^*} (\widehat{\mathbb{V}}^{-1} \widehat{\mathbb{P}}) \mathbf{Y} \frac{\partial \widehat{\psi}_j^*}{\partial \mathbf{Y}^\top} \right).$$

- KROK 2: Ukážeme, že $\frac{\partial}{\partial \psi_j^*} (\widehat{\mathbb{V}}^{-1} \widehat{\mathbb{P}}) = (\widehat{\mathbb{V}}^{-1} \widehat{\mathbb{P}})_{(j)} = -\widehat{\mathbb{A}} \widehat{\mathbb{V}}_{(j)} \widehat{\mathbb{A}}$.
- KROK 3: Pre časť parametrov s nulovými maximálne vierohodnými odhadmi, teda pre $i = 1, \dots, N$, $k = 1, \dots, n_i$ a $j = s + 1, \dots, Q$ sa ukáže, že platí

$$\frac{\partial \widehat{\psi}_j^*}{\partial Y_{ik}} = 0.$$

- KROK 4: Nakoniec pre $j = 1, \dots, s$ ukážeme, že platí

$$\frac{\partial \widehat{\psi}_j^*}{\partial \mathbf{Y}^\top} = \mathbf{e}_j^\top \widehat{\mathbb{B}}^{-1} \widehat{\mathbb{G}},$$

kde \mathbf{e}_j je jednotkový vektor dĺžky s s jednotkou na mieste j a matice \mathbb{B} a \mathbb{G} odpovedajú tvrdeniu.

Dosadením vzťahov z krokov 2 až 4 do vyjadrenia v prvom kroku je znenie tvrdenie dokázané. Venujme sa teda jednotlivým krokom.

KROK 1: Pripomenutím vzťahu $\widehat{\mathbf{Y}}^P = \mathbb{X}\widehat{\boldsymbol{\beta}} + \mathbb{Z}\widehat{\mathbf{b}}$ a vzťahov (3.12) a (1.6) vidíme, že

$$\widehat{\mathbf{Y}}^P = \mathbb{X}\widehat{\boldsymbol{\beta}} + \mathbb{Z}\mathbb{D}_*\mathbb{Z}^\top\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{Y} - \mathbb{X}\widehat{\boldsymbol{\beta}})$$

a taktiež $\mathbb{X}\widehat{\boldsymbol{\beta}} = \mathbb{X}(\mathbb{X}^\top\mathbb{V}^{-1}\mathbb{X})^{-1}\mathbb{X}^\top\mathbb{V}^{-1}\mathbf{Y} = (\mathbb{I}_n - \widehat{\mathbb{P}})\mathbf{Y}$, kde označíme

$$\mathbb{P} = \mathbb{I}_n - \mathbb{X}(\mathbb{X}^\top\mathbb{V}^{-1}\mathbb{X})^{-1}\mathbb{X}^\top\mathbb{V}^{-1}. \quad (3.30)$$

Zo vzťahu pre $\boldsymbol{\Sigma}$ vieme, že $\mathbb{Z}\mathbb{D}_*\mathbb{Z}^\top = \sigma_e^2\mathbb{I}_n - \boldsymbol{\Sigma}$, preto platí

$$\widehat{\mathbf{Y}}^P = (\mathbb{I}_n - \widehat{\mathbb{P}})\mathbf{Y} - \sigma_e^2\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\mathbb{P}}\mathbf{Y} + \widehat{\mathbb{P}}\mathbf{Y} = (\mathbb{I}_n - \sigma_e^2\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\mathbb{P}})\mathbf{Y}.$$

Na dokončenie dôkazu prvého kroku porovnáme vzťah (3.11), podľa ktorého $\sigma_e^2\boldsymbol{\Sigma}^{-1} = \mathbb{V}^{-1}$.

KROK 2: Chceme ukázať, že $(\mathbb{V}^{-1}\mathbb{P})_{(j)} = -\mathbb{A}\mathbb{V}_{(j)}\mathbb{A}$, kde porovnaním definície matice \mathbb{A} v znení rovno vidíme, že $\mathbb{A} = \mathbb{V}^{-1}\mathbb{P}$. Použijeme pravidlo pre derivovanie súčinu, podľa ktorého

$$(\mathbb{V}^{-1}\mathbb{P})_{(j)} = \mathbb{V}_{(j)}^{-1}\mathbb{P} + \mathbb{V}^{-1}\mathbb{P}_{(j)}. \quad (3.31)$$

Ďalej rovnako deriváciou súčinu $\mathbb{S}\mathbb{S}^{-1}$ odvodíme pre ľubovoľnú maticu pravidlo

$$\mathbb{S}_{(j)}^{-1} = -\mathbb{S}^{-1}\mathbb{S}_{(j)}\mathbb{S}^{-1}. \quad (3.32)$$

Vo vyjadrení (3.30) vidíme, že v matici \mathbb{P} závisí na parametri $\boldsymbol{\psi}^*$ len matica \mathbb{V} a navyše

$$\mathbb{P}_{(j)} = -\mathbb{X}(\mathbb{X}^\top\mathbb{V}^{-1}\mathbb{X})^{-1}\mathbb{X}^\top\mathbb{V}^{-1} - \mathbb{X}(\mathbb{X}^\top\mathbb{V}^{-1}\mathbb{X})^{-1}\mathbb{X}^\top\mathbb{V}_{(j)}^{-1}.$$

Dvojnásobným použitím vzťahu (3.31) vypočítame

$$\begin{aligned} (\mathbb{X}^\top\mathbb{V}^{-1}\mathbb{X})_{(j)}^{-1} &= -(\mathbb{X}^\top\mathbb{V}^{-1}\mathbb{X})^{-1}(\mathbb{X}^\top\mathbb{V}^{-1}\mathbb{X})_{(j)}(\mathbb{X}^\top\mathbb{V}^{-1}\mathbb{X})^{-1} = \\ &= (\mathbb{X}^\top\mathbb{V}^{-1}\mathbb{X})^{-1}(\mathbb{X}^\top\mathbb{V}^{-1}\mathbb{V}_{(j)}\mathbb{V}^{-1}\mathbb{X})(\mathbb{X}^\top\mathbb{V}^{-1}\mathbb{X})^{-1}. \end{aligned}$$

Dosadením do predošlého vyjadrenia, kde rovnako dosadíme vzťah pre $\mathbb{V}_{(j)}^{-1}$, dostaneme

$$\begin{aligned} \mathbb{P}_{(j)} &= -\mathbb{X}(\mathbb{X}^\top\mathbb{V}^{-1}\mathbb{X})^{-1}(\mathbb{X}^\top\mathbb{V}^{-1}\mathbb{V}_{(j)}\mathbb{V}^{-1}\mathbb{X})(\mathbb{X}^\top\mathbb{V}^{-1}\mathbb{X})^{-1}\mathbb{X}^\top\mathbb{V}^{-1} + \\ &+ \mathbb{X}(\mathbb{X}^\top\mathbb{V}^{-1}\mathbb{X})^{-1}\mathbb{X}^\top\mathbb{V}^{-1}\mathbb{V}_{(j)}\mathbb{V}^{-1} = \mathbb{X}(\mathbb{X}^\top\mathbb{V}^{-1}\mathbb{X})^{-1}\mathbb{X}^\top\mathbb{V}^{-1}\mathbb{V}_{(j)}\mathbb{V}^{-1}\mathbb{P}. \end{aligned}$$

Toto vyjadrenie spolu so vzťahom pre výpočet $\mathbb{V}_{(j)}^{-1}$ dosadíme do (3.32) a postupne upravíme

$$\begin{aligned} (\mathbb{V}^{-1}\mathbb{P})_{(j)} &= \mathbb{A}_{(j)} = -\mathbb{V}^{-1}\mathbb{V}_{(j)}\mathbb{V}^{-1}\mathbb{P} - \mathbb{V}^{-1}\mathbb{X}(\mathbb{X}^\top\mathbb{V}^{-1}\mathbb{X})^{-1}\mathbb{X}^\top\mathbb{V}^{-1}\mathbb{V}_{(j)}\mathbb{V}^{-1}\mathbb{P} = \\ &= -\mathbb{V}^{-1}(\mathbb{I}_n - \mathbb{X}(\mathbb{X}^\top\mathbb{V}^{-1}\mathbb{X})^{-1}\mathbb{X}^\top\mathbb{V}^{-1})\mathbb{V}_{(j)}\mathbb{V}^{-1}\mathbb{P} = -\mathbb{V}^{-1}\mathbb{P}\mathbb{V}_{(j)}\mathbb{V}^{-1}\mathbb{P}. \end{aligned} \quad (3.33)$$

Keďže $\mathbb{V}^{-1}\mathbb{P} = \mathbb{A}$ znenie druhého kroku je tým dokázané.

KROK 3: Označíme $\overline{\boldsymbol{\psi}}^*$ maximum marginálnej vierohodnosti, ak uvažujeme neohraničený parametrický priestor pre $\boldsymbol{\psi}^* \in \mathbb{R}^Q$. Nakoľko je vierohodnosť normálneho rozdelenia log-konkávna, maximálne vierohodný odhad $\widehat{\boldsymbol{\psi}}^*$, ktorý maximalizuje vierohodnosť na parametrickom priestore Ψ , je projekciou $\overline{\boldsymbol{\psi}}^*$. Tvar

vierohodnosti zohráva kľúčovú rolu aj v tom, že pre $j = s + 1, \dots, s + T_1$ sú zložky odhadu $\bar{\psi}_j^*$ nenulové. Ďalej zo spojitosti a otvorenosti \mathbb{R}^Q vyplýva, že aj body v dostatočne malom okolí bodu $(\bar{\psi}_{s+1}^*, \dots, \bar{\psi}_{s+T_1}^*)^\top$ majú nulu ako projekciu na $[0, \infty)^{T_1}$. Teda pre dostatočne malé $h > 0$ je $\hat{\psi}_j^*(\mathbf{Y} + h\mathbf{e}_{ik}) = 0$, ak ho chápeme ako maximálne vierohodný odhad z vierohodnosti na priestore Ψ po pripočítaní čísla h k pozorovaniu Y_{ik} . Potom teda

$$\lim_{h \rightarrow 0} \frac{\hat{\psi}_j^*(\mathbf{Y} + h\mathbf{e}_{ik}) - \hat{\psi}_j^*}{h} = 0, \quad j = s + 1, \dots, s + T_1; i = 1, \dots, N, k = 1, \dots, n_i.$$

Nulovosť derivácií pre zvyšné $\hat{\psi}_j$ pre $j = s + T_1 + 1, \dots, Q$ vychádza z toho, že funkcie F_t , resp. F určujúce kovariancie sú nulové, ak je aspoň jeden argument nulový a platí to na celom vyššie popísanom okolí. Preto rovnakou úvahou vyplýva, že bude parciálna derivácia nulová.

KROK 4: Na dôkaz posledného kroku využijeme podobu profilovej vierohodnosti pre parametre $\boldsymbol{\psi}^*$, ktorú sme odvodili vo vzťahu (3.13). Pre maximálne vierohodný odhad $\hat{\boldsymbol{\psi}}^*$ platí, že derivácia logaritmu vierohodnosti je nulová. Definujeme preto funkciu

$$h_j(\boldsymbol{\psi}^*, \mathbf{Y}) = \frac{\partial}{\partial \psi_j^*} \left[-2 \log L^M(\boldsymbol{\psi}^* | \hat{\boldsymbol{\beta}}, \hat{\sigma}_e^2) \right].$$

Z vyjadrenia profilovej vierohodnosti parametru $\boldsymbol{\psi}^*$ vo vzťahu (3.13) máme

$$\begin{aligned} -2 \log L^M(\boldsymbol{\psi}^* | \hat{\boldsymbol{\beta}}, \hat{\sigma}_e^2) &= n \log(2\pi/n) + n + \log |\mathbb{V}| + \\ &\quad + n \log \left((\mathbf{Y} - \mathbb{X}\hat{\boldsymbol{\beta}})^\top \mathbb{V}^{-1} (\mathbf{Y} - \mathbb{X}\hat{\boldsymbol{\beta}}) \right). \end{aligned}$$

Ďalej máme $\mathbf{Y} - \mathbb{X}\hat{\boldsymbol{\beta}} = \hat{\mathbb{P}}\mathbf{Y}$. Prvý člen je konštanta a dosadením vzťahu pre deriváciu logaritmu z determinantu matice dostaneme

$$h_j(\boldsymbol{\psi}^*, \mathbf{Y}) = \text{tr} \left(\mathbb{V}^{-1} \mathbb{V}_{(j)} \right) + n \frac{\left(\mathbf{Y}^\top \hat{\mathbb{P}}^\top \mathbb{V}^{-1} \hat{\mathbb{P}} \mathbf{Y} \right)_{(j)}}{\mathbf{Y}^\top \hat{\mathbb{P}}^\top \mathbb{V}^{-1} \hat{\mathbb{P}} \mathbf{Y}}.$$

Pomocou vyjadrenia matice \mathbb{A} môžeme prepísať do tvaru

$$h_j(\boldsymbol{\psi}^*, \mathbf{Y}) = \text{tr} \left(\mathbb{V}^{-1} \mathbb{V}_{(j)} \right) + n \frac{\left(\mathbf{Y}^\top \mathbb{A}^\top \mathbb{V} \mathbb{A} \mathbf{Y} \right)_{(j)}}{\mathbf{Y}^\top \hat{\mathbb{A}}^\top \mathbb{V} \hat{\mathbb{A}} \mathbf{Y}}. \quad (3.34)$$

Deriváciu čitateľa v predošlom vyjadrení počítame ako deriváciu súčinu, čím dostaneme 3 členy, z ktorých budú však dva rovnaké. Platí

$$\left(\mathbf{Y}^\top \mathbb{A}^\top \mathbb{V} \mathbb{A} \mathbf{Y} \right)_{(j)} = 2 \mathbf{Y}^\top \mathbb{A}_{(j)}^\top \mathbb{V} \mathbb{A} \mathbf{Y} + \mathbf{Y}^\top \mathbb{A}^\top \mathbb{V}_{(j)} \mathbb{A} \mathbf{Y}.$$

V druhom kroku výrazu sme dokázali vzťah pre deriváciu $\mathbb{A}_{(j)}$, ktorý ďalej dosadíme a upravíme na

$$\left(\mathbf{Y}^\top \mathbb{A}^\top \mathbb{V} \mathbb{A} \mathbf{Y} \right)_{(j)} = -2 \mathbf{Y}^\top \mathbb{A}^\top \mathbb{V}_{(j)} \mathbb{A}^\top \mathbb{V} \mathbb{A} \mathbf{Y} + \mathbf{Y}^\top \mathbb{A}^\top \mathbb{V}_{(j)} \mathbb{A} \mathbf{Y}.$$

Ďalej jednoduchým počítaním zistíme, že matica \mathbb{P} je idempotentná, pretože

$$\mathbb{X}(\mathbb{X}^\top \mathbb{V}^{-1} \mathbb{X})^{-1} \mathbb{X}^\top \mathbb{V}^{-1} \mathbb{X}(\mathbb{X}^\top \mathbb{V}^{-1} \mathbb{X})^{-1} \mathbb{X}^\top \mathbb{V}^{-1} = \mathbb{X}(\mathbb{X}^\top \mathbb{V}^{-1} \mathbb{X})^{-1} \mathbb{X}^\top \mathbb{V}^{-1}$$

a \mathbb{P} je rozdielom jednotkovej matice a $\mathbb{X}(\mathbb{X}^\top \mathbb{V}^{-1} \mathbb{X})^{-1} \mathbb{X}^\top \mathbb{V}^{-1}$, ktorá je podľa hornej rovnosti idempotentná. Preto môžeme odvodiť vzťah

$$\begin{aligned}\mathbb{V}\mathbb{A} &= \mathbb{P}, \\ \mathbb{P}\mathbb{V}\mathbb{A} &= \mathbb{P}, \\ \mathbb{P}(\mathbb{V}\mathbb{A} - \mathbb{I}_n) &= \mathbf{0}_{n \times n}, \\ \mathbb{V}\mathbb{A} &= \mathbb{I}_n.\end{aligned}$$

Matica \mathbb{A} je symetrická, lebo je rozdielom dvoch symetrických matíc a navyše dosadením predošlej poslednej rovnosti do prvého výrazu vyjadrenia (3.34) odvodíme

$$\left(\mathbf{Y}^\top \mathbb{A}^\top \mathbb{V} \mathbb{A} \mathbf{Y}\right)_{(j)} = -\mathbf{Y}^\top \mathbb{A} \mathbb{V}_{(j)} \mathbb{A} \mathbf{Y}.$$

Celkovo máme teda

$$h_j(\boldsymbol{\psi}^*, \mathbf{Y}) = \text{tr}\left(\mathbb{V}^{-1} \mathbb{V}_{(j)}\right) - n \frac{\mathbf{Y}^\top \mathbb{A} \mathbb{V}_{(j)} \mathbb{A} \mathbf{Y}}{\mathbf{Y}^\top \mathbb{A} \mathbf{Y}}. \quad (3.35)$$

Vieme, že pre každé $j = 1, \dots, s$ je $h_j(\widehat{\boldsymbol{\psi}}, \mathbf{Y}) = 0$. Ak budeme derivovať ešte raz podľa \mathbf{Y} , zistíme retiazkovým pravidlom

$$\mathbf{0} = \frac{\partial}{\partial \mathbf{Y}^\top} h_j(\widehat{\boldsymbol{\psi}}^*, \mathbf{Y}) = \sum_{k=1}^s \frac{\partial h_j(\widehat{\boldsymbol{\psi}}, \mathbf{Y})}{\partial \psi_k^*} \frac{\partial \widehat{\psi}_k^*}{\partial \mathbf{Y}^\top} + \frac{\partial h_j(\widehat{\boldsymbol{\psi}}, \mathbf{Y})}{\partial \mathbf{Y}^\top}.$$

Úpravou poslednej rovnosti dostaneme pre parameter $\boldsymbol{\psi}_s^* = (\psi_1^*, \dots, \psi_s^*)^\top$ vzťah

$$\frac{\partial \widehat{\boldsymbol{\psi}}_s^*}{\partial \mathbf{Y}^\top} = - \left[\frac{\partial h_j(\widehat{\boldsymbol{\psi}}, \mathbf{Y})}{\partial \psi_k^*} \right]_{j,k=1,\dots,s}^{-1} \frac{\partial h(\widehat{\boldsymbol{\psi}}, \mathbf{Y})}{\partial \mathbf{Y}^\top},$$

kde $h = (h_1, \dots, h_s)$. Derivácia podľa \mathbf{Y} nám dá tvar matice \mathbb{G} , pretože

$$\frac{\partial h_j(\widehat{\boldsymbol{\psi}}, \mathbf{Y})}{\partial \mathbf{Y}^\top} = \mathbf{e}_j^\top \frac{\partial h(\widehat{\boldsymbol{\psi}}, \mathbf{Y})}{\partial \mathbf{Y}^\top}.$$

Deriváciou kvadratických foriem odvodíme, že

$$\frac{\partial h_j(\widehat{\boldsymbol{\psi}}, \mathbf{Y})}{\partial \mathbf{Y}^\top} = -n \frac{2(\mathbf{Y}^\top \mathbb{A} \mathbf{Y}) \mathbf{Y}^\top \mathbb{A} \mathbb{V}_{(j)} \mathbb{A} - 2(\mathbf{Y}^\top \mathbb{A} \mathbb{V}_{(j)} \mathbb{A} \mathbf{Y}) \mathbf{Y}^\top \mathbb{A}}{(\mathbf{Y}^\top \mathbb{A} \mathbf{Y})^2},$$

čo je vzťah udávajúci riadok matice \mathbb{G} až na násobenie faktorom $-n/(\mathbf{Y}^\top \mathbb{A} \mathbf{Y})^2$. Prvky \mathbb{B} stačí voliť ako prvky Hessovej matice pre parameter $\boldsymbol{\psi}_s^*$, a síce

$$\begin{aligned}\frac{\partial h_j(\widehat{\boldsymbol{\psi}}, \mathbf{Y})}{\partial \psi_k^*} &= \text{tr}\left(\mathbb{V}_k^{-1} \mathbb{V}_{(j)} + \mathbb{V}^{-1} \mathbb{V}_{(jk)}\right) - n \frac{(\mathbf{Y}^\top \mathbb{A} \mathbf{Y}) \mathbf{Y}^\top (\mathbb{A} \mathbb{V}_{(j)} \mathbb{A})_{(k)} \mathbf{Y}}{(\mathbf{Y}^\top \mathbb{A} \mathbf{Y})^2} + \\ &+ n \frac{(\mathbf{Y}^\top \mathbb{A} \mathbb{V}_{(j)} \mathbb{A} \mathbf{Y}) \mathbf{Y}^\top \mathbb{A}_{(k)} \mathbf{Y}}{(\mathbf{Y}^\top \mathbb{A} \mathbf{Y})^2} = \frac{n}{(\mathbf{Y}^\top \mathbb{A} \mathbf{Y})^2} [r_{jk} - t_{jk} - s_{jk}],\end{aligned}$$

kde výrazy r_{jk} , s_{jk} a t_{jk} sú v znení tvrdenia. Na určenie r_{jk} použijeme vzťah (3.32), podľa ktorého je

$$\text{tr}\left(\mathbb{V}_{(k)}^{-1} \mathbb{V}_{(j)} + \mathbb{V}^{-1} \mathbb{V}_{(jk)}\right) = \text{tr}\left(\mathbb{V}^{-1} \mathbb{V}_{(jk)} - \mathbb{V}^{-1} \mathbb{V}_k \mathbb{V}^{-1} \mathbb{V}_{(j)}\right).$$

Na odvodenie s_{jk} stačí dosadiť znenie z druhého kroku dôkazu, pretože máme $\mathbb{A}_{(k)} = -\mathbb{A}\mathbb{V}_{(k)}\mathbb{A}$. To taktiež použijeme k tomu, aby sme odvodili t_{jk} , pretože deriváciou súčinu dostaneme vzťah

$$\left(\mathbb{A}\mathbb{V}_{(j)}\mathbb{A}\right)_{(k)} = 2\mathbb{A}_{(k)}\mathbb{V}_{(j)}\mathbb{A} + \mathbb{A}\mathbb{V}_{(jk)}\mathbb{A} = \mathbb{A}\mathbb{V}_{(jk)}\mathbb{A} - 2\mathbb{A}\mathbb{V}_{(k)}\mathbb{A}\mathbb{V}_{(j)}\mathbb{A}.$$

Týmto je ukázaný aj posledný krok dôkazu. □

Poznámka. Výraz $n - \text{tr}(\hat{\mathbb{A}})$ môžeme v znení nahradiť aj výrazom $\text{tr}(\hat{\mathbb{H}})$. Pretože vzťah sme v dôkaze získali ako stopu matice $\mathbb{I}_n - \mathbb{A}$ a bolo ukázané, že to je odhad projekčnej matice vektoru \mathbf{Y} na $\hat{\mathbf{Y}}^P$. Rovnako je definovaná matica \mathbb{H} .

Nestranný odhad sa dá nájsť aj v prípade, keď je σ_e^2 neznáme, no matica náhodných efektov po preškáľovaní známa. Tento výsledok je uvedený v článku Vaida a Blanchard (2005).

Veta 15. *Za predpokladu, že matica $\sigma_e^2\mathbb{D}_*^{-1}$ je známa a parameter σ_e^2 neznámy, nestranný odhad cAI je daný rovnosťou (3.18), v ktorej je korekčný člen ρ rovný*

$$\rho_0 \equiv \rho = \frac{n(n-p-1)}{(n-p)(n-p-2)} (\text{tr}(\mathbb{H}) + 1) + \frac{n(p+1)}{(n-p)(n-p-2)},$$

kde pripomíname, že p je počet pevných efektov modelu, $n = n_1 + \dots + n_N$ a \mathbb{H} je projekčná matica vektora \mathbf{Y} na predikciu $\hat{\mathbf{Y}}^P$.

Dôkaz. Viď článok Vaida a Blanchard (2005), Theorem 2. □

Poznámka. Predpoklad, že rozptylová matica náhodných efektov po preškáľovaní, teda $\sigma_e^{-1}\mathbb{D}_*$, je známa, odpovedá tomu, že považujeme projekčnú maticu \mathbb{H} za známu aj napriek neistote v parametri σ_e^2 .

Poznámka. Pri predpokladoch Vety 15 sa dá úpravami ρ_0 ukázať (viď článok Vaida a Blanchard (2005), Proposition 1)

$$\frac{\rho_0}{\text{tr}(\mathbb{H}) + 1} \xrightarrow{N \rightarrow \infty} 1.$$

Asymptoticky nestranný odhad cAI je preto pri znalosti \mathbb{H} vzťah (3.18) s korekčným členom $\rho = \text{tr}(\mathbb{H}) + 1$. Oproti prípadu, keď sú všetky variančné parametre známe (viď Veta 11), zväčšíme člen o jedna kvôli neistote v odhadovaní σ_e^2 a máme asymptoticky nestranný odhad.

3.2.3 Asymptotická nestrannosť AIC

Ďalej sa budeme venovať tomu, v akom prípade je marginálne kritérium aspoň asymptoticky nestranným odhadom Akaikovej informácie pre lineárne zmiešané modely. Vychádzame z článku Kubokawa (2011) a najprv zavedieme značenie podobne ako v spomínanom článku, z ktorého máme výsledky.

Variančné parametre LME modelu budeme v tejto časti označovať parametrom $\boldsymbol{\tau} = (\sigma_e^2, \boldsymbol{\psi}^\top)^\top$. Ako sme zaviedli v časti o marginálnom AIC, predpokladáme,

že ich je spolu $Q + 1$, niekedy ich budeme indexovať $\tau_0, \tau_1, \dots, \tau_Q$, pričom vždy je $\tau_0 = \sigma_e^2$. Ďalej preberieme značenie pre parciálne derivácie podľa parametrov

$$\mathbf{A}_{(i)}(\boldsymbol{\tau}) = \frac{\partial \mathbf{A}(\boldsymbol{\tau})}{\partial \tau_i}, \quad \mathbf{A}_{(ij)}(\boldsymbol{\tau}) = \frac{\partial^2 \mathbf{A}(\boldsymbol{\tau})}{\partial \tau_i \partial \tau_j}, \quad \mathbf{A}_{(ijk)}(\boldsymbol{\tau}) = \frac{\partial^3 \mathbf{A}(\boldsymbol{\tau})}{\partial \tau_i \partial \tau_j \partial \tau_k},$$

kde $\mathbf{A}(\boldsymbol{\tau})$ môže byť skalár, vektor alebo nejaká matica. Výsledok po parciálnom derivovaní je rovnako skalár, vektor alebo matica a každou jej zložkou je parciálna derivácia príslušného rádu podľa premenných uvedených v dolnom indexe v zátvorke.

Ďalej pracujeme s vlastnými číslami rozptylovej matice marginálneho rozdelenia \mathbf{Y} , ktorú značíme $\boldsymbol{\Sigma}$ a jej tvar nájdeme pod vzťahom (1.1). Najprv označíme $\lambda_1 \leq \dots \leq \lambda_n$ vlastné čísla matice $\boldsymbol{\Sigma}$, ktoré sú nezáporné, pretože predpokladáme, že pracujeme s pozitívne semidefinitnou maticou. Navyiac pre $s = 1, \dots, n$ označíme vlastné čísla $\lambda_s^i, \lambda_s^{ij}$ a λ_s^{ijk} matíc $\boldsymbol{\Sigma}_{(i)}, \boldsymbol{\Sigma}_{(ij)}$, resp. $\boldsymbol{\Sigma}_{(ijk)}$ tak, že sú zoradené $|\lambda_1^i| \leq \dots \leq |\lambda_n^i|, |\lambda_1^{ij}| \leq \dots \leq |\lambda_n^{ij}|$ a nakoniec $|\lambda_1^{ijk}| \leq \dots \leq |\lambda_n^{ijk}|$ pre $0 \leq i, j, k \leq Q$.

Zavedieme predpoklady potrebné na odvodenie asymptotickej nestrannosti odhadov Akaikovej informácie v LME modeloch. Budeme pracovať so symbolmi O a O_p , ktoré vyjadrujú asymptotické vlastnosti.

(P1) Prvky matice \mathbb{X} a \mathbb{D} sú ohraničené, σ_e^2, p a q sú konečné, matica $\mathbb{X}^\top \mathbb{X}$ je pozitívne definitná a $\mathbb{X}^\top \mathbb{X}/n$ konverguje k nejakej pozitívne definitnej matici.

(P2) Matica $\boldsymbol{\Sigma}$ je trikrát spojito diferencovateľná podľa parametra $\boldsymbol{\tau}$ a ďalej pre všetky $0 \leq i, j, k \leq Q$ platí

$$\lim_{n \rightarrow \infty} \lambda_1 > 0, \quad \lim_{n \rightarrow \infty} \lambda_n < \infty, \quad \lim_{n \rightarrow \infty} |\lambda_n^i| < \infty, \quad \lim_{n \rightarrow \infty} |\lambda_n^{ij}| < \infty, \quad \lim_{n \rightarrow \infty} |\lambda_n^{ijk}| < \infty.$$

(P3) Odhad $\hat{\boldsymbol{\tau}} = \hat{\boldsymbol{\tau}}(\mathbf{Y})$ variančných parametrov je

(a) symetrický, t.j. $\hat{\boldsymbol{\tau}}(-\mathbf{Y}) = \hat{\boldsymbol{\tau}}(\mathbf{Y})$,

(b) invariantný voči posunu regresorov, teda $\hat{\boldsymbol{\tau}}(\mathbf{Y} + \mathbb{X}\boldsymbol{\alpha}) = \hat{\boldsymbol{\tau}}(\mathbf{Y})$ pre ľubovoľný p -rozmerný vektor $\boldsymbol{\alpha}$.

(P4) Predpokladáme, že existuje rozvoj

$$\hat{\boldsymbol{\tau}} - \boldsymbol{\tau} = \hat{\boldsymbol{\tau}}^* + \hat{\boldsymbol{\tau}}^{**} + O_p(n^{-3/2}),$$

kde $\hat{\boldsymbol{\tau}}^* = O_p(n^{-1/2})$, $\hat{\boldsymbol{\tau}}^{**} = O_p(n^{-1})$ a $\mathbf{E} \hat{\boldsymbol{\tau}}^* = \mathbf{0}$.

(P5) Členy $\hat{\boldsymbol{\tau}}^* = (\hat{\boldsymbol{\tau}}_0^*, \dots, \hat{\boldsymbol{\tau}}_Q^*)^\top$ a $\hat{\boldsymbol{\tau}}^{**} = (\hat{\boldsymbol{\tau}}_0^{**}, \dots, \hat{\boldsymbol{\tau}}_Q^{**})^\top$ splňujú pre $0 \leq i, j \leq Q$

$$\begin{aligned} \mathbf{E} \operatorname{tr} \left(\frac{\partial^2 \hat{\boldsymbol{\tau}}_i^{**}}{\partial \mathbf{Y} \partial \mathbf{Y}^\top} \right) &= O(n^{-1}), \\ \mathbf{E} \operatorname{tr} \left(\frac{\partial^2 \hat{\boldsymbol{\tau}}_i^*}{\partial \mathbf{Y} \partial \mathbf{Y}^\top} \hat{\boldsymbol{\tau}}_j^* \right) &= O(n^{-1}), \\ \mathbf{E} \operatorname{tr} \left(\left(\frac{\partial \hat{\boldsymbol{\tau}}_i^*}{\partial \mathbf{Y}} \right) \left(\frac{\partial \hat{\boldsymbol{\tau}}_j^*}{\partial \mathbf{Y}} \right)^\top \right) &= O(n^{-1}). \end{aligned}$$

Poznámka. Predpoklad (P2), zjednodušene povedané, tvrdí, že pri dostatočnom počte pozorovaní je rozptylová matica Σ pozitívne definitná s konečnými prvkami a matice vzniknuté deriváciami jednotlivých zložiek do tretieho rádu podľa τ majú taktiež konečné prvky.

Poznámka. Predpoklad (P4) môžeme v rade prípadov zaručiť napríklad Taylorovým rozvojom, po ktorom požadujeme, aby boli jednotlivé členy rozumné. Ako rozumné presne určuje spomínaný predpoklad.

Poznámka. Všetky tieto predpoklady sú potrebné pre asymptotickú neustrannosť marginálneho Akaikovho kritéria (mAIC). V našom prípade sú postačujúce aj pre podmienené Akaikove kritérium (cAIC). V článku Kubokawa (2011) pracujú so všeobecnejším tvarom matice podmieneného rozdelenia vektora \mathbf{Y} pri známych hodnotách pevných efektov. Pre túto maticu potom požadujú splnenie ekvivalentného predpokladu ako predpoklad (P2). V našom prípade uvažujeme túto maticu diagonálnu s prvkami σ_e^2 na diagonále, preto stačí predpoklad $\sigma_e^2 > 0$.

Pred formuláciou vety o asymptoticky neustrannom odhade AI sformulujeme pomocné lemmy, ktoré sú potrebné k dôkazu. V prvej z nich budeme potrebovať deriváciu funkcie (zobrazenia) z \mathbb{R}^n do \mathbb{R}^n . Pre $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ absolútne spojitú, $\mathbf{f}(\mathbf{Y}) = (f_1(\mathbf{Y}), \dots, f_n(\mathbf{Y}))^\top$, budeme rozumieť

$$\frac{d\mathbf{f}(\mathbf{Y})}{d\mathbf{Y}} = \sum_{i=1}^n \frac{\partial f_i(\mathbf{Y})}{\partial Y_i},$$

kde $\partial f_i(\mathbf{Y})/\partial Y_i$ je parciálna derivácia zložky funkcie \mathbf{f} podľa jednej premennej vektora \mathbf{Y} .

Lemma 16.

(i) *Nech má vektor \mathbf{Y} mnohorozmerné normálne rozdelenie so strednou hodnotou $\boldsymbol{\mu}$ a jednotkovou rozptylovou maticou \mathbb{I}_n , píšeme $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \mathbb{I}_n)$. Nech je funkcia $\mathbf{h} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ absolútne spojitá. Potom platia vzťahy*

$$\mathbb{E} \left\| \mathbf{Y} + \mathbf{h}(\mathbf{Y}) - \boldsymbol{\mu} \right\|^2 = \mathbb{E} \left[\left\| \mathbf{Y} - \boldsymbol{\mu} \right\|^2 + 2 \frac{d\mathbf{h}(\mathbf{Y})}{d\mathbf{Y}} + \left\| \mathbf{h}(\mathbf{Y}) \right\|^2 \right], \quad (3.36)$$

$$\mathbb{E} \left[(\mathbf{Y} - \boldsymbol{\mu})^\top \mathbf{h}(\mathbf{Y}) \right] = \mathbb{E} \frac{d\mathbf{h}(\mathbf{Y})}{d\mathbf{Y}}. \quad (3.37)$$

(ii) *Nech má vektor $\mathbf{Z} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$ a $\mathbf{g} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ je absolútne spojitá funkcia. Potom*

$$\mathbb{E} \left[(\mathbf{Z} - \boldsymbol{\mu})^\top \mathbf{g}(\mathbf{Z}) \right] = \mathbb{E} \left[\frac{d}{d\mathbf{Z}} (\Sigma \mathbf{g}(\mathbf{Z})) \right].$$

Dôkaz. Rovnosť (3.36) je dokázaná na strane 351 v publikácii Stein (1973). Druhá rovnosť prvého bodu vychádza z lineariry strednej hodnoty a nasledujúcej identity

$$\left\| \mathbf{Y} + \mathbf{h}(\mathbf{Y}) - \boldsymbol{\mu} \right\|^2 = \left\| \mathbf{Y} - \boldsymbol{\mu} \right\|^2 + 2 (\mathbf{Y} - \boldsymbol{\mu})^\top \mathbf{h}(\mathbf{Y}) + \left\| \mathbf{h}(\mathbf{Y}) \right\|^2.$$

Platnosť druhého bodu vyplynie z toho, že vyjadríme $\mathbf{g}(\mathbf{Z}) = \mathbf{g}(\mathbf{h}(\mathbf{Y}))$, pre vektor $\mathbf{Y} = \Sigma^{-1/2}(\mathbf{Z} - \boldsymbol{\mu}) \sim \mathcal{N}_n(\mathbf{0}, \mathbb{I}_n)$ a funkciu $\mathbf{h}(\mathbf{Y}) = \Sigma^{1/2}\mathbf{Y} + \boldsymbol{\mu}$. Vektor \mathbf{Y} splňuje podmienky bodu (i), podľa ktorého pre funkciu $\mathbf{g} \circ \mathbf{h}$ platí

$$\mathbb{E} \left[\mathbf{Y}^\top \mathbf{g}(\mathbf{h}(\mathbf{Y})) \right] = \mathbb{E} \frac{d\mathbf{g}(\mathbf{h}(\mathbf{Y}))}{d\mathbf{Y}}.$$

Derivovaním zloženej funkcie dostaneme

$$\begin{aligned} \frac{d\mathbf{g}(\mathbf{h}(\mathbf{Y}))}{d\mathbf{Y}} &= \sum_{i=1}^n \frac{\partial g_i(\mathbf{Z})}{\partial Z_i} \frac{\partial [\boldsymbol{\Sigma}^{1/2}\mathbf{Y} + \boldsymbol{\mu}]_i}{\partial Y_i} = \sum_{i=1}^n \frac{\partial g_i(\mathbf{Z})}{\partial Z_i} [\boldsymbol{\Sigma}^{1/2}]_{i,i} = \\ &= \frac{d}{d\mathbf{Z}} (\boldsymbol{\Sigma}^{1/2}\mathbf{g}(\mathbf{Z})), \end{aligned} \quad (3.38)$$

kde $[\mathbf{x}]_i$ vyjadruje zložku x_i vektora \mathbf{x} a $[\mathbb{A}]_{i,i}$ prvok $a_{i,i}$ z diagonály matice \mathbb{A} . Vyjadrenie vektora \mathbf{Y} vložíme do vzťahu (3.38) a dostaneme

$$\mathbb{E} [(\mathbf{Z} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1/2} \mathbf{g}(\mathbf{Z})] = \mathbb{E} \left[\frac{d}{d\mathbf{Z}} (\boldsymbol{\Sigma}^{1/2} \mathbf{g}(\mathbf{Z})) \right].$$

Matica $\boldsymbol{\Sigma}$ nezávisí na hodnotách \mathbf{Z} , preto je predchádzajúca rovnosť ekvivalentná zneniu bodu (ii). □

Lemma 17. *Nech má vektor $\mathbf{Z} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, \mathbb{A} je matrica konštant typu $n \times n$, ktorej zložky nezávisia na hodnote vektora \mathbf{Z} . Buď $h : \mathbb{R}^n \rightarrow \mathbb{R}$ dvakrát diferencovateľná funkcia vzhľadom ku \mathbf{Z} . Potom*

$$\mathbb{E} [\mathbf{u}^\top \mathbb{A} \mathbf{u} h(\mathbf{Z})] = \text{tr}(\boldsymbol{\Sigma} \mathbb{A}) \mathbb{E} [h(\mathbf{Z})] + \text{tr} \left(\boldsymbol{\Sigma} \mathbb{A} \boldsymbol{\Sigma} \mathbb{E} \left[\frac{\partial^2 h(\mathbf{Z})}{\partial \mathbf{Z} \partial \mathbf{Z}^\top} \right] \right),$$

kde $\mathbf{u} = \mathbf{Z} - \boldsymbol{\mu}$.

Dôkaz. Stačí dvakrát použiť znenie Lemmy 16. Najprv použijeme znenie s funkciou $\mathbf{g}_1(\mathbf{Z}) = \mathbb{A} \mathbf{u} h(\mathbf{Z})$ a dostaneme

$$\mathbb{E} [\mathbf{u}^\top \mathbb{A} \mathbf{u} h(\mathbf{Z})] = \mathbb{E} \left[\frac{d}{d\mathbf{Z}} (\boldsymbol{\Sigma} \mathbb{A} \mathbf{u} h(\mathbf{Z})) \right].$$

Derivovaním súčinu odvodíme

$$\frac{d}{d\mathbf{Z}} (\boldsymbol{\Sigma} \mathbb{A} \mathbf{u} h(\mathbf{Z})) = \sum_{i=1}^n \frac{\partial [\boldsymbol{\Sigma} \mathbb{A} \mathbf{u}]_i}{\partial Z_i} h(\mathbf{Z}) + \sum_{i=1}^n [\boldsymbol{\Sigma} \mathbb{A} \mathbf{u}]_i \frac{\partial h(\mathbf{Z})}{\partial Z_i}.$$

Vieme, že $\partial u_i / \partial Z_i = 1$ pre všetky $i = 1, \dots, n$. Preto odvodíme vzťah

$$\mathbb{E} [\mathbf{u}^\top \mathbb{A} \mathbf{u} h(\mathbf{Z})] = \text{tr}(\boldsymbol{\Sigma} \mathbb{A}) \mathbb{E} [h(\mathbf{Z})] + \mathbb{E} \left[\mathbf{u}^\top \mathbb{A}^\top \boldsymbol{\Sigma} \frac{\partial h(\mathbf{Z})}{\partial \mathbf{Z}} \right].$$

Na druhý člen predošlej rovnosti použijeme predchádzajúcu lemmu s funkciou $\mathbf{g}_2 = \mathbb{A}^\top \boldsymbol{\Sigma} (\partial h(\mathbf{Z}) / \partial \mathbf{Z})$ a odvodíme tak druhý člen zo znenia. □

Pri dôkazoch o asymptotickej nestrannosti kritéria budeme pracovať s Taylorovým rozvojom, preto zavedieme značenie pre niektoré používané sumy

$$\sum_i \equiv \sum_{i=0}^Q, \quad \sum_{i,j} \equiv \sum_{i=0}^Q \sum_{j=0}^Q \quad \text{a} \quad \sum_{i,j,k} \equiv \sum_{i=0}^Q \sum_{j=0}^Q \sum_{k=0}^Q.$$

Veta 18. *Predpokladajme, že sú splnené (P1)–(P5), potom platí*

$$\mathbf{E}_{\mathbf{Y}} \left[-2 \log L^M(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\psi}}, \hat{\sigma}_e^2) \right] = \text{AI} - 2p - 2h_M(\boldsymbol{\tau}) + O(n^{-1/2}),$$

kde

$$h_M(\boldsymbol{\tau}) = \frac{1}{2} \sum_i \mathbf{E}_{\mathbf{Y}} \left[\text{tr} \left(\boldsymbol{\Sigma}^{(i)} \frac{\partial^2 \hat{\boldsymbol{\tau}}_i^*}{\partial \mathbf{Y} \partial \mathbf{Y}^\top} \right) \right].$$

Dôkaz. Dôkaz sa dá nájsť v článku Kubokawa (2011), Theorem 2.1. Kroky s úpravou vierohodnosti oproti článku podrobnejšie zdôvodníme v nasledujúcom dôkaze. Pripomenieme vzťah pre logaritmus používanej vierohodnosti, ktorý vychádza z (3.9)

$$-2 \log L^M(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\psi}}, \hat{\sigma}_e^2) = n \log(2\pi) + \log |\hat{\boldsymbol{\Sigma}}| + (\mathbf{Y} - \mathbb{X}\hat{\boldsymbol{\beta}})^\top \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{Y} - \mathbb{X}\hat{\boldsymbol{\beta}}).$$

Rovnako pripomenieme Definíciu 6 Akaikovej informácie

$$\text{AI} = \mathbf{E}_{\mathbf{Y}} \mathbf{E}_{\mathbf{Y}^*} \left[n \log(2\pi) + \log |\hat{\boldsymbol{\Sigma}}| + (\mathbf{Y}^* - \mathbb{X}\hat{\boldsymbol{\beta}})^\top \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{Y}^* - \mathbb{X}\hat{\boldsymbol{\beta}}) \right].$$

Pretože $\hat{\boldsymbol{\Sigma}}$ závisí len na vektore \mathbf{Y} , pre vychýlenie platí

$$\begin{aligned} \text{B} \equiv \mathbf{E}_{\mathbf{Y}} \left[-2 \log L^M(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\psi}}, \hat{\sigma}_e^2) \right] - \text{AI} &= \mathbf{E}_{\mathbf{Y}} \left[(\mathbf{Y} - \mathbb{X}\hat{\boldsymbol{\beta}})^\top \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{Y} - \mathbb{X}\hat{\boldsymbol{\beta}}) - \right. \\ &\quad \left. - \mathbf{E}_{\mathbf{Y}^*} (\mathbf{Y}^* - \mathbb{X}\hat{\boldsymbol{\beta}})^\top \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{Y}^* - \mathbb{X}\hat{\boldsymbol{\beta}}) \right]. \end{aligned}$$

Ďalej pričítaním a odčítaním skutočnej strednej hodnoty $\mathbb{X}\boldsymbol{\beta}$ v zátvorkách použijeme identitu

$$\begin{aligned} (\mathbf{Y}^\perp - \mathbb{X}\hat{\boldsymbol{\beta}})^\top \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{Y}^\perp - \mathbb{X}\hat{\boldsymbol{\beta}}) &= (\mathbf{Y}^\perp - \mathbb{X}\boldsymbol{\beta})^\top \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{Y}^\perp - \mathbb{X}\boldsymbol{\beta}) + \\ &\quad + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^\top \mathbb{X}^\top \hat{\boldsymbol{\Sigma}}^{-1} \mathbb{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + 2 (\mathbf{Y}^\perp - \mathbb{X}\boldsymbol{\beta})^\top \hat{\boldsymbol{\Sigma}}^{-1} \mathbb{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}), \end{aligned} \quad (3.39)$$

kde za \mathbf{Y}^\perp môžeme dosadiť \mathbf{Y} alebo \mathbf{Y}^* . Použitím Lemmy 17, kde za funkciu $h(\mathbf{Y}^*)$ volíme identicky jednotku, vypočítame

$$\mathbf{E}_{\mathbf{Y}^*} (\mathbf{Y}^* - \mathbb{X}\boldsymbol{\beta})^\top \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{Y}^* - \mathbb{X}\boldsymbol{\beta}) = \text{tr} \left(\boldsymbol{\Sigma} \hat{\boldsymbol{\Sigma}}^{-1} \right). \quad (3.40)$$

Odhady $\hat{\boldsymbol{\beta}}$ a $\hat{\boldsymbol{\Sigma}}$ sú nezávislé na \mathbf{Y}^* , preto platí

$$\mathbf{E}_{\mathbf{Y}^*} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^\top \mathbb{X}^\top \hat{\boldsymbol{\Sigma}}^{-1} \mathbb{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) = (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^\top \mathbb{X}^\top \hat{\boldsymbol{\Sigma}}^{-1} \mathbb{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}), \quad (3.41)$$

$$\mathbf{E}_{\mathbf{Y}^*} (\mathbf{Y}^* - \mathbb{X}\boldsymbol{\beta})^\top \hat{\boldsymbol{\Sigma}}^{-1} \mathbb{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) = 0. \quad (3.42)$$

Dosadením identity (3.39) do výpočtu vychýlenia pre \mathbf{Y} a \mathbf{Y}^* s použitím výpočtu stredných hodnôt vo vzťahoch (3.40), (3.41) a (3.42) upravíme vychýlenie do tvaru

$$\text{B} = \mathbf{E}_{\mathbf{Y}} \left[(\mathbf{Y} - \mathbb{X}\boldsymbol{\beta})^\top \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{Y} - \mathbb{X}\boldsymbol{\beta}) + 2 (\mathbf{Y} - \mathbb{X}\boldsymbol{\beta})^\top \hat{\boldsymbol{\Sigma}}^{-1} \mathbb{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) - \text{tr} \left(\boldsymbol{\Sigma} \hat{\boldsymbol{\Sigma}}^{-1} \right) \right].$$

Pripomenutím vzťahu (1.5) pre výpočet odhadu pevných efektov zistíme, že

$$\left(\mathbb{X}^\top \hat{\boldsymbol{\Sigma}}^{-1} \mathbb{X} \right)^{-1} \left(\mathbb{X}^\top \hat{\boldsymbol{\Sigma}}^{-1} \right) (\mathbf{Y} - \mathbb{X}\boldsymbol{\beta}) = \hat{\boldsymbol{\beta}} - \left(\mathbb{X}^\top \hat{\boldsymbol{\Sigma}}^{-1} \mathbb{X} \right)^{-1} \left(\mathbb{X}^\top \hat{\boldsymbol{\Sigma}}^{-1} \mathbb{X} \right) \boldsymbol{\beta} = \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}.$$

Pre prehľadnosť označme ďalej $\mathbf{u} = \mathbf{Y} - \mathbb{X}\boldsymbol{\beta}$ a z predošlej identity odvodíme, že vychýlenie je rovné

$$\mathbf{B} = \mathbb{E}_{\mathbf{Y}} \left[\mathbf{u}^\top \widehat{\boldsymbol{\Sigma}}^{-1} \mathbf{u} - 2 \mathbf{u}^\top \left\{ \widehat{\boldsymbol{\Sigma}}^{-1} \mathbb{X} \left(\mathbb{X}^\top \widehat{\boldsymbol{\Sigma}}^{-1} \mathbb{X} \right)^{-1} \mathbb{X}^\top \widehat{\boldsymbol{\Sigma}}^{-1} \right\} \mathbf{u} - \text{tr} \left(\boldsymbol{\Sigma} \widehat{\boldsymbol{\Sigma}}^{-1} \right) \right].$$

Vyjadrenie vychýlenia označíme ako rozdiel dvoch členov, ktoré budeme potom upravovať takto

$$\mathbf{B} = \mathbf{B}_1 - 2 \mathbf{B}_2, \quad (3.43)$$

kde

$$\begin{aligned} \mathbf{B}_1 &= \mathbb{E}_{\mathbf{Y}} \left[\mathbf{u}^\top \widehat{\boldsymbol{\Sigma}}^{-1} \mathbf{u} - \text{tr} \left(\boldsymbol{\Sigma} \widehat{\boldsymbol{\Sigma}}^{-1} \right) \right], \\ \mathbf{B}_2 &= \mathbb{E}_{\mathbf{Y}} \left[\mathbf{u}^\top \left\{ \widehat{\boldsymbol{\Sigma}}^{-1} \mathbb{X} \left(\mathbb{X}^\top \widehat{\boldsymbol{\Sigma}}^{-1} \mathbb{X} \right)^{-1} \mathbb{X}^\top \widehat{\boldsymbol{\Sigma}}^{-1} \right\} \mathbf{u} \right]. \end{aligned}$$

Pre výpočet \mathbf{B}_1 použijeme Taylorov rozvoj, podľa ktorého existuje nejaké $\boldsymbol{\tau}^\perp$ medzi $\widehat{\boldsymbol{\tau}}$ a $\boldsymbol{\tau}$ (teda $\|\boldsymbol{\tau}^\perp - \boldsymbol{\tau}\| \leq \|\widehat{\boldsymbol{\tau}} - \boldsymbol{\tau}\|$) také, že

$$\begin{aligned} \widehat{\boldsymbol{\Sigma}}^{-1} &= \boldsymbol{\Sigma}^{-1} + \sum_i \left(\boldsymbol{\Sigma}^{-1} \right)_{(i)} (\widehat{\tau}_i - \tau_i) + \frac{1}{2} \sum_{i,j} \left(\boldsymbol{\Sigma}^{-1} \right)_{(ij)} (\widehat{\tau}_i - \tau_i) (\widehat{\tau}_j - \tau_j) + \\ &\quad + \frac{1}{6} \sum_{i,j,k} \left(\boldsymbol{\Sigma}^{-1} \right)_{(ijk)}^{\boldsymbol{\tau}=\boldsymbol{\tau}^\perp} (\widehat{\tau}_i - \tau_i) (\widehat{\tau}_j - \tau_j) (\widehat{\tau}_k - \tau_k). \end{aligned} \quad (3.44)$$

Z vlastností operátora stopy vieme, že

$$\mathbf{B}_1 = \mathbb{E}_{\mathbf{Y}} \left[\text{tr} \left(\widehat{\boldsymbol{\Sigma}}^{-1} \mathbf{u} \mathbf{u}^\top \right) - \text{tr} \left(\widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Sigma} \right) \right] = \mathbb{E}_{\mathbf{Y}} \left[\text{tr} \left(\widehat{\boldsymbol{\Sigma}}^{-1} (\mathbf{u} \mathbf{u}^\top - \boldsymbol{\Sigma}) \right) \right]. \quad (3.45)$$

Podľa predpokladu (P4) je každý člen $\widehat{\tau}_i - \tau_i = O_p(n^{-1/2})$. Pre kvadratickú formu máme $\mathbf{u}^\top \left(\boldsymbol{\Sigma}^{-1} \right)_{(ijk)}^{\boldsymbol{\tau}=\boldsymbol{\tau}^\perp} \mathbf{u} = O_p(n)$. Ďalej vieme, že

$$\mathbb{E}_{\mathbf{Y}} \left[\text{tr} \left(\boldsymbol{\Sigma}^{-1} (\mathbf{u} \mathbf{u}^\top - \boldsymbol{\Sigma}) \right) \right] = \mathbb{E}_{\mathbf{Y}} \left[\mathbf{u}^\top \boldsymbol{\Sigma}^{-1} \mathbf{u} - \text{tr}(\mathbb{I}_n) \right] = 0,$$

lebo $\mathbf{u}^\top \boldsymbol{\Sigma} \mathbf{u}$ má χ_n^2 rozdelenie a stredná hodnota je teda rovná n . Preto použitím Taylorovho rozvoja (3.44) môžeme vyjadriť

$$\begin{aligned} \mathbf{B}_1 &= \sum_i \mathbb{E}_{\mathbf{Y}} \left[\text{tr} \left(\left(\boldsymbol{\Sigma}^{-1} \right)_{(i)} (\mathbf{u} \mathbf{u}^\top - \boldsymbol{\Sigma}) \right) (\widehat{\tau}_i - \tau_i) \right] + \\ &\quad + \frac{1}{2} \sum_{i,j} \mathbb{E}_{\mathbf{Y}} \left[\text{tr} \left(\left(\boldsymbol{\Sigma}^{-1} \right)_{(ij)} (\mathbf{u} \mathbf{u}^\top - \boldsymbol{\Sigma}) \right) (\widehat{\tau}_i - \tau_i) (\widehat{\tau}_j - \tau_j) \right] + O(n^{-1/2}). \end{aligned}$$

Kvôli konečnosti prvkov jednotlivých matíc podľa predpokladu (P2) a z vyjadrení v predpoklade (P4) môžeme \mathbf{B}_1 upraviť na tvar, kam dosadíme τ_i^* , resp. τ_i^{**}

$$\begin{aligned} \mathbf{B}_1 &= \sum_i \mathbb{E}_{\mathbf{Y}} \left[\text{tr} \left(\left(\boldsymbol{\Sigma}^{-1} \right)_{(i)} (\mathbf{u} \mathbf{u}^\top - \boldsymbol{\Sigma}) \right) (\widehat{\tau}_i^* + \widehat{\tau}_i^{**}) \right] + \\ &\quad + \frac{1}{2} \sum_{i,j} \mathbb{E}_{\mathbf{Y}} \left[\text{tr} \left(\left(\boldsymbol{\Sigma}^{-1} \right)_{(ij)} (\mathbf{u} \mathbf{u}^\top - \boldsymbol{\Sigma}) \right) \widehat{\tau}_i^* \widehat{\tau}_j^* \right] + O(n^{-1/2}). \end{aligned} \quad (3.46)$$

Počítajme členy z prvej sumy predošlej rovnosti (3.46)

$$\begin{aligned} \mathbf{E}_Y \left[\text{tr} \left((\boldsymbol{\Sigma}^{-1})_{(i)} (\mathbf{u}\mathbf{u}^\top - \boldsymbol{\Sigma}) \right) (\hat{\tau}_i^* + \hat{\tau}_i^{**}) \right] &= \mathbf{E}_Y \left[\mathbf{u}^\top (\boldsymbol{\Sigma}^{-1})_{(i)} \mathbf{u} (\hat{\tau}_i^* + \hat{\tau}_i^{**}) \right] - \\ &\quad - \text{tr} \left((\boldsymbol{\Sigma}^{-1})_{(i)} \boldsymbol{\Sigma} \right) \mathbf{E}_Y (\hat{\tau}_i^* + \hat{\tau}_i^{**}). \end{aligned} \quad (3.47)$$

Prvý člen na pravej strane predchádzajúcej rovnosti je podľa Lemmy 17 rovný

$$\text{tr} \left((\boldsymbol{\Sigma}^{-1})_{(i)} \boldsymbol{\Sigma} \right) \mathbf{E}_Y (\hat{\tau}_i^* + \hat{\tau}_i^{**}) + \mathbf{E}_Y \left[\text{tr} \left(\boldsymbol{\Sigma} (\boldsymbol{\Sigma}^{-1})_{(i)} \boldsymbol{\Sigma} \frac{\partial^2 (\hat{\tau}_i^* + \hat{\tau}_i^{**})}{\partial \mathbf{Y} \partial \mathbf{Y}^\top} \right) \right].$$

Keďže $\partial^2 \hat{\tau}_i^{**} / (\partial \mathbf{Y} \partial \mathbf{Y}^\top) = O_p(n^{-1})$ podľa predpokladu (P5) a podľa predpokladu (P2) sú pre dostatočne veľké n prvky matíc konečné, môžeme pravú stranu výrazu (3.47) písať v tvare

$$\mathbf{E}_Y \left[\text{tr} \left(\boldsymbol{\Sigma} (\boldsymbol{\Sigma}^{-1})_{(i)} \boldsymbol{\Sigma} \frac{\partial^2 \hat{\tau}_i^*}{\partial \mathbf{Y} \partial \mathbf{Y}^\top} \right) \right] + O(n^{-1/2}).$$

Rovnakým použitím Lemmy 17 sa ukáže, že pre členy pri dvojnej sume rovnosti (3.46) máme

$$\mathbf{E}_Y \left[\text{tr} \left((\boldsymbol{\Sigma}^{-1})_{(ij)} (\mathbf{u}\mathbf{u}^\top - \boldsymbol{\Sigma}) \right) \hat{\tau}_i^* \hat{\tau}_j^* \right] = \mathbf{E}_Y \left[\text{tr} \left(\boldsymbol{\Sigma} (\boldsymbol{\Sigma}^{-1})_{(ij)} \boldsymbol{\Sigma} \frac{\partial^2 \hat{\tau}_i^* \hat{\tau}_j^*}{\partial \mathbf{Y} \partial \mathbf{Y}^\top} \right) \right].$$

Člen na pravej strane tejto rovnosti je rádu $O(n^{-1})$ podľa predpokladu (P5), preto môžeme celkovo písať

$$B_1 = \sum_i \mathbf{E}_Y \left[\text{tr} \left(\boldsymbol{\Sigma} (\boldsymbol{\Sigma}^{-1})_{(i)} \boldsymbol{\Sigma} \frac{\partial^2 \hat{\tau}_i^*}{\partial \mathbf{Y} \partial \mathbf{Y}^\top} \right) \right] + O(n^{-1/2}).$$

Derivovaním súčinnu $\boldsymbol{\Sigma} \boldsymbol{\Sigma}^{-1}$ podľa τ_i odvodíme

$$\boldsymbol{\Sigma}_{(i)} \boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma} (\boldsymbol{\Sigma}^{-1})_{(i)} = \mathbf{0}_{n \times n}.$$

Preto dosadením identity $\boldsymbol{\Sigma} (\boldsymbol{\Sigma}^{-1})_{(i)} = -\boldsymbol{\Sigma}_{(i)} \boldsymbol{\Sigma}^{-1}$ do predošlej rovnosti máme

$$B_1 = - \sum_i \mathbf{E}_Y \left[\text{tr} \left(\boldsymbol{\Sigma}_{(i)} \frac{\partial^2 \hat{\tau}_i^*}{\partial \mathbf{Y} \partial \mathbf{Y}^\top} \right) \right] + O(n^{-1/2}). \quad (3.48)$$

Pre výpočet B_2 označme maticu $\mathbb{P} = \boldsymbol{\Sigma}^{-1} \mathbb{X} \left(\mathbb{X}^\top \boldsymbol{\Sigma}^{-1} \mathbb{X} \right)^{-1} \mathbb{X}^\top \boldsymbol{\Sigma}^{-1}$. Potom platí, že $B_2 = \mathbf{E}_Y \left[\mathbf{u}^\top \hat{\mathbb{P}} \mathbf{u} \right]$. Použijeme Taylorov rozvoj, podľa ktorého existuje $\boldsymbol{\tau}^+$ také, že $\|\boldsymbol{\tau}^+ - \boldsymbol{\tau}\| \leq \|\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}\|$ a platí

$$B_2 = \mathbf{E}_Y \left[\mathbf{u}^\top \mathbb{P} \mathbf{u} \right] + \sum_i \mathbf{E}_Y \left[\mathbf{u}^\top \mathbb{P}_{(i)}^{\boldsymbol{\tau} = \boldsymbol{\tau}^+} \mathbf{u} (\hat{\tau}_i - \tau_i) \right].$$

Z predpokladov (P1) a (P2) sa ukáže, že $\mathbf{u}^\top \mathbb{P}_{(i)}^{\boldsymbol{\tau} = \boldsymbol{\tau}^+} \mathbf{u}$ je ohraničené v pravdepodobnosti, teda rádu $O_p(1)$ a podľa (P4) je $\hat{\tau}_i - \tau_i = O_p(n^{-1/2})$. Preto vieme, že

$$B_2 = \mathbf{E}_Y \left[\mathbf{u}^\top \mathbb{P} \mathbf{u} \right] + O(n^{-1/2}).$$

Ďalej využijeme, že poznáme rozdelenie normálneho vektoru po transformácii

$$\mathbf{v} = \mathbb{X}^\top \boldsymbol{\Sigma}^{-1} \mathbf{u} \sim \mathcal{N}_p(\mathbf{0}, \mathbb{X}^\top \boldsymbol{\Sigma}^{-1} \mathbb{X}).$$

Z toho vyplýva, že

$$\mathbf{v}^\top (\mathbb{X}^\top \boldsymbol{\Sigma}^{-1} \mathbb{X})^{-1} \mathbf{v} = \mathbf{u}^\top \mathbb{P} \mathbf{u} \sim \chi_p^2.$$

Nakoniec preto dostávame, že

$$B_2 = p + O(n^{-1/2}). \quad (3.49)$$

Dosadením vyjadrení (3.48) a (3.49) do (3.43) zistíme, že $h_M(\boldsymbol{\tau})$ v znení vety vyplynie z výrazu B_1 . □

Ak je h diferencovateľná, potom sa dá podľa článku Kubokawa (2011) ukázať, že po dosadení odhadu platí

$$\mathbb{E}_Y h_M(\hat{\boldsymbol{\tau}}) = h_M(\boldsymbol{\tau}) + O(n^{-1/2}).$$

Preto je výraz $-2 \log L(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\psi}}, \hat{\sigma}_e^2) + 2p + 2h_M(\hat{\boldsymbol{\tau}})$ asymptoticky nestranným odhadom Akaikovej informácie LME modelu. V kapitole 4 používaného článku je ukázané, že pri použití maximálne vierohodného odhadu $\hat{\boldsymbol{\tau}} = (\hat{\sigma}_e^2, \hat{\boldsymbol{\psi}}^\top)^\top$ platí pre výraz $h_M(\hat{\boldsymbol{\tau}}) = Q + 1$, ak je splnený predpokald naviac, a síce

(P6) Označme maticu $\mathbb{A}_2 = (a_{ij})$ typu $(Q + 1) \times (Q + 1)$ takú, že

$$a_{ij} = \text{tr}(\boldsymbol{\Sigma}_{(i)} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{(j)} \boldsymbol{\Sigma}^{-1}).$$

Predpokladáme, že \mathbb{A}_2 je pozitívne definitná a taktiež \mathbb{A}_2/n konverguje k nejakej pozitívne definitnej matici.

Prvky matice \mathbb{A}_2 sú jedným z členov v Hessovej matici pre variančné parametre, ktorú sme počítali v Tvrdení 14. Kritérium mAIC definované vo výraze (3.16) je preto asymptoticky nestranný odhad Akaikovej informácie.

Aj pre podmienenú Akaikovu informáciu existuje odhad, ktorý je asymptoticky nestranný. Taktiež využíva Taylorov rozvoj a derivácie matice $\boldsymbol{\Sigma}$. Presný tvar sa dá nájsť v článku Kubokawa (2011), kde sa pracuje so všeobecnejším tvarom matice podmieneného rozdelenia a jej Taylorovým rozvojom.

4. Simulačná štúdia

Ďalšia kapitola sa venuje simulačnej štúdiu, pomocou ktorej sme zisťovali vlastnosti odvodených kritérií z predchádzajúcej kapitoly. Uvažované boli modely závislosti odozvy \mathbf{Y} na jednom regresore, pod ktorým si môžeme predstaviť čas, nakoľko sú LME modely častým nástrojom pri práci s longitudinálnymi dátami. Pracovali sme s odozvou N na sebe nezávislých subjektov (zhlukov), pričom na každom z nich bolo prevedených n_i pozorovaní odozvy v časoch t_{ij} pre $i = 1, \dots, N$ a $j = 1, \dots, n_i$.

Popíšeme postup generovania dátového súboru v jednom opakovaní simulácie. Zvolili sme šesť pozorovaní ako priemerný očakávaný počet pozorovaní pre jeden subjekt. Pre každý subjekt sa určilo prvé pozorovanie na začiatku, $t_{i1} = 0$. Časy ďalších pozorovaní sme zvolili približne v časoch 5, 10, 15, 20 a 25. Ku každej z predchádzajúcich hodnôt bolo pripočítané náhodné číslo z rovnomerného rozdelenia na intervale $(-1,1)$. Ak takto zvolený čas pozorovania prekročil hodnotu 25 (mohlo sa stať v poslednom prípade), posledné pozorovanie sa neuvažovalo. Máme tak simuláciu aj pre nevyvážené veľkosti zhlukov a dôvodom bol aj uvažovaný funkčný vzťah medzi odozvou a časom, ktorý je následne popísaný.

Vždy sme na vygenerovaných dátach odhadovali skupinu lineárnych zmiešaných modelov, ktoré majú štruktúru z Definície 1, pričom jeden z nich bol správny. Modely sa od seba líšili voľbou spomínaného funkčného vzťahu medzi odozvou a časom, presnejšie povedané líšili sa tvarom podmienenej strednej hodnoty odozvy pri daných hodnotách času. Použité prípady boli

$$\begin{aligned} \text{(a)} \quad m_a(t_{ij}) &= \mathbb{E}[Y_{ij}|t_{ij}] = \beta_0 + \beta_1 t_{ij}, \\ \text{(b)} \quad m_b(t_{ij}) &= \mathbb{E}[Y_{ij}|t_{ij}] = \beta_0 + \beta_1 \sin\left(\frac{2\pi}{25}t_{ij}\right). \end{aligned}$$

Períodu funkcie sínus sme zvolili tak, aby na uvažovanom intervale bol jeden priebeh sínusoidy a hodnoty pevných efektov boli vždy $\beta_0 = 3$ a $\beta_1 = 1$. Ďalej sme rozlišovali modely podľa variančnej štruktúry. Pre každý z prípadov (a) a (b) sa bralo do úvahy päť rôznych možností. Prvou je nezávislosť v rámci pozorovaní na subjekte, čo vytvára klasický lineárny model. Ďalšie dve zahŕňajú prípady, keď je náhodný buď absolútny člen alebo koeficient smernice. Posledné dve uvažujú oba náhodné členy, ktoré môžu byť buď na sebe nezávislé alebo nejak korelované. Spoločne to môžeme zhrnúť tak, že pre každý subjekt sme uvažovali vektor náhodných efektov $\mathbf{b}_i = (b_{0i}, b_{1i})^\top$, ktoré mali normálne rozdelenie s nulovou strednou hodnotou a variančnou maticou \mathbb{D} . Matica \mathbb{D} mala jeden z nasledujúcich tvarov

$$\mathbb{D}_1 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \mathbb{D}_2 = \begin{pmatrix} \tau_0^2 & 0 \\ 0 & 0 \end{pmatrix}, \mathbb{D}_3 = \begin{pmatrix} 0 & 0 \\ 0 & \tau_1^2 \end{pmatrix}, \mathbb{D}_4 = \begin{pmatrix} \tau_0^2 & 0 \\ 0 & \tau_1^2 \end{pmatrix}, \mathbb{D}_5 = \begin{pmatrix} \tau_0^2 & \tau_{01} \\ \tau_{01} & \tau_1^2 \end{pmatrix}$$

a parametre $\tau_0^2, \tau_1^2 \in (0, \infty)$ a $\tau_{01} \neq 0$. Podmienené rozdelenie $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^\top$ pri znalosti hodnôt náhodných efektov bolo následne uvažované ako normálne so strednou hodnotou μ_a , resp. μ_b podľa prípadu (a), resp. (b) a s variančnou maticou $\sigma_e^2 \mathbb{I}_{n_i}$. Parameter σ_e^2 nazývame reziduálny rozptyl. Konkrétne tvary podmienených stredných hodnôt hodnotami času a náhodných efektov sú

$$\begin{aligned} \text{(a)} \quad \mu_a &= \mathbb{E}[Y_{ij}|t_{ij}, \mathbf{b}_i] = m_a(t_{ij}) + b_{0i} + b_{1i} t_{ij}, \\ \text{(b)} \quad \mu_b &= \mathbb{E}[Y_{ij}|t_{ij}, \mathbf{b}_i] = m_b(t_{ij}) + b_{0i} + b_{1i} \sin\left(\frac{2\pi}{25}t_{ij}\right). \end{aligned}$$

Celkovo sme teda uvažovali desať možných modelov. Budeme ich označovať postupne v poradí číslami 1 až 5 pre prípad lineárnej závislosti (a) a matíc \mathbb{D} s indexmi 1 až 5. Modely 6 až 10 budú mať sínusovú závislosť na čase (b) a matice \mathbb{D} postupne s indexmi 1 až 5.

Následne sme previedli dve štúdie, prvá z nich uvažovala vždy jeden z desiatich správnych modelov a sledovala, aké úspešné sú kritériá pri označovaní správneho modelu. Druhá sa venovala modelom s dvoma korelovanými náhodnými efektmi a sledovala vplyv korelácie na podiel označených správnych modelov pomocou kritérií. Obe simulácie sme prevádzali vo výpočtovom prostredí R, R Core Team (2016), kde sme používali funkciu `lme` z balíčka `nlme` Pinheiro a kol. (2016) pre odhadovanie parametrov uvažovaných LME modelov.

4.1 Štúdia č. 1

Najprv sme prevádzali simuláciu pre sériu modelov s hodnotami rozptylov náhodných efektov $\tau_0^2 = 0,5$ a $\tau_1^2 = 0,2$. V poslednej piatej možnej matici sme navyše zvolili $\tau_{01} = -0,15$, čo odpovedá korelácii $-0,47$ medzi náhodnými efektmi. Zvolený bol jednotkový reziduálny rozptyl. Zhrnieme, ako vyzerali skutočné hodnoty matíc \mathbb{D} pri simulácii

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0,5 & 0 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 \\ 0 & 0,2 \end{pmatrix}, \quad \begin{pmatrix} 0,5 & 0 \\ 0 & 0,2 \end{pmatrix}, \quad \begin{pmatrix} 0,50 & -0,15 \\ -0,15 & 0,20 \end{pmatrix}.$$

Pre každú kombináciu jednej z matíc a prípadu (a), resp. (b) bolo vygenerovaných 1000 pozorovaní odozvy pre tri rôzne počty zhlukov $N \in \{20, 50, 75\}$. Na týchto dátach sa následne odhadli všetky uvažované modely spomínanou funkciou `lme`. Pre každý z nich boli určené hodnoty Akaikových kritérií `mAIC` a `cAIC`. Prvé z nich pomocou vzťahu (3.16) a druhé zo vzťahu (3.18), kde je ρ odhadnuté pri predpokladanej znalosti reziduálneho rozptylu podľa Vety 13 a Ψ_0 sme počítali pomocou Tvrdenia 14. Nakoniec sme pre každý dátový súbor vybrali model s minimálnou hodnotou kritéria a pozreli sme sa na percentuálne zastúpenie výberu jednotlivých modelov pri 1000 simuláciách. Predovšetkým nás zaujímalo, ako často je týmto vybraný model správny, prípadne aké modely jednotlivé kritériá uprednostňujú.

V Tabulke 4.1 vidíme percentuálne podiely výberu správneho modelu. Pre marginálne Akaikovo kritérium sa zdá, že v prípadoch, keď nie je žiaden náhodný efekt alebo je len jeden, správny model je rozoznaný vo väčšine prípadov už aj pri menšom rozsahu výberu (stĺpce 1, 2, 3, 6, 7 a 8). Najčastejšie bol pre tieto modely vybraný správny model č. 3 s lineárnou závislosťou a jediným náhodným efektom, ktorým bol čas. Pri dvadsiatich zhlukoch bol tento model správne označený až v 946 prípadoch. Podmienené kritérium je na tom v týchto modeloch percentuálne horšie, no platí preňho, že sa takmer vždy percento vybratia správneho modelu zvyšovalo s rastúcim počtom zhlukov. Na druhej strane sa `cAIC` pri výbere modelov s dvomi náhodnými členmi (stĺpce 4, 5, 9 a 10) marginálnemu kritériu prinajmenšom vyrovnáva, dokonca vo väčšine prípadov (až na 10 stĺpec) má vyššie percentuálne zastúpenie výberu správneho modelu. Vidíme, že v prípade, keď je správny model č. 10, ani jedno z kritérií nedáva dobré výsledky. Ide o najzložitejšiu uvažovanú štruktúru náhodných efektov so sínusovou závislosťou.

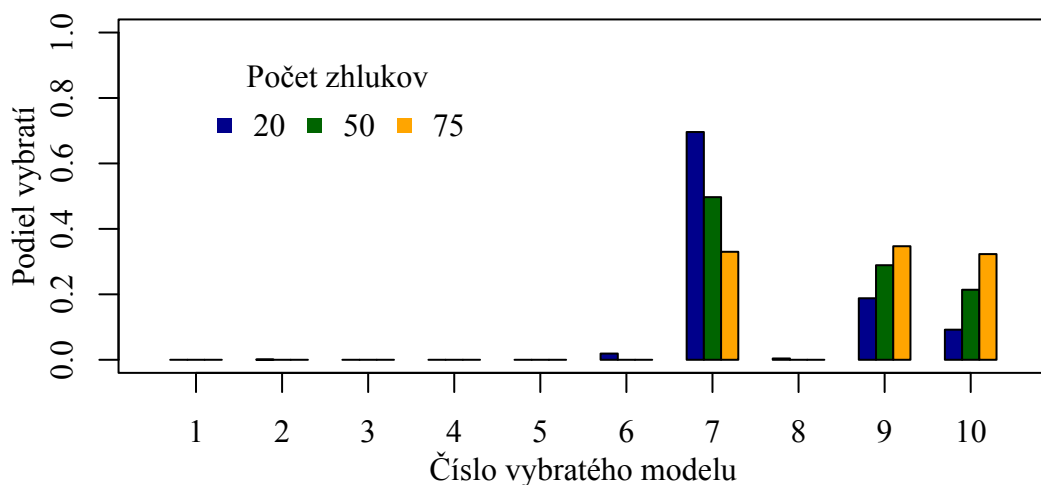
	1	2	3	4	5	6	7	8	9	10
N	Percentuálne zastúpenie výberu správneho modelu pomocou mAIC									
20	89,2	89,9	94,6	43,7	25,8	87,2	91,4	31,5	26,8	9,2
50	90,1	90,7	92,5	76,5	47,2	88,1	92,9	57,3	48,6	21,4
75	89,4	91,1	91,4	82,6	64,9	87,4	92,4	73,2	62,4	32,3
N	Percentuálne zastúpenie výberu správneho modelu pomocou cAIC									
20	47,2	77,2	84,2	63,5	32,9	43,0	73,0	40,5	44,6	6,2
50	53,5	85,4	84,1	81,2	63,1	50,1	85,4	65,8	65,6	19,4
75	52,2	84,6	81,3	84,0	76,5	50,7	85,7	79,6	75,1	31,1

Tabuľka 4.1: Prehľad percentuálneho zastúpenia výberu správneho modelu pre obe kritéria. Stĺpce označujú číslo správneho modelu, prvých päť má marginálnu závislosť lineárnu a ďalších päť sínusovú.

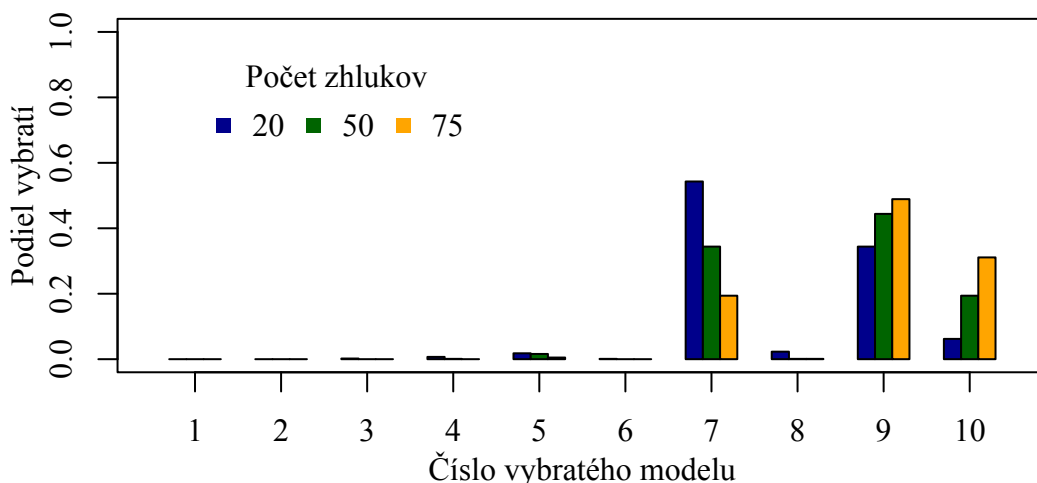
V prílohe A.3 sú uvedené obrázky pre každé nastavenie parametrov a obe kritéria zvlášť. Na tomto mieste vyberieme obrázky venujúce sa práve poslednému modelu, ktorý sa zdal byť problémový.

Na Obrázkoch 4.1 a 4.2 môžeme porovnať, ktoré modely boli vyberané najčastejšie, ak sme generovali z modelu so sínusovou závislosťou a korelovanými náhodnými efektami. Obe kritéria pre najmenší počet zhlukov ($N = 20$) uprednostňovali model č. 7 s náhodným iba absolútnym členom a pre najvyšší počet zhlukov ($N = 75$) bol preferovaný model č. 9 s nezávislými náhodnými efektami, aj keď pre mAIC je výsledok porovnateľný s vyberaním správneho modelu č. 10. Pre 50 zhlukov sa kritéria líšili, mAIC uprednostňovalo model č. 7 s jedným náhodným efektom a cAIC s dvoma nezávislými. Podiel vybratí správneho modelu pre obe kritériá s rastúcim počtom zhlukov narastal.

Celkovo sa podľa obrázkov v prílohe A.3 zdá, že marginálne kritérium dobre rozpoznáva štruktúru pevných efektov, no následne vyberá modely s menším počtom efektov náhodných, resp. lineárny model. Naopak, podmienené kritérium



Obr. 4.1: Zastúpenie vybratí jednotlivých modelov kritériom mAIC v prípade, že je správny model č.10



Obr. 4.2: Zastúpenie vybratí jednotlivých modelov kritériom cAIC v prípade, že je správny model č.10

zase častejšie vyberá modely, kde je variančná štruktúra zložitejšia a majú aspoň jeden náhodný člen. Vidíme, že s rastúcim počtom pozorovaní je marginálne kritérium aj v prípadoch s náhodnými efektami lepšie, čo môže odpovedať jeho asymptotickým vlastnostiam. Rovnako to je s podmieneným kritériom vzhľadom k počtu pozorovaných subjektov.

4.2 Štúdia č. 2

Na základe výsledkov sme sa rozhodli bližšie preskúmať, ako sa správa výber správneho modelu v prípadoch s dvoma korelovanými náhodnými efektmi. Zaujímalo nás, či súvisí problém v simulácii s desiatym modelom predošlej časti s hodnotou kovariancie, resp. korelácie medzi náhodnými efektami.

Zvolili sme preto k matici z predošlej časti odpovedajúcej modelom s najzložitejšou štruktúrou náhodných efektov ďalších päť. Použité boli tri hodnoty kovariancie, z nich každá s kladným a záporným znamienkom. Konkrétne sme mali matice

$$\begin{pmatrix} 0,50 & \pm 0,05 \\ \pm 0,05 & 0,20 \end{pmatrix}, \quad \begin{pmatrix} 0,50 & \pm 0,15 \\ \pm 0,15 & 0,20 \end{pmatrix}, \quad \begin{pmatrix} 0,50 & \pm 0,30 \\ \pm 0,30 & 0,20 \end{pmatrix}.$$

Odpovedajú postupne hodnotám korelácie 0,16; 0,47 a 0,95. Budeme ich označovať postupne číslami 1, 2 a 3. Okrem prípadu so sínusovou závislosťou sme taktiež simulovali lineárnu závislosť, aby sa dali tieto prípady porovnať. Celkovo tak máme pre každý počet zhlukov, zvolený rovnako ako v prvej časti (20, 50 alebo 70), jednu zo šiestich matíc a dve možné závislosti, čo spolu dáva 36 rôznych nastavení. Pri každom nastavení bolo generovaných 1 000 dátových súborov a odhadnuté rovnaké LME modely ako v prvej časti. Potom sme sledovali, koľkokrát sú vyberané jednotlivé modely pomocou rovnakých kritérií ako v prvej simulácii.

V Tabuľke 4.2 vidíme, že pre model so sínusovou závislosťou pretrvávajú problémy pri výbere správneho modelu. Pri vyššej hodnote korelácie medzi oboma hlavnými efektmi je obomi kritériami vybratý správny model menej než v 10 %

lineárna závislosť - správny model 5			sínusová závislosť - správny model 10																																			
kladná korelácia			záporná korelácia			kladná korelácia			záporná korelácia																													
1	2	3	1	2	3	1	2	3	1	2	3																											
N												Percentuálne zastúpenie výberu správneho modelu pomocou mAIC																										
20	16,0	26,2	35,9	17,7	25,8	45,9	8,4	8,7	2,9	7,7	9,2	2,9	50	19,4	50,2	80,8	19,7	47,2	86,9	13,3	21,9	4,9	13,2	21,4	6,1	75	22,2	60,7	93,6	21,3	64,9	95,8	18,7	30,4	7,4	16,8	32,3	8,4
N												Percentuálne zastúpenie výberu správneho modelu pomocou cAIC																										
20	10,6	9,3	7,8	17,5	32,9	64,3	5,3	6,8	2,3	4,6	6,2	2,8	50	10,4	25,4	45,0	26,6	63,1	94,2	12,3	21,2	4,7	11,5	19,4	6,0	75	11,9	36,4	67,1	33,8	76,5	98,4	17,4	29,9	8,2	15,3	31,1	8,2

Tabuľka 4.2: Prehľad percentuálneho zastúpenia výberu správneho modelu pre model s korelovanými náhodnými efektami, pre tri rôzne veľkosti korelácie s kladnou a zápornou hodnotou. Hodnoty absolútnej hodnoty korelácie sú postupne 0,16; 0,47 a 0,95 označené v stĺpcoch číslami 1, 2 a 3.

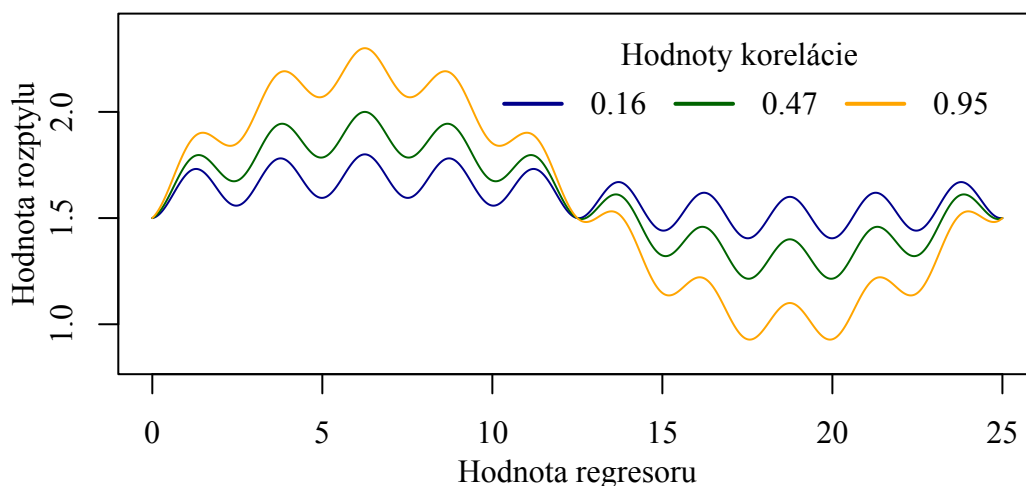
prípadoch pre kladnú aj zápornú hodnotu korelácie a rovnako pri všetkých počtoch zhlukov. Okrem toho ani jedno z kritérií pri všetkých uvažovaných hodnotách korelácií nepresahuje podielom označenia správneho modelu hodnotu 0,35. Oproti tomu je model s lineárnou závislosťou oveľa viackrát správne vybraný pomocou oboch kritérií. V každom prípade sa prejavujú asymptotické vlastnosti a počet správnych označení modelu č. 5 je rastúci s počtom zhlukov. V tomto prípade je pri vyšších hodnotách absolútnej hodnoty korelácie medzi náhodnými efektmi častejšie identifikovaný správny model. Pri 75 zhlukoch a korelácií $-0,95$ označilo podmienené kritérium dokonca 98,4 % správne.

Na základe Tabuľky 4.2 usudzujeme, že efektívita označovania správneho modelu pri modeloch s dvomi korelovanými náhodnými efektami je ovplyvnená hodnotou korelácie. Pre lineárnu závislosť rastie percento označovania správneho modelu so zvyšujúcou sa absolútnou hodnotou korelácie. Pre sínusovú závislosť naopak percento klesá pri vyšších hodnotách korelácie. Avšak problém, ktorý bol v prvej časti, netkvie vo veľkosti korelácie, pretože sme pozorovali rovnako nízke podiely výberu správneho modelu č. 10.

Problémom pravdepodobne bude, ako nami zvolený správny model určuje rozptyl jednotlivých zložiek odozvy, prípadne korelácie medzi jednotlivými pozorovaniami. Pozrieme sa na marginálny rozptyl. Podľa Tvrdenia 1 platí, že $\text{var } Y_{ij} = \sigma_e^2 + \mathbf{z}_{ij}^\top \mathbb{D} \mathbf{z}_{ij}$, kde Y_{ij} je zložka odozvy, \mathbb{D} je uvažovaná variančná matica náhodných efektov, \mathbf{z}_{ij} je vektor hodnôt regresorov náhodnej časti pre pozorovanie j v zhluke i a σ_e^2 hodnota reziduálneho rozptylu. V modeli č. 10 uvažujeme okrem absolútneho členu aj hodnotu sínusu použitého regresoru t_{ij} , preto je konkrétne pri zvolených hodnotách parametrov

$$\text{var } Y_{ij} = 1 + 0,5 + 2 \tau_{01} \sin\left(\frac{2\pi}{25} t_{ij}\right) + 0,2 \sin^2\left(\frac{2\pi}{25} t_{ij}\right),$$

kde τ_{01} sú hodnoty uvažovaných kovariancií. Na Obrázku 4.3 vidíme, ako sa mení rozptyl v našich prípadoch na uvažovanom intervale od 0 do 25, kde ležia hodnoty regresoru. Vidíme, že rozptyl ako funkcia hodnôt regresoru celkom výrazne kmitá.



Obr. 4.3: Závislosť rozptylu jednotlivých zložiek odozvy, ak sme uvažovali správny modelu č. 10 pre tri rôzne hodnoty korelácií náhodných efektov použité v druhej časti simulácií.

Čím je vyššia hodnota korelácie, tým kmitá viac, preto sa percento odhalenia správneho modelu znižovalo s rastúcou kovarianciou. Dôsledkom môže byť horšie rozpoznávanie správneho modelu pri simulácii dát z modelu č. 10. V prípade modelu č. 5, kde stredná hodnota závisí lineárne, je rozptyl kvadratickou funkciou hodnôt regresoru. Z toho dôvodu sa nevyskytli ani pri korelovaných náhodných efektoch výraznejšie problémy s niektorým z kritérií.

4.3 Zhrnutie výsledkov

Simulačná štúdia skúmala, ako vhodné sú marginálne a podmienené Akaikovo kritérium pri rozoznávaní správnej variančnej štruktúry náhodných efektov medzi modelmi s absolútnym členom a hodnotami jedného regresoru. Taktiež sme porovnávali, ako sú schopné rozlíšiť štruktúru pevných efektov medzi lineárnou a sínusovou závislosťou.

Ukázalo sa, že obe kritériá v každom prípade odhalili správnu štruktúru pevných efektov. Marginálne kritérium má tendenciu vyberať modely s menším počtom náhodných efektov, preto dosahovalo lepšie výsledky než podmienené pri modeloch s najviac jedným náhodným efektom. Naopak podmienené Akaikovo kritérium dosahovalo lepšie výsledky pri modeloch s dvomi náhodnými efektmi, z ktorých jeden bol absolútny člen.

Pre modely s dvomi náhodnými efektmi sa však ukázalo, že je podstatné, akú variančnú štruktúru predpokladá zahrnutý funkčný vzťah hodnôt regresoru. Ak skutočný rozptyl zložiek odozvy závisel na hodnotách regresoru kmitavou krivkou, ktorá sa v druhej polovici svojimi hodnotami pohybovala okolo hodnôt reziduálneho rozptylu, obe kritériá mali problém s rozpoznaním správneho modelu. Vyberali modely len s jedným náhodným efektom vo väčšine prípadov.

Pri menej kolísavej závislosti rozptylu, v našom prípade napríklad pri kvadratickej, obe kritériá označili viac než 40 % modelov správne a prejavili sa asymptotické vlastnosti kritérií. Pri vyššom počte zhlukov, resp. subjektov bolo vyššie percento označenia správneho modelu.

Záver

Práca sa venovala ohodnocovaniu predikčných vlastností lineárnych zmiešaných modelov. Tento proces sa používa pri výbere vhodného modelu na základe pozorovaní napríklad skupinovo závislých alebo longitudinálnych dát, ak požadujeme, aby vybraný model dobre predikoval hodnotu odozvy (závislej veličiny) na základe pozorovaní regresorov (nezávislých veličín).

Po definícii lineárneho zmiešaného modelu a pripomenutí jeho vlastností a dôsledkov, ktoré vyplývajú pre variančnú štruktúru pozorovaní odozvy, sa práca venovala kvantifikácii chyby predikcie. Najprv bol vysvetlený rozdiel medzi marginálnym a podmieneným prístupom predikcie, ktorý sa dá zovšeobecniť aj na iné hierarchicky usporiadané modely. Ďalej sme sa venovali výpočtu strednej štvorcovej chyby predikcie (MSEP) v lineárnom zmiešanom modeli, kde sa ukázala potreba odhadovať parametre modelu. Preto sme sa venovali spôsobu, ako sa všeobecné postupy, penalizácia kovarianciou a krosvalidácia, navrhnuté v článku Efron (2004) dajú použiť pri výpočte MSEP v lineárnom zmiešanom modeli.

Následne sa v práci sústredíme na použitie Akaikovho informačného kritéria pri výbere modelov. Spomenuté kritérium bolo pôvodne navrhnuté pre prácu s dátami, ktoré sú na sebe nezávislé. Lineárny zmiešaný model tento predpoklad porušuje, preto bolo potrebné zisťovať vlastnosti kritéria v lineárnom zmiešanom modeli, prípadne ho upraviť. Odvodenie marginálneho Akaikovho kritéria kopíruje pôvodný postup odvodenia Akaikovho kritéria. Tento prístup využíva vierohodnosť celého vektoru odozvy. Vychádzajúc z článku Kubokawa (2011) je v práci ukázané, za akých predpokladov je kritérium asymptoticky neustranným odhadom Akaikovej informácie.

Okrem toho sme sa venovali podmienenej Akaikovej informácii (cAI), ktorá pracuje s vierohodnosťou zhlukov pri znalosti náhodných efektov lineárneho zmiešaného modelu. Využitím článkov Vaida a Blanchard (2005) a Greven a Kneib (2010) sme ukázali, ako sa počíta podmienené Akaikovo kritérium. Taktiež sme sa venovali tomu, ako vyzerá korekčný člen podmienenej vierohodnosti, ktorý ju upravuje na neustranný odhad cAI. Rozlíšili sme niekoľko prípadov v závislosti na znalosti alebo potrebe odhadovania variančných parametrov modelu.

Nakoniec sme pomocou simulácií zisťovali, ako sa správajú odvodené marginálne a podmienené Akaikovo kritérium pri výbere správneho modelu. Ukázali sa asymptotické vlastnosti oboch kritérií, a síce pri vyššom počte zhlukov lineárneho zmiešaného modelu odhalili kritériá správny model častejšie než pri menšom počte zhlukov. Hlavný rozdiel medzi kritériami bol v tom, že marginálne kritérium sa lepšie správalo pri modeloch s menším počtom náhodných efektov, kým podmienené dosahovalo lepšie výsledky pre zložitejšie variančné štruktúry. Taktiež sa preukázal nedostatok oboch kritérií v prípade, keď bol uvažovaný marginálny rozptyl odozvy výrazne kmitajúcou funkciou hodnôt použitého regresoru.

Námetom pre ďalšie štúdium je odvodiť tvary kritérií v prípadoch, keď potrebujeme pracovať s odozvou, o ktorej sa nedá predpokladať, že je normálne rozdelená, prípadne je diskretná. Preto by bolo vhodné podobným prístupom študovať zovšeobecnený lineárny zmiešaný model, prípadne pomocou simulácií skúmať už dostupné odvodené kritéria pre konkrétne tvary rozdelení odozvy.

Zoznam použitej literatúry

- AKAIKE, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **AC-19**(6), 716–723.
- EFRON, B. (2004). The estimation of prediction error: Covariance penalties and cross-validation [with comments, rejoinder]. *Journal of the American Statistical Association*, **99**(467), 619–642.
- GREVEN, S. a KNEIB, T. (2010). On the behaviour of marginal and conditional AIC in linear mixed models. *Biometrika*, **97**(4), 773–789.
- HENDERSON, C. R. (1984). *Applications of Linear Models in Animal Breeding*. University of Guelph, 3rd edition. URL <http://cgil.uoguelph.ca/pub/Henderson>.
- KUBOKAWA, T. (2011). Conditional and unconditional methods for selecting variables in linear mixed models. *Journal of Multivariate Analysis*, **102**(3), 641–660.
- LAIRD, N. M. a WARE, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, **38**(4), 963–974.
- LIANG, H., WU, H. a ZOU, G. (2008). A note on conditional AIC for linear mixed-effects models. *Biometrika*, **95**(3), 773–778.
- PATTERSON, H. D. a THOMPSON, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, **58**(3), 545–554.
- PINHEIRO, J., BATES, D., DEBROY, S., SARKAR, D. a R CORE TEAM (2016). *nlme: Linear and Nonlinear Mixed Effects Models*. URL <http://CRAN.R-project.org/package=nlme>. R package version 3.1-128.
- R CORE TEAM (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- STEIN, C. (1973). Estimation of the mean of a multivariate normal distribution. In HÁJEK, J., editor, *Proceedings of the Prague Symposium on Asymptotic Statistics*, volume 2, pages 345–381. Charles University, Prague.
- STEIN, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, **9**(6), 1135–1151.
- VAIDA, F. a BLANCHARD, S. (2005). Conditional Akaike information for mixed-effects models. *Biometrika*, **92**(2), 351–370.

A. Prílohy

A.1 Prehľad teórie maximálnej vierohodnosti

Predpokladajme, že máme náhodný výber vektorov $\mathbf{X}_1, \dots, \mathbf{X}_n$ rozdelených rovnako ako generický vektor $\mathbf{X} = (X_1, \dots, X_k)^\top$, ktorý má hustotu f_θ vzhľadom ku nejakej σ -konečnej miere μ . Predpokladajme, že poznáme jej tvar až na hodnotu parametra $\theta = (\theta_1, \dots, \theta_p)^\top \in \Theta$. Jeho skutočnú hodnotu označme θ_0 .

Ďalej definujeme vierohodnosť

$$L_n(\theta) = \prod_{i=1}^n f_\theta(\mathbf{X}_i).$$

Maximálne vierohodný odhad parametra θ_0 hľadáme ako

$$\hat{\theta}_n = \arg \max_{\theta \in \Theta} L_n(\theta). \quad (\text{A.1})$$

Na odvodenie vlastností tohto odhadu sa štandardne zavádzajú tzv. podmienky regularity, ktoré zhrnieme v nasledujúcich bodoch:

- (R0) Pre ľubovoľné dva parametre $\theta_1, \theta_2 \in \Theta$ platí, že $f_{\theta_1}(\mathbf{x}) = f_{\theta_2}(\mathbf{x})$ pre skoro všetky \mathbf{x} vzhľadom ku miere μ práve vtedy, keď $\theta_1 = \theta_2$.
- (R1) Počet parametrov modelu je konštantný, rovný p .
- (R2) Nosič $S = \{\mathbf{x} \in \mathbb{R}^k; f_\theta(\mathbf{x}) > 0\}$ nezávisí na hodnote parametra θ .
- (R3) Parameter θ_0 (skutočná hodnota) je vnútorným bodom parametrického priestoru Θ .
- (R4) Hustota $f_\theta(\mathbf{x})$ je trikrát diferencovateľná podľa parametra θ na nejakom otvorenom okolí U bodu θ_0 . Navyiac pre každé $j, k, l \in \{1, \dots, p\}$ existuje (integrovateľná) funkcia $M_{jkl}(\mathbf{x})$ taká, že

$$\sup_{\theta \in U} \left| \frac{\partial^3 \log f_\theta(\mathbf{x})}{\partial \theta_j \partial \theta_k \partial \theta_l} \right| \leq M_{jkl}(\mathbf{x})$$

pre skoro všetky \mathbf{x} vzhľadom ku μ a $\mathbf{E}_{\theta_0} M_{jkl} < \infty$.

- (R5) Fisherova informačná matica

$$I(\theta) = \mathbf{E}_\theta \left[\frac{\partial \log f_\theta(\mathbf{X})}{\partial \theta} \frac{\partial \log f_\theta(\mathbf{X})}{\partial \theta^\top} \right]$$

je konečná, regulárna a pozitívne definitná pre θ_0 .

- (R6) Poradie derivácie a integrálu môžeme zameniť vo výrazoch typu

$$\frac{\partial}{\partial \theta} \int h_\theta(\mathbf{x}) d\mu(\mathbf{x}) = \int \frac{\partial}{\partial \theta} h_\theta(\mathbf{x}) d\mu(\mathbf{x}),$$

kde $h_\theta(\mathbf{x})$ je rovné buď $f_\theta(\mathbf{x})$ alebo $\partial f_\theta(\mathbf{x}) / \partial \theta$.

Poznámka. Podmienka (R0) hovorí iba o identifikácii parametra. Podmienka (R4) by mohla byť slabšia, avšak nie je naším cieľom ju zjednodušať.

Označme skórovú funkciu $\mathbf{U}_\theta(\mathbf{X}_i)$ pozorovania i , skórovú štatistiku $\mathbf{U}_n(\boldsymbol{\theta})$ a pozorovanú informačnú maticu $I_n(\boldsymbol{\theta})$.

$$\mathbf{U}_\theta(\mathbf{X}_i) = \frac{\partial \log f_\theta(\mathbf{X}_i)}{\partial \boldsymbol{\theta}}, \quad \mathbf{U}_n(\boldsymbol{\theta}) = \sum_{i=1}^n \mathbf{U}_\theta(\mathbf{X}_i), \quad I_n(\boldsymbol{\theta}) = -\frac{1}{n} \frac{\partial \mathbf{U}_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}.$$

Maximálne vierohodný odhad môžeme hľadať následne riešením sústavy odhadovacích rovníc

$$\mathbf{U}_n(\hat{\boldsymbol{\theta}}_n) \stackrel{!}{=} \mathbf{0}. \quad (\text{A.2})$$

Nasledujúca veta zhrňuje základné výsledky o asymptotických vlastnostiach maximálne vierohodného odhadu.

Veta 19. *Ak sú splnené podmienky regularity (R0)–(R6), potom*

(i) *S pravdepodobnosťou blížiacou sa k jednej pre $n \rightarrow \infty$ existuje konzistentné riešenie $\hat{\boldsymbol{\theta}}_n$ odhadovacích rovníc (A.2) také, že*

$$\sqrt{n} (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = I^{-1}(\boldsymbol{\theta}_0) \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{U}_{\boldsymbol{\theta}_0}(\mathbf{X}_i) + o_P(1),$$

z čoho ďalej máme

$$\sqrt{n} (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}_p(\mathbf{0}_p, I^{-1}(\boldsymbol{\theta}_0)).$$

(ii) *Pre konzistentný odhad \hat{I}_n matice $I(\boldsymbol{\theta}_0)$ platí*

$$\begin{aligned} W_n &= n (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top \hat{I}_n (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow[n \rightarrow \infty]{d} \chi_p^2, \\ R_n &= \frac{1}{n} \mathbf{U}_n^\top(\boldsymbol{\theta}_0) \hat{I}_n^{-1} \mathbf{U}_n(\boldsymbol{\theta}_0) \xrightarrow[n \rightarrow \infty]{d} \chi_p^2. \end{aligned}$$

(iii) *Platí*

$$LR_n = 2 \log \left(\frac{L_n(\hat{\boldsymbol{\theta}}_n)}{L_n(\boldsymbol{\theta}_0)} \right) \xrightarrow[n \rightarrow \infty]{d} \chi_p^2.$$

Poznámka. Možné konzistentné odhady, ktoré potrebujeme v bode (ii) predošlej vety sú

$$I(\hat{\boldsymbol{\theta}}_n), \quad I_n(\hat{\boldsymbol{\theta}}_n) \quad \text{alebo} \quad \frac{1}{n} \sum_{i=1}^n \mathbf{U}_{\hat{\boldsymbol{\theta}}_n}(\mathbf{X}_i) \mathbf{U}_{\hat{\boldsymbol{\theta}}_n}^\top(\mathbf{X}_i),$$

pričom na konzistenciu posledného z nich potrebujeme, aby boli všetky parciálne derivácie druhého rádu logaritmu hustoty f_θ ohraničené funkciou $M_{jk}(\mathbf{x})$ na nejakom okolí bodu $\boldsymbol{\theta}_0$ pre skoro všetky \mathbf{x} vzhľadom ku miere μ a $\mathbf{E}_{\boldsymbol{\theta}_0} M_{jk}^2(\mathbf{X}_1) < \infty$.

A.2 Dôkaz Lemmy 12

Lemma 12. Nech Y je náhodná veličina s normálnym rozdelením $\mathcal{N}(\mu, \sigma^2)$ a g je reálna diferencovateľná funkcia taká, že $\mathbf{E} |g'(Y)| < \infty$. Označme

$$h(y) = g\left(\frac{y - \mu}{\sigma}\right).$$

Potom platí

$$\mathbf{E} h'(Y) = \mathbf{E} \left[\frac{Y - \mu}{\sigma^2} h(Y) \right].$$

Dôkaz. Vzťah sa najprv ukáže pre $X \sim \mathcal{N}(0,1)$. Chceme teda ukázať rovnosť

$$\mathbf{E} g'(X) = \mathbf{E} [Xg(X)]. \quad (\text{A.3})$$

Vyjdeme z ľavej strany a budeme počítať integrál. Označme $\phi(x)$ hustotu štandardného normálneho rozdelenia.

$$\mathbf{E} g'(X) = \int_{-\infty}^{\infty} g'(x)\phi(x)dx \quad (\text{A.4})$$

Ďalej použijeme vzťah

$$\phi'(x) = -x\phi(x),$$

z čoho riešením diferenciálnej rovnice platí

$$\phi(x) = \begin{cases} \int_x^{\infty} x\phi(x)dx, & \text{pre } x > 0, \\ -\int_{-\infty}^x x\phi(x)dx, & \text{pre } x < 0. \end{cases} \quad (\text{A.5})$$

Dosadením (A.5) do (A.4) dostaneme

$$\mathbf{E} g'(X) = \int_0^{\infty} g'(x) \left(\int_x^{\infty} y\phi(y)dy \right) dx - \int_{-\infty}^0 g'(x) \left(\int_{-\infty}^x y\phi(y)dy \right) dx.$$

Ďalej použitím Fubiniovej vety na oba integrály na pravej strane dostaneme

$$\mathbf{E} g'(X) = \int_0^{\infty} y\phi(y) \left(\int_0^y g'(x)dx \right) dy - \int_{-\infty}^0 y\phi(y) \left(\int_{-\infty}^x g'(x)dx \right) dy.$$

Zintegrovaním derivácie $g'(x)$ upravíme rovnosť do tvaru

$$\mathbf{E} g'(X) = \int_0^{\infty} y\phi(y)(g(y) - g(0))dy - \int_{-\infty}^0 y\phi(y)(g(0) - g(y))dy.$$

Vyňatím -1 z druhého integrálu upravíme na

$$\mathbf{E} g'(X) = \int_{-\infty}^{\infty} y\phi(y)(g(y) - g(0))dy = \mathbf{E} [Xg(X)] - g(0)\mathbf{E}X.$$

V tomto prípade je $\mathbf{E}X = 0$, lebo dokazujeme (A.3), ktorá je špeciálnym prípadom pre štandardné normálne rozdelenie. Pomocou X sa normálne rozdelenie so strednou hodnotou μ a rozptylom σ^2 získa transformáciou $Y = \mu + \sigma X$. Deriváciou funkcie h odvodíme

$$\mathbf{E} h'(Y) = \frac{1}{\sigma} \mathbf{E} g' \left(\frac{Y - \mu}{\sigma} \right) = \frac{1}{\sigma} \mathbf{E} g'(X). \quad (\text{A.6})$$

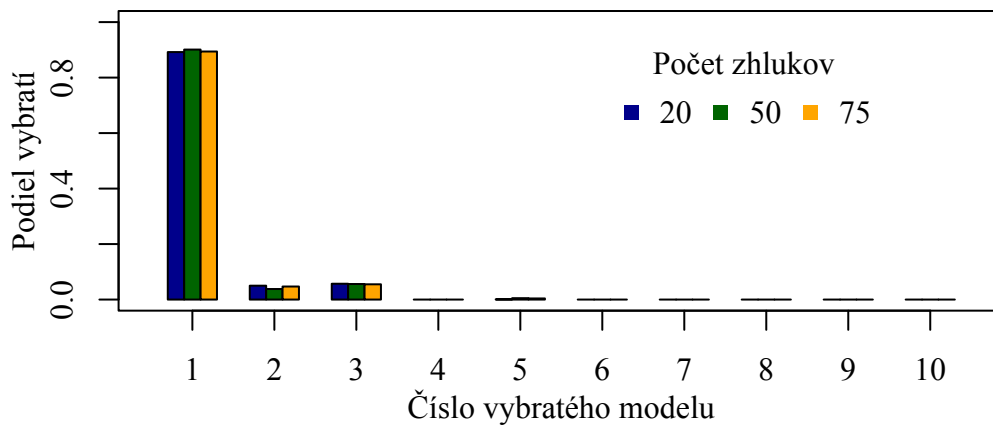
Taktiež platí

$$\mathbb{E} g'(X) = \mathbb{E} [Xg(X)] = \mathbb{E} \left[\frac{Y - \mu}{\sigma} g \left(\frac{Y - \mu}{\sigma} \right) \right] = \mathbb{E} \left[\frac{Y - \mu}{\sigma} h(Y) \right], \quad (\text{A.7})$$

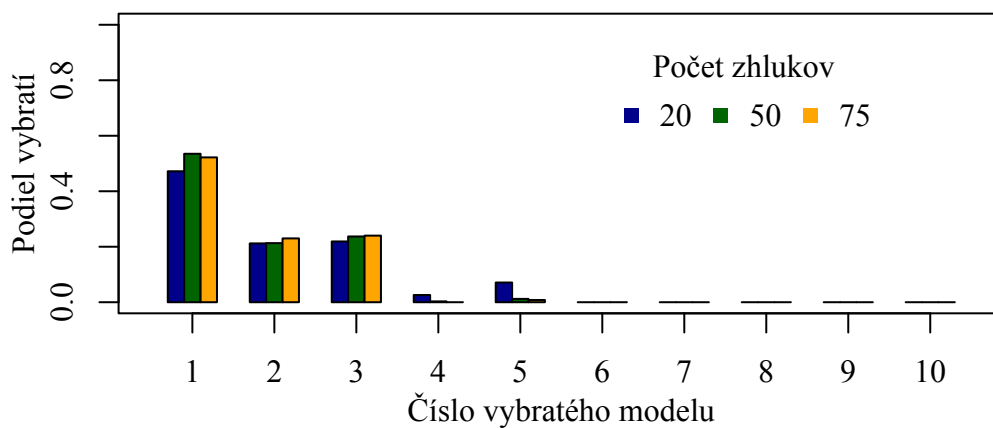
kde prvá rovnosť platí, pretože X má štandardné normálne rozdelenie a ďalšie rovnosti vychádzajú z definície funkcie h a transformácie $Y = \mu + \sigma X$. Dosadením vzťahu (A.7) do (A.6) sa získa tvrdenie. □

A.3 Grafické výstupy simulačnej štúdie č. 1

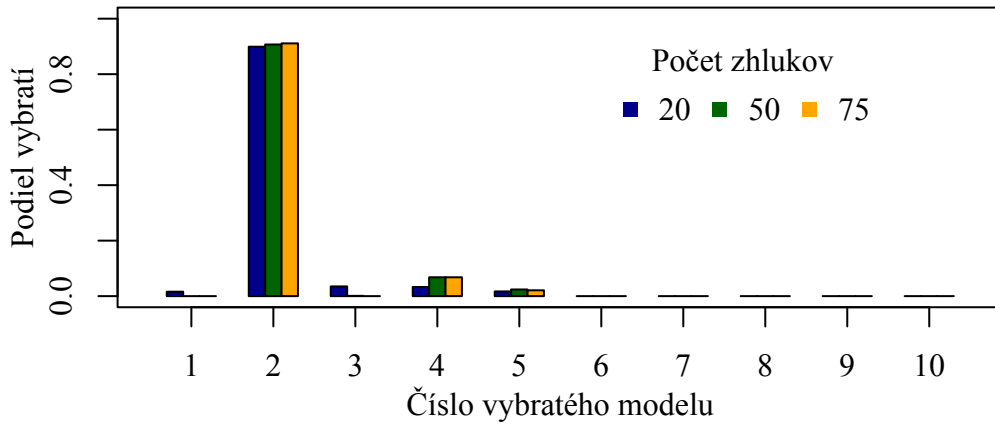
Zastúpenie vybratí jednotlivých modelov kritériom mAIC
v prípade, že je správny model č. 1.



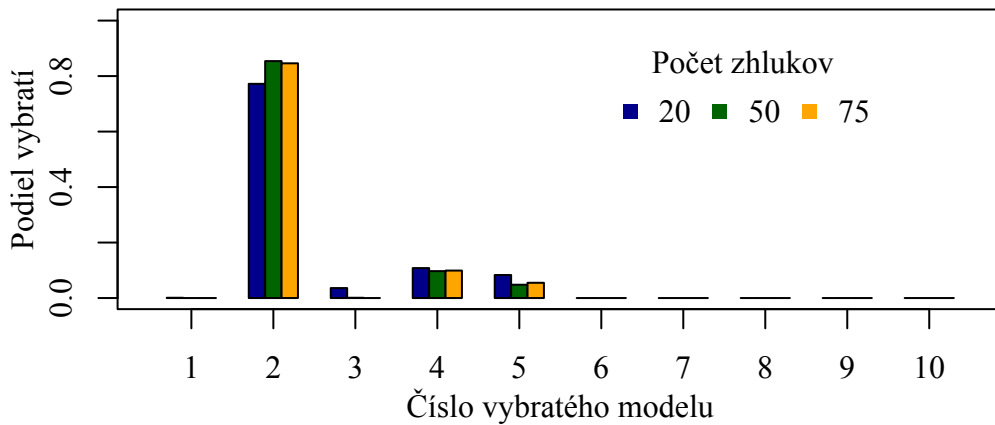
Zastúpenie vybratí jednotlivých modelov kritériom cAIC
v prípade, že je správny model č. 1.



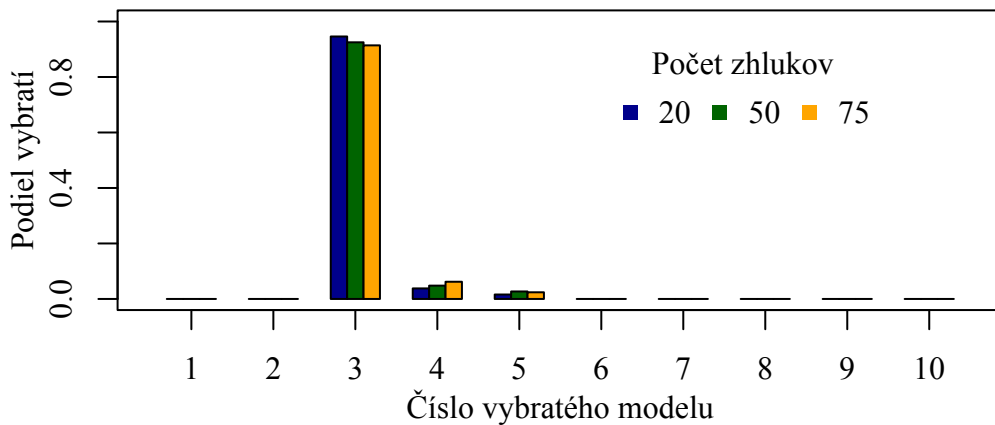
Zastúpenie vybratí jednotlivých modelov kritériom mAIC
v prípade, že je správny model č. 2.



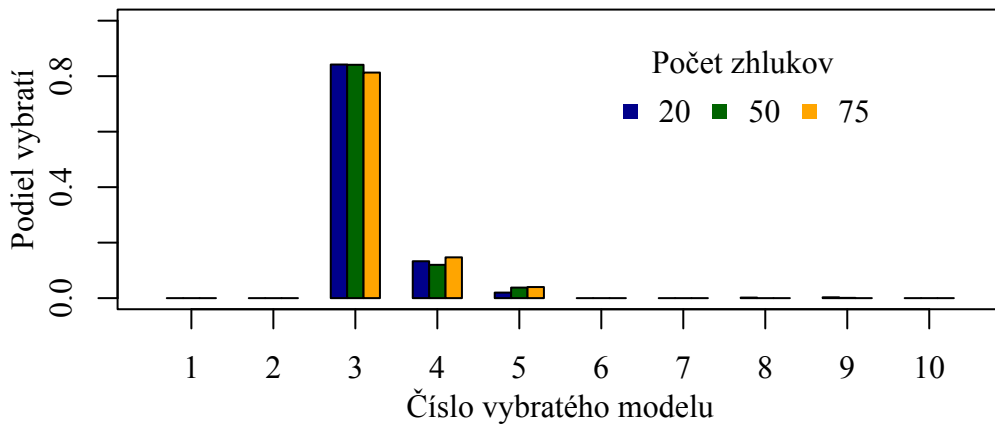
Zastúpenie vybratí jednotlivých modelov kritériom cAIC
v prípade, že je správny model č. 2.



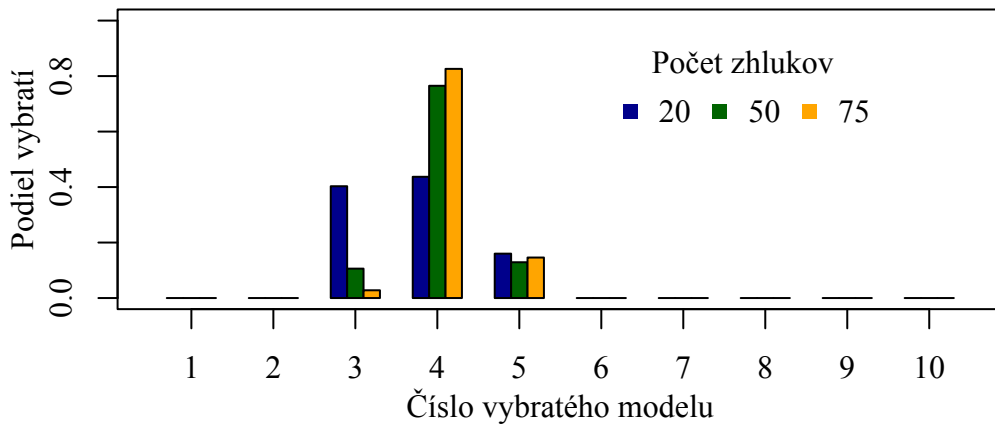
Zastúpenie vybratí jednotlivých modelov kritériom mAIC
v prípade, že je správny model č. 3.



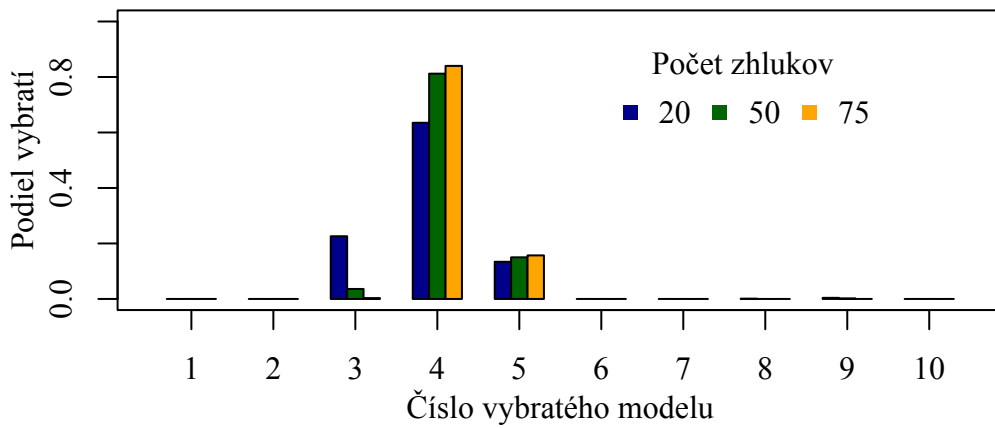
Zastúpenie vybratí jednotlivých modelov kritériom cAIC
v prípade, že je správny model č. 3.



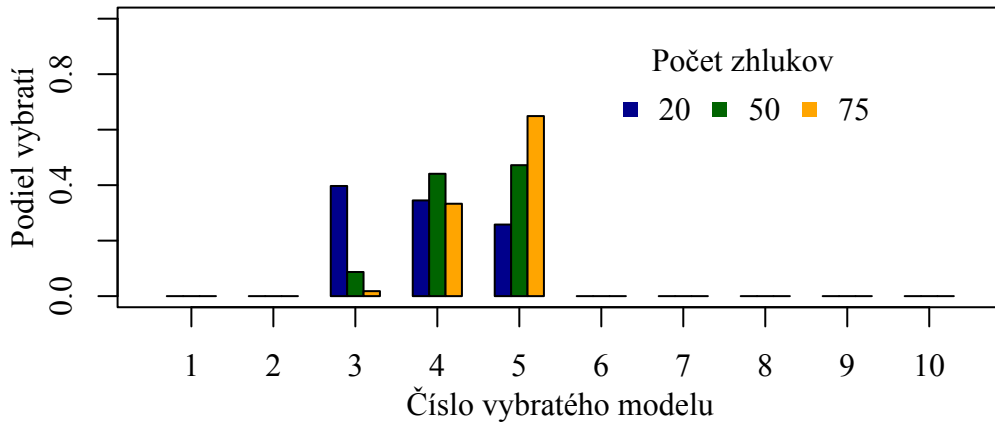
Zastúpenie vybratí jednotlivých modelov kritériom mAIC
v prípade, že je správny model č. 4.



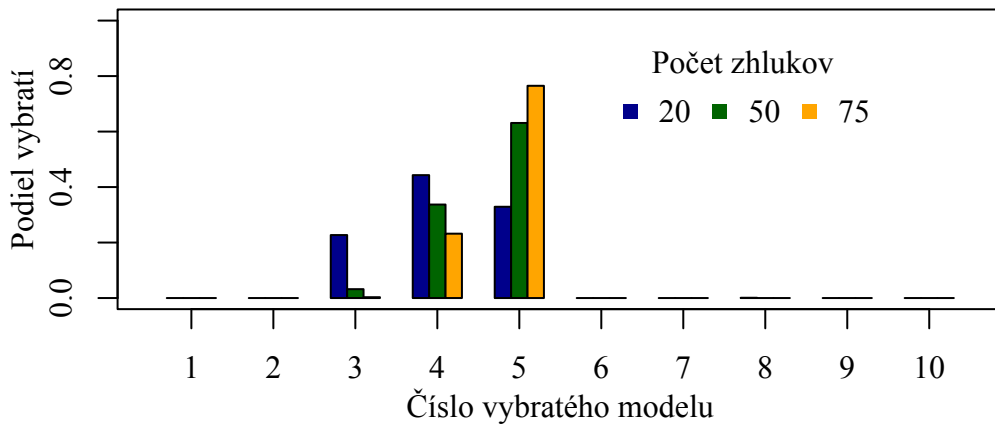
Zastúpenie vybratí jednotlivých modelov kritériom cAIC
v prípade, že je správny model č. 4.



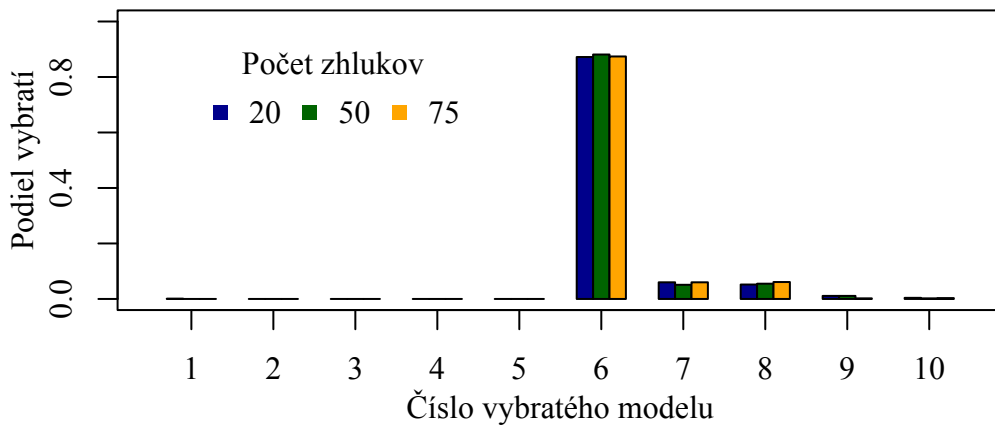
Zastúpenie vybratí jednotlivých modelov kritériom mAIC
v prípade, že je správny model č. 5.



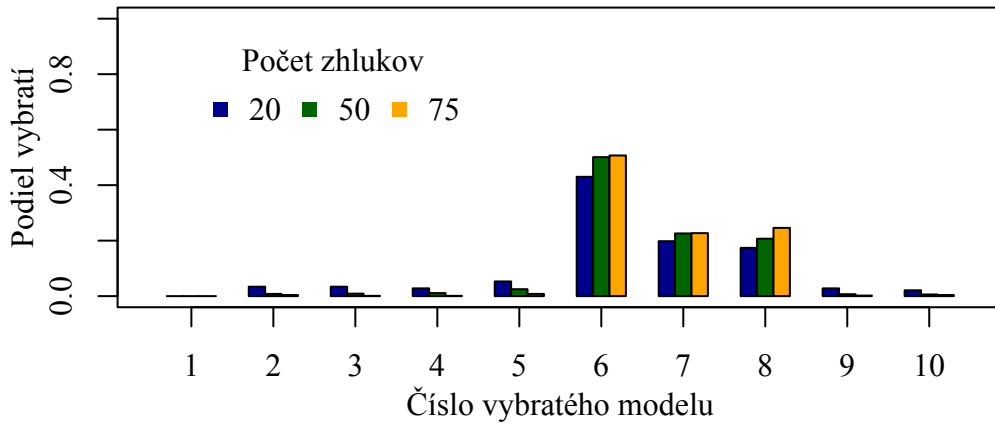
Zastúpenie vybratí jednotlivých modelov kritériom cAIC
v prípade, že je správny model č. 5.



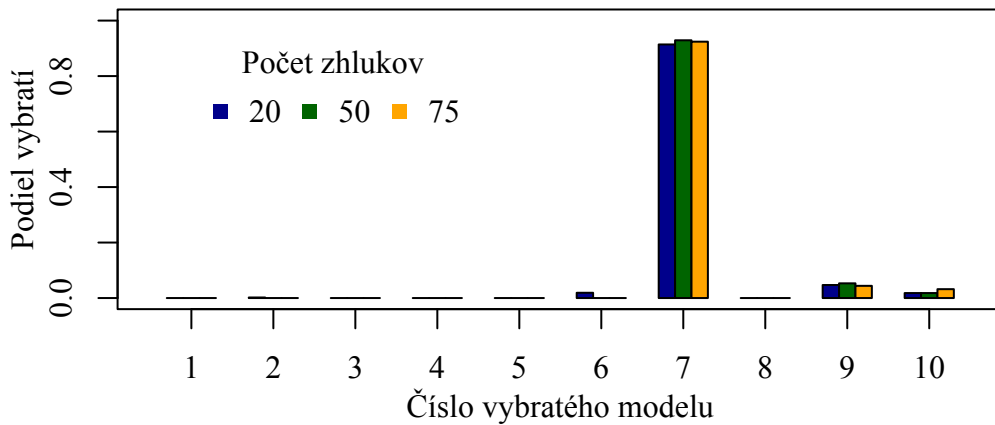
Zastúpenie vybratí jednotlivých modelov kritériom mAIC
v prípade, že je správny model č. 6.



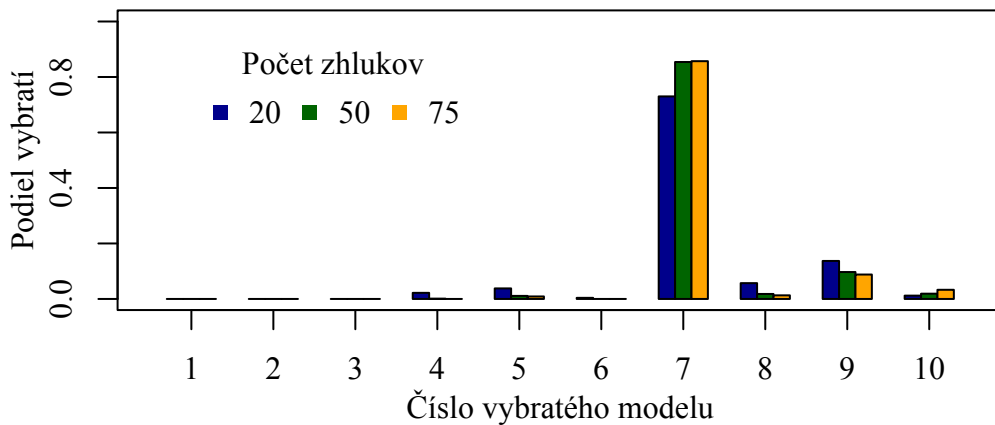
Zastúpenie vybratí jednotlivých modelov kritériom cAIC
v prípade, že je správny model č. 6.



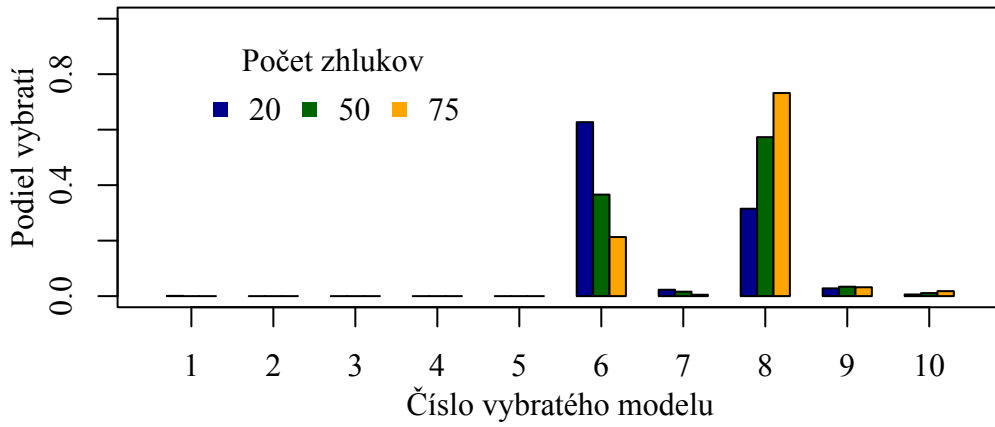
Zastúpenie vybratí jednotlivých modelov kritériom mAIC
v prípade, že je správny model č. 7.



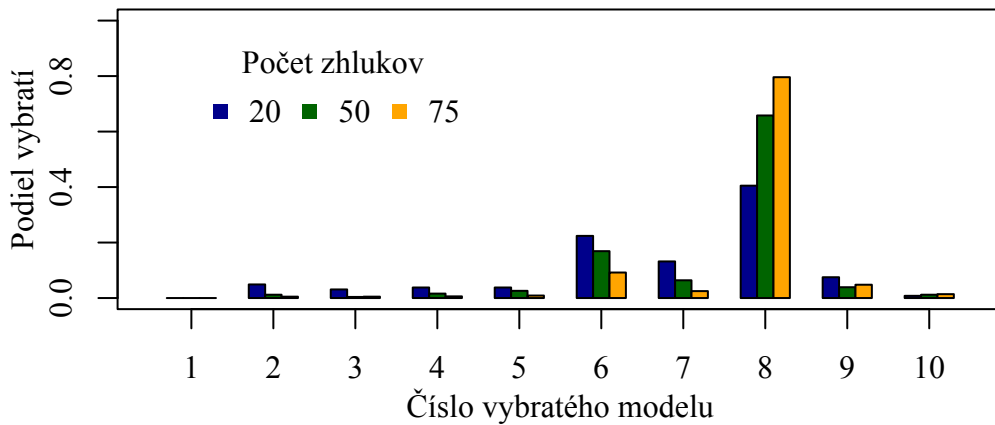
Zastúpenie vybratí jednotlivých modelov kritériom cAIC
v prípade, že je správny model č. 7.



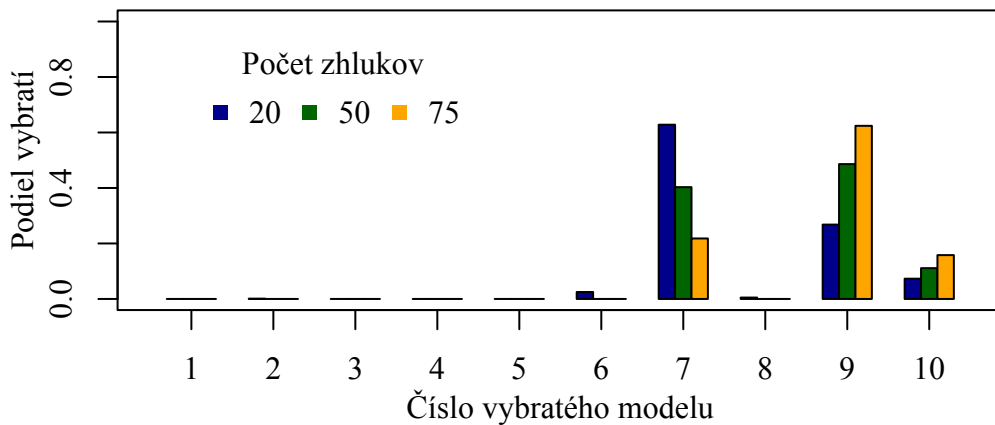
Zastúpenie vybratí jednotlivých modelov kritériom mAIC
v prípade, že je správny model č. 8.



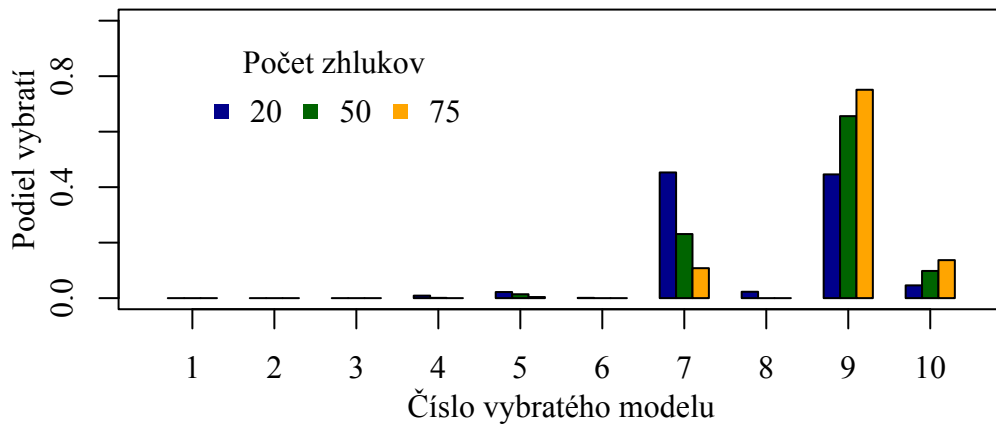
Zastúpenie vybratí jednotlivých modelov kritériom cAIC
v prípade, že je správny model č. 8.



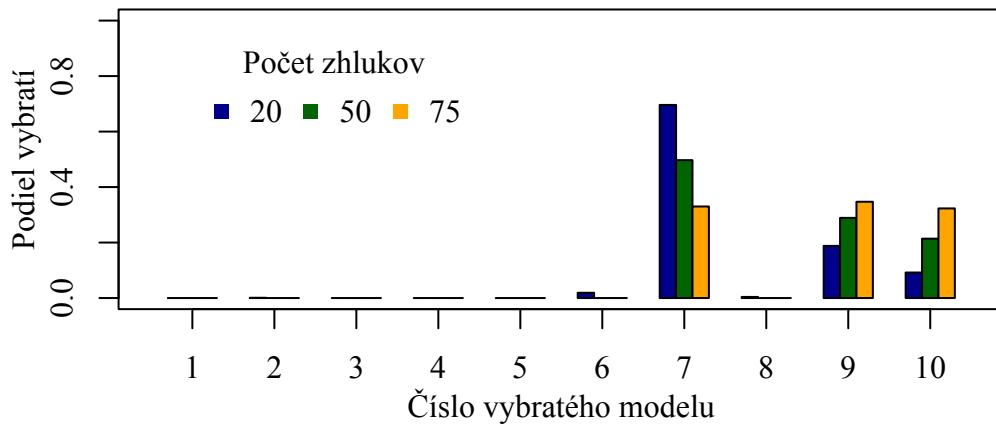
Zastúpenie vybratí jednotlivých modelov kritériom mAIC
v prípade, že je správny model č. 9.



Zastúpenie vybratí jednotlivých modelov kritériom cAIC
v prípade, že je správny model č. 9.



Zastúpenie vybratí jednotlivých modelov kritériom mAIC
v prípade, že je správny model č. 10.



Zastúpenie vybratí jednotlivých modelov kritériom cAIC
v prípade, že je správny model č. 10.

