

Exaktné testy a konfidenčné oblasti pre parametre normálneho lineárneho modelu s dvomi variančnými komponentami

Viktor WITKOVSKÝ¹

Ústav merania SAV, Bratislava
witkovsky@savba.sk



ROBUST 2012

17. letná škola JČMF, 9.9. - 14.9.2012, Němčičky, ČR

¹Práca vznikla v spolupráci s J. Volaufovou (School of Public Health, LSU Health Sciences Center, New Orleans, USA) a vďaka podpore grantov APVV-0096-10, VEGA 2/0019/10 a VEGA 2/0038/12

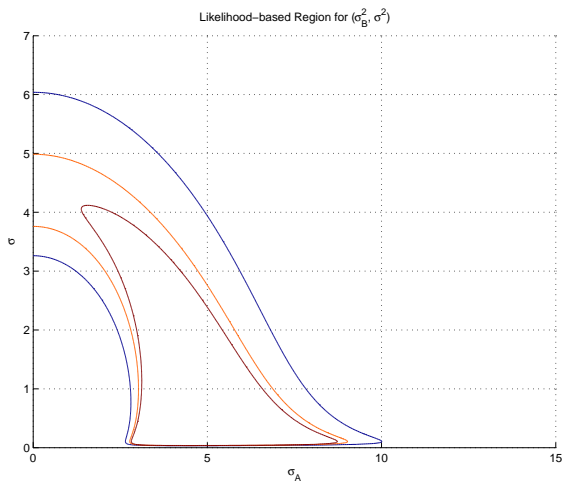
V príspevku popíšeme metódu konštrukcie konfidenčných oblastí založených na invertovaní exaktných testov založených na funkcii podielu vierohodností (LRT/RLRT) pre parametre normálneho regresného modelu s dvomi variančnými komponentami

$$Y \sim N_n(X\beta, \sigma^2 W(\lambda)),$$

kde X je známa ($n \times p$) matica, $\beta \in R^p$ je neznámy vektor parametrov a $\sigma^2 W(\lambda) = \sigma^2(I_n + \lambda V)$ je kovariančná matica, pričom V je známa nezáporne definitná matica, ktorá závisí od neznámych parametrov $\sigma^2 > 0$ a $\lambda \geq 0$.

Práca nadväzuje na publikácie:

- *Crainiceanu - Ruppert (JRSS 2004)*,
- *Chvosteková - Witkovský (MSR 2009)*,
- *Volaufová - Witkovský (TMMP 2012)*,
- *Witkovský - Volaufová (LINSTAT 2012)*.



Obr.: Ilustračný príklad: Exaktné konfidenčné oblasti pre variančné komponenty normálneho lineárneho modelu s dvomi variančnými komponentami

1 Úvod

2 Odvodenie exaktného nulového rozdelenia štatistík LRT/RLRT

- Značenie a štandardné výsledky
- Hypotéza 1: $H_0^1 : \theta = \theta_0$ a $\lambda = \lambda_0$
- Hypotéza 2: $H_0^2 : \theta = \theta_0$ a $\lambda = \lambda_0$ a $\sigma^2 = \sigma_0^2$
- Hypotéza 3: $H_0^3 : \lambda = \lambda_0$
- Hypotéza 4: $H_0^4 : \lambda = \lambda_0$ a $\sigma^2 = \sigma_0^2$
- Pravdepodobnosť nulového odhadu ML/REML parametra λ

3 Štandard asymptotické aproximácie

4 Empirická evidencia

5 Neštandardné asymptotické aproximácie

- Špeciálne prípady
- Vyvážený ANOVA model jednoduchého triedenia s náhodnými efektami

Uvažujme normálny lineárny regresný model s dvomi variančnými komponentami

$$Y \sim N_n(X\beta, \sigma^2 W(\lambda)),$$

kde

- Y je n -rozmerný vektor pozorovaní;
- X je známa ($n \times p$) matica, pričom $\text{rank}(X) = r_X \leq p$;
- $\beta \in \mathbb{R}^p$ je neznámy vektor parametrov;
- $\sigma^2 W(\lambda) = \sigma^2(I_n + \lambda V)$ je kovariančná matica, ktorá závisí od parametrov $\sigma^2 > 0$ a $\lambda \geq 0$.
- V je známa n.n.d. matica, $R(V) \not\subseteq R(X)$.
Často platí, že $V = ZZ'$ pre nejakú ($n \times k$) maticu Z , pričom $\text{rank}(V) = r_V \leq k$.

Špeciálnym prípadom modelu je **lineárny zmiešaný model** (LMM) s dvomi variančnými komponentami:

$$Y = X\beta + Zu + \varepsilon,$$

kde

- $u \sim N(0, \sigma_B^2 I_k)$ a $\varepsilon \sim N(0, \sigma^2 I_n)$, $u \perp \varepsilon$.
- $\lambda = \frac{\sigma_B^2}{\sigma^2}$ reprezentuje podiel variančných komponentov.

V príspevku budeme prezentovať tvar štatistík podielom vierohodností (LRT/RLRT) a ich exaktné rozdelenie (za platnosti nulovej hypotézy) pre testovanie týchto hypotéz:

Hypotéza 1: $H_0^1 : \theta = \theta_0 \text{ a } \lambda = \lambda_0$

Hypotéza 2: $H_0^2 : \theta = \theta_0 \text{ a } \lambda = \lambda_0 \text{ and } \sigma^2 = \sigma_0^2$

Hypotéza 3: $H_0^3 : \lambda = \lambda_0$

Hypotéza 4: $H_0^4 : \lambda = \lambda_0 \text{ a } \sigma^2 = \sigma_0^2$

kde

- $\theta = H' \beta, \theta_0 = H' \beta_0,$

- $H \in R(X')$ je známa matica, pričom $\text{rank}(H) = r_H \leq r_X.$

Exaktné (nulové) rozdelenie LRT/RLRT štatistík pre testovanie hypotéz H_0^1 a H_0^3 odvodili *Crainiceanu a Ruppert (JRSS 2004)*.

V príspevku uvidíme alternatívnu reprezentáciu testovacích štatistík a ich nulové rozdelenie, založené na výsledkoch: *Volaufová (Probastat 2011)*, *Volaufová (IWMS 2011)* a *Volaufová and Witkovský (TMMP 2012)*.

- Vo všeobecnosti neexistuje jednoduchá analytická reprezentácia exaktných distribučných funkcií (CDF) štatistík LRT/RLRT, ktorá by bola vhodná na (rýchle) numerické výpočty.
- Kvantily exaktných rozdelení LRT/RLRT možno odhadnúť z empirických distribučných funkcií, získaných pomocou simulácií.
- Takýto spôsob výpočtu je výpočtovo náročný, keďže v každom kroku simulácie je potrebná numerická optimalizácia (hľadanie maxima v 1-rozmernom priestore).

- Pre štatistiky LRT/RLRT budeme používať nasledovné označenie:

$$LR^j(Y | H_0^j \equiv \vartheta_0^j) = -2 \left(\sup_{H_0^j} \ell(\beta, \lambda, \sigma^2 | Y) - \sup \ell(\beta, \lambda, \sigma^2 | Y) \right) \quad i = 1, 2, 3, 4$$

$$LR_R^j(Y | H_0^j \equiv \vartheta_0^j) = -2 \left(\sup_{H_0^j} \ell_R(\lambda, \sigma^2 | Y) - \sup \ell_R(\lambda, \sigma^2 | Y) \right), \quad j = 3, 4$$

- $\vartheta_0^1 = (\theta_0, \lambda_0)$, $\vartheta_0^2 = (\theta_0, \lambda_0, \sigma_0^2)$, $\vartheta_0^3 = \lambda_0$, $\vartheta_0^4 = (\lambda_0, \sigma_0^2)$,
 - $\ell(\beta, \lambda, \sigma^2 | Y)$, resp. $\ell_R(\lambda, \sigma^2 | Y)$, je logaritmus funkcie vierohodnosti modelu (log-likelihood resp. restricted log-likelihood).
- Motiváciou a konečným cieľom tejto práce je konštrukcia simultánnych konfidenčných oblastí pre parametre modelu pomocou invertovania exaktných testov LRT/RLRT:

$$C^i(Y) = \{ \vartheta_0^i : LR^i(Y | \vartheta_0^i) \leq LR_{1-\alpha}^i(\vartheta_0^i) \}, \quad i = 1, 2, 3, 4$$

$$C_R^j(Y) = \{ \vartheta_0^j : LR_R^j(Y | \vartheta_0^j) \leq LR_{R,1-\alpha}^j(\vartheta_0^j) \}, \quad j = 3, 4$$

- $LR_{1-\alpha}^i(\vartheta_0^i)$ a $LR_{R,1-\alpha}^j(\vartheta_0^j)$ sú $(1 - \alpha)$ -kvantily nulového rozdelenia $LR^i(Y | H_0^i)$ resp. $LR_R^j(Y | H_0^j)$.
- Vo všeobecnosti, exaktné nulové rozdelenie $LR^i(Y | H_0^i)$ resp. $LR_R^j(Y | H_0^j)$ závisí od parametrov $\vartheta_0^i \in H_0^i$. Ako však ukážeme, exaktné rozdelenie LRT/RLRT štatistík pre testovanie uvedených hypotéz závisí len od hodnoty $\lambda_0 \in H_0^i$.

Budeme používať nasledujúce označenia:

- $M_X = I_n - P_X$, kde $P_X = X(X'X)^{-1}X'$ a $\text{rank}(M_X) = r_{M_X} = n - r_X$.
- Keďže $R(H) \subset R(X')$, potom existuje matica G , $R(G) \subset R(X)$ taká, že $H = X'G$, napr. $G = X(X'X)^{-1}H$.
- $M_G = I_n - P_G$, kde $P_G = G(G'G)^{-1}G'$, pričom $\text{rank}(P_G) = r_H$.
- $X_* = M_G X = X - X_{**}$, kde $X_{**} = P_G X$, $\text{rank}(X_{**}) = r_H$ a $\text{rank}(X_*) = r_X - r_H$.
- $M_{X_*} = I - P_{X_*}$, kde $P_{X_*} = X_*(X_*'X_*)^{-1}X_*'$ a $\text{rank}(P_{X_*}) = r_X - r_H$ a $\text{rank}(M_{X_*}) = n - r_X + r_H$.
- $M_{X_*} - M_X = P_{X_{**}} = X_{**}(X_{**}'X_{**})^{-1}X_{**}'$, kde $\text{rank}(P_{X_{**}}) = r_H$.
- **Ekvivalencia podpriestorov**

Afinná množina stredných hodnôt náhodného vektora Y je posunutý lineárny podpriestor, parametrizovaný vektorom parametrov γ bez ďalších obmedzení, t.j.

$$\{X\beta : H'\beta = H'\beta_0\} = X\beta_0 + \{X_*\gamma : \gamma \in \mathbb{R}^{r_X - r_H}\},$$

kde $X_* = M_G X$, and any $\beta_0 \in H_0^i$, $i = 1, 2$.

Viac podrobnosti možno nájsť napr. v prácach *Christensen (1987)*, alebo *LaMotte (1998)*.

Spektrálny rozklad matice V a výpočet podielu determinantov $|W(\lambda_0)|/|W(\lambda)|$.

Nech $0 \leq \xi_1 < \xi_2 < \dots < \xi_s$ označujú rôzne vlastné čísla matice V s násobnosťami ν_{ξ_i} pre ξ_i .

Platí, že nenulové vlastné čísla matíc V a $Z'Z$ sú rovnaké.

Potom, pre $W(\lambda) = I_n + \lambda V$ dostávame

$$\begin{aligned} \log(|W(\lambda)|) &= \sum_{i=1}^s \nu_{\xi_i} \log(1 + \lambda \xi_i) \\ \log\left(\frac{|W(\lambda_0)|}{|W(\lambda)|}\right) &= \sum_{i=1}^s \nu_{\xi_i} \log\left(\frac{1 + \lambda_0 \xi_i}{1 + \lambda \xi_i}\right) \end{aligned}$$

Uvedené výrazy závisia iba od nenulových vlastných čísel ξ_i matice V .

- Pre ľubovoľné $\lambda \geq 0$ označme

$$\hat{\sigma}_\lambda^2 = \frac{1}{n} Y'(M_X W(\lambda) M_X)^+ Y.$$

- Rozdelenie odhadu $\hat{\sigma}_\lambda^2$.

Nech $B'VB = \sum_{i=0}^r \mu_i E_i$ je spektrálny rozklad, kde B je $n \times (n - r_X)$ -matica, že $BB' = M_X$, $B'B = I_{n-r_X}$, $0 = \mu_0 < \mu_1 < \dots < \mu_r$ označujú rôzne vlastné čísla matice $B'VB$ s násobnosťami ν_{μ_i} pre μ_i a E_i sú navzájom ortogonálne projekčné matice. Platí, že nenulové čísla matíc $B'VB$ a $Z'M_X Z$ sú rovnaké. Podľa *Olsen, Seely, and Birkes (1976)* platí, že

$$Y'BE_iB'Y \sim \sigma_0^2(1 + \lambda_0\mu_i)Q_{\nu_{\mu_i}}, \quad i = 0, 1, \dots, r$$

je množina $r + 1$ navzájom nezávislých náhodných premenných, pričom

$Q_{\nu_{\mu_i}} \sim \chi_{\nu_{\mu_i}}^2$, kde λ_0, σ_0^2 sú skutočné hodnoty parametrov.

Odtiaľ pomocou vzťahov $(M_X W(\lambda) M_X)^+ = B(B'W(\lambda)B)^{-1}B'$ a $W(\lambda) = I_n + \lambda V$ dostávame

$$\frac{n\hat{\sigma}_\lambda^2}{\sigma_0^2} \sim \sum_{i=0}^r \frac{(1 + \lambda\mu_i)}{(1 + \lambda_0\mu_i)} Q_{\nu_{\mu_i}}.$$

- Pre $\lambda_0 \in H_0^i$, $i = 1, 2$ označme

$$\hat{\sigma}_{\lambda_0}^2 = \frac{1}{n} Y' (M_X W(\lambda_0) M_X)^+ Y,$$

$$\tilde{\sigma}_{\lambda_0}^2 = \frac{1}{n} Y_*' (M_{X_*} W(\lambda_0) M_{X_*})^+ Y_*,$$

kde $X_* = M_G X$ a $Y_* = Y - X\beta_0$, pre ľubovoľné $\beta_0 \in H_0^i$ (akákoľvek voľba $\beta_0 = (H')^{-1}\theta_0$).

- Rozdelenie odhadu $\tilde{\sigma}_{\lambda_0}^2 - \hat{\sigma}_{\lambda_0}^2$ za platnosti H_0^i , $i = 1, 2$.

Keďže platí $R(G) \subset R(X)$, dostávame

$$\begin{aligned} \frac{n\tilde{\sigma}_{\lambda_0}^2}{\sigma_0^2} - \frac{n\hat{\sigma}_{\lambda_0}^2}{\sigma_0^2} &= \frac{1}{\sigma_0^2} Y_*' [(M_{X_*} W(\lambda_0) M_{X_*})^+ - (M_X W(\lambda_0) M_X)^+] Y_* \\ &= Y_*' (P_{X_{**}} W(\lambda_0) P_{X_{**}})^+ Y_* \stackrel{H_0^i}{\sim} Q_{r_H}, \end{aligned}$$

kde $Q_{r_H} \sim \chi_{r_H}^2$ a λ_0, σ_0^2 označuje skutočné hodnoty parametrov. Navyše, Q_{r_H} je náhodná premenná nezávislá so všetkými $Q_{\nu_{\mu_i}}$.

Viac možno nájsť napr. v *Crainiceanu and Ruppert (JRSS 2004)* a *Volaufová (IWMS 2011)*.

Test podielom vierohodností LRT hypotézy 1: $H_0^1 : \theta = \theta_0$ a $\lambda = \lambda_0$

Na základe predpokladov o modeli a uvedených tvrdení dostávame:



$$LR^1(Y | H_0^1) = n \log \left(\hat{\sigma}_{\lambda_0}^2 \right) + \sup_{\lambda \geq 0} \left\{ -n \log \left(\hat{\sigma}_{\lambda}^2 \right) + \sum_{i=1}^s \nu_{\xi_i} \log \left(\frac{1 + \lambda_0 \xi_i}{1 + \lambda \xi_i} \right) \right\}$$



$$LR^1(Y | H_0^1) \sim n \log \left(Q_{r_H} + \sum_{i=0}^r Q_{\nu_i} \right) + \sup_{\lambda \geq 0} \left\{ -n \log \left(\sum_{i=0}^r \frac{1 + \lambda_0 \mu_i}{1 + \lambda \mu_i} Q_{\nu_i} \right) + \sum_{i=1}^s \nu_{\xi_i} \log \left(\frac{1 + \lambda_0 \xi_i}{1 + \lambda \xi_i} \right) \right\}$$

Test podielom vierohodností LRT hypotézy 2:

$$H_0^2 : \theta = \theta_0 \text{ a } \lambda = \lambda_0 \text{ a } \sigma^2 = \sigma_0^2$$



$$LR^2(Y | H_0^2) = \frac{n\tilde{\sigma}_{\lambda_0}^2}{\sigma_0^2} - n(1 - \log(n)) \\ + \sup_{\lambda \geq 0} \left\{ -n \log \left(\frac{n\hat{\sigma}_{\lambda}^2}{\sigma_0^2} \right) + \sum_{i=1}^s \nu_{\xi_i} \log \left(\frac{1 + \lambda_0 \xi_i}{1 + \lambda \xi_i} \right) \right\}$$



$$LR^2(Y | H_0^2) \sim Q_{r_H} + \sum_{i=0}^r Q_{\nu_i} - n(1 - \log(n)) \\ + \sup_{\lambda \geq 0} \left\{ -n \log \left(\sum_{i=0}^r \frac{1 + \lambda_0 \mu_i}{1 + \lambda \mu_i} Q_{\nu_i} \right) + \sum_{i=1}^s \nu_{\xi_i} \log \left(\frac{1 + \lambda_0 \xi_i}{1 + \lambda \xi_i} \right) \right\}$$

Test podielom vierohodností LRT hypotézy 3: $H_0^3 : \lambda = \lambda_0$

■

$$LR^3(Y | H_0^3) = n \log \left(\hat{\sigma}_{\lambda_0}^2 \right) + \sup_{\lambda \geq 0} \left\{ -n \log \left(\hat{\sigma}_{\lambda}^2 \right) + \sum_{i=1}^s \nu_{\xi_i} \log \left(\frac{1 + \lambda_0 \xi_i}{1 + \lambda \xi_i} \right) \right\}$$

■

$$LR^3(Y | H_0^3) \sim n \log \left(\sum_{i=0}^r Q_{\nu_i} \right) + \sup_{\lambda \geq 0} \left\{ -n \log \left(\sum_{i=0}^r \frac{1 + \lambda_0 \mu_i}{1 + \lambda \mu_i} Q_{\nu_i} \right) + \sum_{i=1}^s \nu_{\xi_i} \log \left(\frac{1 + \lambda_0 \xi_i}{1 + \lambda \xi_i} \right) \right\}$$

Test podielom vierohodností LRT hypotézy 4: $H_0^4 : \lambda = \lambda_0$ a $\sigma^2 = \sigma_0^2$

■

$$LR^4(Y | H_0^4) = \frac{n\hat{\sigma}_{\lambda_0}^2}{\sigma_0^2} - n(1 - \log(n)) \\ + \sup_{\lambda \geq 0} \left\{ -n \log \left(\frac{n\hat{\sigma}_{\lambda}^2}{\sigma_0^2} \right) + \sum_{i=1}^s \nu_{\xi_i} \log \left(\frac{1 + \lambda_0 \xi_i}{1 + \lambda \xi_i} \right) \right\}$$

■

$$LR^4(Y | H_0^4) \sim \sum_{i=0}^r Q_{\nu_i} - n(1 - \log(n)) \\ + \sup_{\lambda \geq 0} \left\{ -n \log \left(\sum_{i=0}^r \frac{1 + \lambda_0 \mu_i}{1 + \lambda \mu_i} Q_{\nu_i} \right) + \sum_{i=1}^s \nu_{\xi_i} \log \left(\frac{1 + \lambda_0 \xi_i}{1 + \lambda \xi_i} \right) \right\}$$

Test podielom vierohodností RLRT hypotézy 3: $H_0^3 : \lambda = \lambda_0$ 

$$LR_R^3(Y | H_0^3) = (n - r_X) \log(\hat{\sigma}_{\lambda_0}^2) + \sup_{\lambda \geq 0} \left\{ -(n - r_X) \log(\hat{\sigma}_\lambda^2) + \sum_{i=1}^r \nu_i \log\left(\frac{1 + \lambda_0 \mu_i}{1 + \lambda \mu_i}\right) \right\}$$



$$LR_R^3(Y | H_0^3) \sim (n - r_X) \log\left(\sum_{i=0}^r Q_{\nu_i}\right) + \sup_{\lambda \geq 0} \left\{ -(n - r_X) \log\left(\sum_{i=0}^r \frac{1 + \lambda_0 \mu_i}{1 + \lambda \mu_i} Q_{\nu_i}\right) + \sum_{i=1}^r \nu_i \log\left(\frac{1 + \lambda_0 \mu_i}{1 + \lambda \mu_i}\right) \right\}$$

Test podielom vierohodností RLRT hypotézy 4: $H_0^4 : \lambda = \lambda_0$ a $\sigma^2 = \sigma_0^2$

■

$$LR_R^4(Y | H_0^4) = \frac{n\hat{\sigma}_{\lambda_0}^2}{\sigma_0^2} - (n - r_X)(1 - \log(n - r_X)) \\ + \sup_{\lambda \geq 0} \left\{ -(n - r_X) \log \left(\frac{n\hat{\sigma}_{\lambda}^2}{\sigma_0^2} \right) + \sum_{i=1}^r \nu_i \log \left(\frac{1 + \lambda_0 \mu_i}{1 + \lambda \mu_i} \right) \right\}$$

■

$$LR_R^4(Y | H_0^4) \sim \sum_{i=0}^r Q_{\nu_i} - (n - r_X)(1 - \log(n - r_X)) \\ + \sup_{\lambda \geq 0} \left\{ -(n - r_X) \log \left(\sum_{i=0}^r \frac{1 + \lambda_0 \mu_i}{1 + \lambda \mu_i} Q_{\nu_i} \right) + \sum_{i=1}^r \nu_i \log \left(\frac{1 + \lambda_0 \mu_i}{1 + \lambda \mu_i} \right) \right\}$$

$$\Pr(\hat{\lambda}_{ML} = 0 \mid \lambda_0) \text{ \& } \Pr(\hat{\lambda}_{REML} = 0 \mid \lambda_0)$$

Označme

$$f_{LRT}(\lambda \mid \lambda_0) = \left\{ -n \log \left(\sum_{i=0}^r \frac{1 + \lambda_0 \mu_i}{1 + \lambda \mu_i} Q_{\nu_i} \right) + \sum_{i=1}^s \nu_{\xi_i} \log \left(\frac{1 + \lambda_0 \xi_i}{1 + \lambda \xi_i} \right) \right\}$$

$$f_{RLRT}(\lambda \mid \lambda_0) = \left\{ -(n - r_X) \log \left(\sum_{i=0}^r \frac{1 + \lambda_0 \mu_i}{1 + \lambda \mu_i} Q_{\nu_i} \right) + \sum_{i=1}^r \nu_i \log \left(\frac{1 + \lambda_0 \mu_i}{1 + \lambda \mu_i} \right) \right\}$$

Potom $\hat{\lambda}_{ML} = \arg \sup_{\lambda \geq 0} \{f_{LRT}(\lambda \mid \lambda_0)\}$, a $\hat{\lambda}_{REML} = \arg \sup_{\lambda \geq 0} \{f_{RLRT}(\lambda \mid \lambda_0)\}$.

Podmienenu pravdepodobnosť toho, že ML/REML odhad parametra λ bude 0 pokiaľ skutočný parameter je λ_0 , možno odhadnúť:

$$\begin{aligned} \Pr(\hat{\lambda}_{ML} = 0 \mid \lambda_0) &\approx \Pr \left(\frac{\partial f_{LRT}(\lambda \mid \lambda_0)}{\partial \lambda} \Big|_{\lambda=0} \leq 0 \right) \\ &= \Pr \left(\sum_{i=0}^r (\mu_i - \bar{\xi})(1 + \lambda_0 \mu_i) Q_{\nu_i} \leq 0 \right) \end{aligned}$$

$$\Pr(\hat{\lambda}_{REML} = 0 \mid \lambda_0) \approx \Pr \left(\sum_{i=0}^r (\mu_i - \bar{\mu})(1 + \lambda_0 \mu_i) Q_{\nu_i} \leq 0 \right)$$

kde $\bar{\xi} = \frac{1}{n} \sum_{i=1}^s \nu_{\xi_i} \xi_i$ a $\bar{\mu} = \frac{1}{n - r_X} \sum_{i=1}^r \nu_i \mu_i$.

- Za predpokladu, že požadované predpoklady platia a **testovaná hodnota λ_0 je vo vnútri parametrického priestoru**, t.j. $\lambda_0 > 0$, štandardnou asymptotickou aproximáciou nulového rozdelenia danej LRT/RLRT štatistiky pre dostatočne veľký rozsah výberu je

$$LR(Y | H_0) \xrightarrow{D} \chi_d^2,$$

kde d je počet reštrikcií na parametre modelu.

- Pre λ_0 na hranici parametrického priestoru, t.j. $\lambda_0 = 0$, *Stram and Lee (Biometrics 1994)* navrhli aproximáciu nulového rozdelenia danej LRT/RLRT štatistiky pre dostatočne veľký rozsah výberu ako zmes rozdelení

$$LR(Y | H_0) \xrightarrow{D} 0.5\chi_{d-1}^2 : 0.5\chi_d^2,$$

kde d je počet reštrikcií na parametre modelu.

- *Pinheiro and Bates (2000)* navrhli aproximáciu nulového rozdelenia danej LRT/RLRT štatistiky pre λ_0 na hranici parametrického priestoru ako zmes rozdelení s rozdielnými váhami:

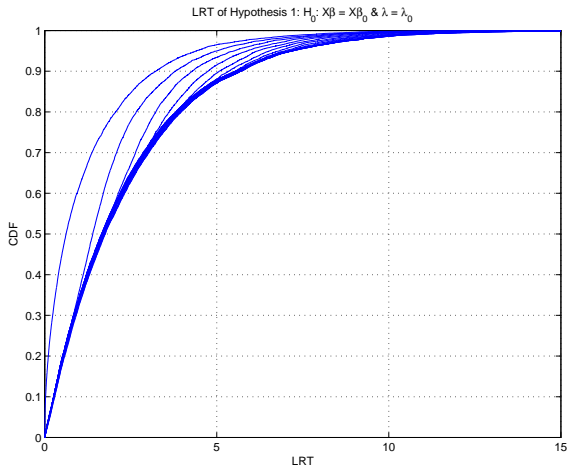
$$LR(Y | H_0) \xrightarrow{D} 0.65\chi_{d-1}^2 : 0.35\chi_d^2,$$

kde d je počet reštrikcií na parametre modelu.

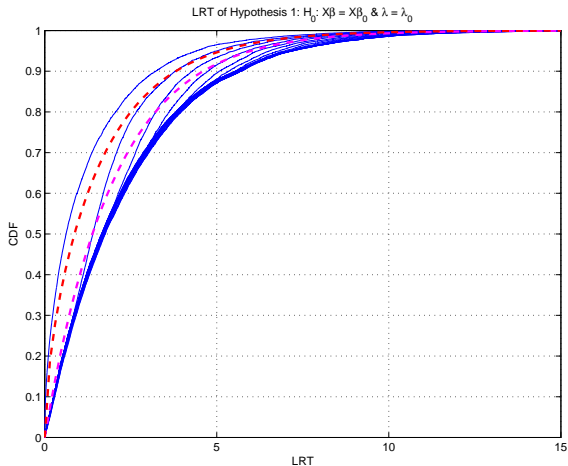
Vlastnosti exaktného nulového rozdelenia LRT/RLRT štatistík budeme ilustrovať na špecifickom príkaldе: Uvažujme model

$$Y = X\beta + Zu + \varepsilon,$$

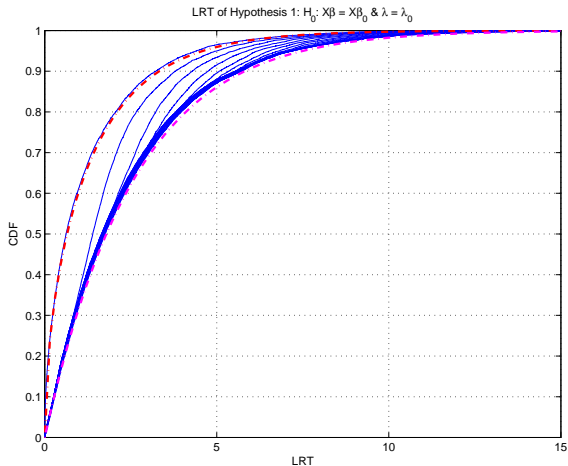
- Uvažujeme tu model vhodný pre výbery z dvoch populácií s rôznym počtom výberových jednotiek (m_1, m_2), tu $m = m_1 + m_2 = 5 + 7 = 12$ výberových jednotiek (subjektov), pričom na subjekte i máme T_i opakovaných pozorovaní v rôznych časoch, $T_i \in \{1, 3, 3, 6, 9\}$ pre prvú populáciu, a $T_i \in \{3, 5, 9, 11, 5, 4, 3\}$ pre druhú populáciu.
- Pozorovania Y , modelujeme ako $Y = [X_1, X_2](\beta_1', \beta_2')'$, kde Y je $n = \sum_{i=1}^m T_i = 62$ dimenzionálny vektor, X_1 reprezentuje s vysvetľujúcu premennú (funkciu času) a stĺpce X_2 reprezentujú pomocné (dummy) premenné pre rozlíšenie dvoch populácií. Teda β_1 je skalár, a β_2 je 2-rozmerný vektor. V tomto modeli uvažujeme ako $H_0^{1,2}$ hypotézu, že $\beta_1 = 0$.
- Odozva Y bola modelovaná pomocou modelu s náhodnými koeficientami (intercepts), s $(n \times m)$ maticou Z plnej hodnosti, kde stĺpce reprezentovali pomocné premenné pre m výberových jednotiek (subjektov), t.j. $V = ZZ'$, pričom $r_V = 12$, teda pomocou LMM s dvomi variančnými komponentami. Viac podrobnosti o štruktúre daného modelu možno nájsť v práci *Volaufová and Witkovský (TMMP 2012)*.

LRT hypotézy 1: $H_0^1: \theta = \theta_0$ a $\lambda = \lambda_0$ 

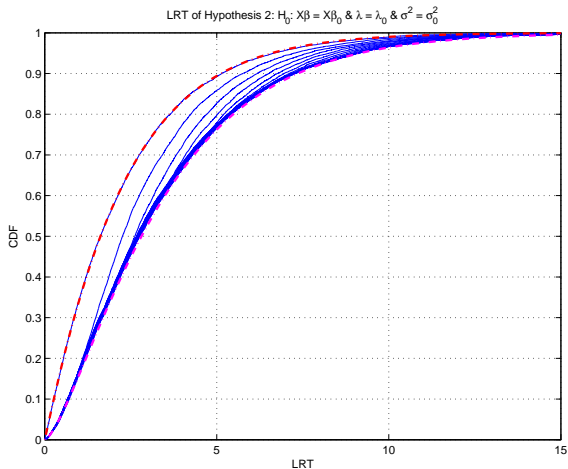
Obr.: Exaktné a približné nulové rozdelenia štatistiky $LR^1(Y | H_0^1)$ pre rôzne hodnoty λ_0 , pričom $h = r_H = 1$. Tenké čiary zobrazujú exaktné CDF pre $\lambda_0 = 0 : 0.05 : 1$ (zľava). Hrube čiary zobrazujú aproximácie: (i) zmes rozdelení $0.5\chi^2(h)$: $0.5\chi^2(h+1)$ a rozdelením $\chi^2(h+1)$, (ii) zmes rozdelení $w\chi^2(h) : (1-w)\chi^2(h+1)$, kde $w = \Pr(\hat{\lambda}_{ML} = 0 | \lambda_0 = 0) = 0.7362$ a necentrálnym rozdelením $\chi_{h+1,\delta}^2$ s odhadnutým parametrom necentrality $\delta = 0.45$.

LRT hypotézy 1: $H_0^1: \theta = \theta_0$ a $\lambda = \lambda_0$ 

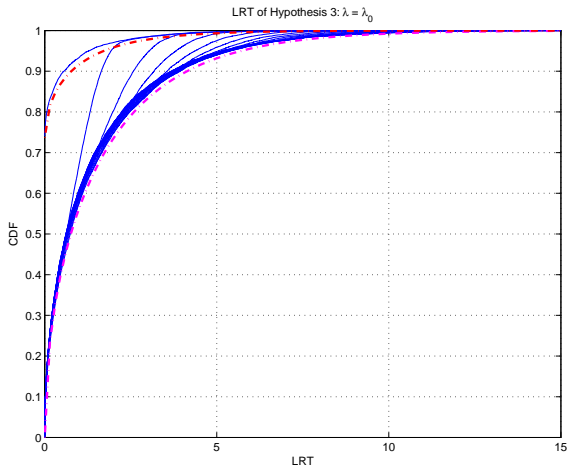
Obr.: Exaktné a približné nulové rozdelenia štatistiky $LR^1(Y | H_0^1)$ pre rôzne hodnoty λ_0 , pričom $h = r_H = 1$. Tenké čiary zobrazujú exaktné CDF pre $\lambda_0 = 0 : 0.05 : 1$ (zľava). Hrube čiary zobrazujú aproximácie: (i) zmes rozdelení $0.5\chi^2(h)$: $0.5\chi^2(h+1)$ a rozdelením $\chi^2(h+1)$, (ii) zmes rozdelení $w\chi^2(h) : (1-w)\chi^2(h+1)$, kde $w = \Pr(\hat{\lambda}_{ML} = 0 | \lambda_0 = 0) = 0.7362$ a necentrálnym rozdelením $\chi_{h+1,\delta}^2$ s odhadnutým parametrom necentrality $\delta = 0.45$.

LRT hypotézy 1: $H_0^1: \theta = \theta_0$ a $\lambda = \lambda_0$ 

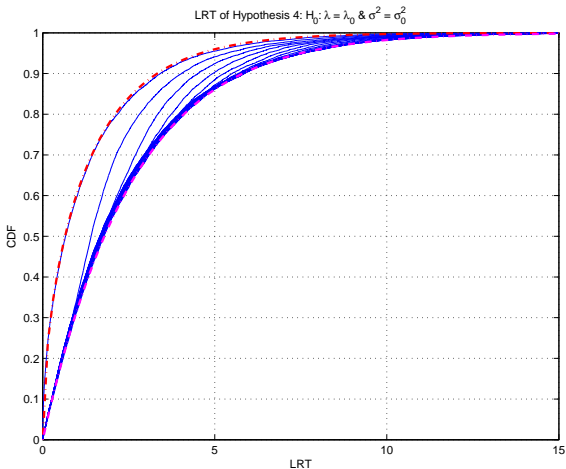
Obr.: Exaktné a približné nulové rozdelenia štatistiky $LR^1(Y | H_0^1)$ pre rôzne hodnoty λ_0 , pričom $h = r_H = 1$. Tenké čiary zobrazujú exaktné CDF pre $\lambda_0 = 0 : 0.05 : 1$ (zľava). Hrube čiary zobrazujú aproximácie: (i) zmes rozdelení $0.5\chi^2(h)$: $0.5\chi^2(h+1)$ a rozdelením $\chi^2(h+1)$, (ii) zmes rozdelení $w\chi^2(h) : (1-w)\chi^2(h+1)$, kde $w = \Pr(\hat{\lambda}_{ML} = 0 | \lambda_0 = 0) = 0.7362$ a necentrálnym rozdelením $\chi_{h+1,\delta}^2$ s odhadnutým parametrom necentrality $\delta = 0.45$.

LRT hypotézy 2: $H_0^2 : \theta = \theta_0$ a $\lambda = \lambda_0$ a $\sigma^2 = \sigma_0^2$ 

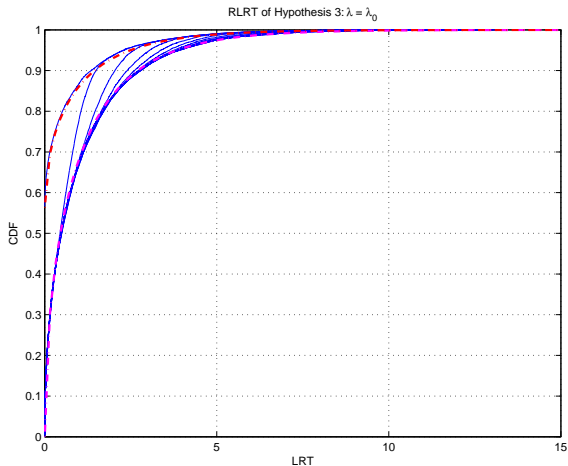
Obr.: Exaktné a približné nulové rozdelenia štatistiky $LR^2(Y | H_0^2)$ pre rôzne hodnoty λ_0 , pričom $h = r_H = 1$. Tenké čiary zobrazujú exaktné CDF pre $\lambda_0 = 0 : 0.05 : 1$ (zľava). Hrube čiary zobrazujú aproximácie: zmes rozdelení $w\chi^2(h+1) : (1-w)\chi^2(h+2)$, kde $w = \Pr(\hat{\lambda}_{ML} = 0 | \lambda_0 = 0) = 0.7362$ a necentrálnym rozdelením $\chi_{h+2, \delta}^2$ s odhadnutým parametrom necentrality $\delta = 0.45$.

LRT hypotézy 3: $H_0^3 : \lambda = \lambda_0$ 

Obr.: Exaktné a približné nulové rozdelenia štatistiky $LR^3(Y | H_0^3)$ pre rôzne hodnoty λ_0 . Tenké čiary zobrazujú exaktné CDF pre $\lambda_0 = 0 : 0.05 : 1$ (zľava). Hrubé čiary zobrazujú aproximácie: zmes rozdelení $w\chi^2(0) : (1-w)\chi^2(1)$, kde $w = \Pr(\hat{\lambda}_{ML} = 0 | \lambda_0 = 0) = 0.7362$ a necentrálnym rozdelením $\chi^2_{1,\delta}$ s odhadnutým parametrom necentrality $\delta = 0.45$.

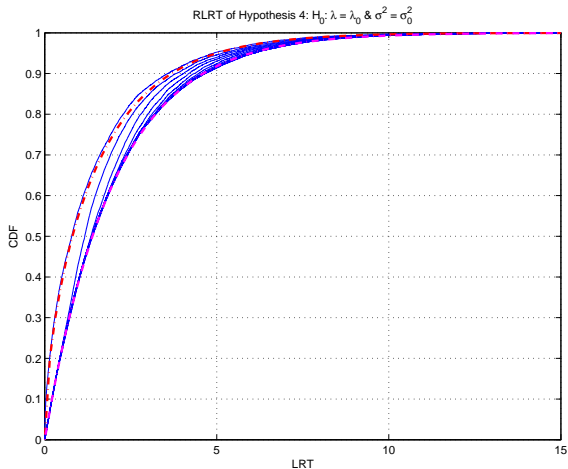
LRT hypotézy 4: $H_0^4 : \lambda = \lambda_0$ a $\sigma^2 = \sigma_0^2$ 

Obr.: Exaktné a približné nulové rozdelenia štatistiky $LR^4(Y | H_0^4)$ pre rôzne hodnoty λ_0 . Tenké čiary zobrazujú exaktné CDF pre $\lambda_0 = 0 : 0.05 : 1$ (zľava). Hrubé čiary zobrazujú aproximácie: zmes rozdelení $w\chi^2(1) : (1 - w)\chi^2(2)$, kde $w = \Pr(\hat{\lambda}_{ML} = 0 | \lambda_0 = 0) = 0.7362$ a necentrálnym rozdelením $\chi_{2,\delta}^2$ s odhadnutým parametrom necentrality $\delta = 0.45$.

RLRT hypotézy 3: $H_0^3 : \lambda = \lambda_0$ 

Obr.: Exaktné a približné nulové rozdelenia štatistiky $LR_R^3(Y | H_0^3)$ pre rôzne hodnoty λ_0 . Tenké čiary zobrazujú exaktné CDF pre $\lambda_0 = 0 : 0.05 : 1$ (zľava). Hrubé čiary zobrazujú aproximácie: zmes rozdelení $w\chi^2(0) : (1-w)\chi^2(1)$, kde $w = \Pr(\hat{\lambda}_{REML} = 0 | \lambda_0 = 0) = 0.5635$ a centrálnym rozdelením χ_1^2 .

Test podielom vierohodností RLRT hypotézy 4: $H_0^4 : \lambda = \lambda_0$ a $\sigma^2 = \sigma_0^2$



Obr.: Exaktné a približné nulové rozdelenia štatistiky $LR_R^3(Y | H_0^3)$ pre rôzne hodnoty λ_0 . Tenké čiary zobrazujú exaktné CDF pre $\lambda_0 = 0 : 0.05 : 1$ (zľava). Hrube čiary zobrazujú aproximácie: zmes rozdelení $w\chi^2(1) : (1 - w)\chi^2(2)$, kde $w = \Pr(\hat{\lambda}_{REML} = 0 | \lambda_0 = 0) = 0.5635$ a centrálnym rozdelením χ^2_2 .

Exaktné nulové rozdelenie LR/RLRT štatistík závisí na vlastných číslach $\mu_{i,n}$ a $\xi_{j,n}$ matíc modelu pre n -rozmerný vektor Y .

Pre testovanie hypotéz H_0^1 a H_0^3 , pričom $\lambda_0 = 0$, *Crainiceanu and Rupert (2004)* navrhli asymptotickú aproximáciu, založenú na predpokladoch:

- Nech $\text{rank}(V) = r_V$ je pevne daná a nech

$$\lim_{n \rightarrow \infty} \frac{\mu_{i,n}}{n^\gamma} = \mu_{i,\infty}$$

$$\lim_{n \rightarrow \infty} \frac{\xi_{i,n}}{n^\gamma} = \xi_{i,\infty}$$

pre nejaké $\gamma \geq 0$.

- Potom

$$LR^1(Y | H_0^1) \xrightarrow{D} Q_{r_H} + \sup_{\bar{\lambda} \geq 0} \left\{ \sum_{i=1}^r \frac{\bar{\lambda} \mu_{i,\infty}}{1 + \bar{\lambda} \mu_{i,\infty}} Q_{\nu_i} - \sum_{j=1}^s \nu_{\xi_j} \log(1 + \bar{\lambda} \xi_{j,\infty}) \right\}$$

$$LR^3(Y | H_0^3) \xrightarrow{D} \sup_{\bar{\lambda} \geq 0} \left\{ \sum_{i=1}^r \frac{\bar{\lambda} \mu_{i,\infty}}{1 + \bar{\lambda} \mu_{i,\infty}} Q_{\nu_i} - \sum_{j=1}^s \nu_{\xi_j} \log(1 + \bar{\lambda} \xi_{j,\infty}) \right\},$$

$$LR_R^3(Y | H_0^3) \xrightarrow{D} \sup_{\bar{\lambda} \geq 0} \left\{ \sum_{i=1}^r \frac{\bar{\lambda} \mu_{i,\infty}}{1 + \bar{\lambda} \mu_{i,\infty}} Q_{\nu_i} - \sum_{i=1}^r \nu_{\mu_i} \log(1 + \bar{\lambda} \mu_{i,\infty}) \right\}.$$

Pre ilustráciu uvažujme vyvážený ANOVA model jednoduchého triedenia s náhodnými efektami, s K úrovňami a J pozorovaniami pre každú úroveň. Potom, $n = KJ$, $\text{rank}(V) = r_V = K$ a zodpovedajúce vlastné čísla sú

- $\mu_{0,n} = 0$ s násobnosťou $\nu_{\mu,0} = n - K$,
- $\mu_{1,n} = J$ s násobnosťou $\nu_{\mu,1} = K - 1$,
- $\xi_{1,n} = 0$ s násobnosťou $\nu_{\xi,1} = n - K$,
- $\xi_{2,n} = J$ s násobnosťou $\nu_{\xi,2} = K$.

Za platnosti nulovej hypotézy $H_0^3 : \lambda_0 = 0$ exaktné rozdelenie RLRT štatistiky $LR_R^3(Y | H_0^3)$ je

$$LR_R^3(Y | H_0^3) \sim n \log(Q_{K-1} + Q_{n-K}) + \inf_{\lambda \geq 0} \left\{ -n \log \left(\frac{Q_{K-1}}{1 + \lambda J} + Q_{n-K} \right) - K \log(1 + \lambda J) \right\}$$

Crainiceanu and Rupert (2004) navrhli asymptotickú aproximáciu rozdelenia pre $J \rightarrow \infty$ a konštantnú hodnotu K ako

$$LR_R^3(Y | H_0^3) \xrightarrow{D} \{(Q_{K-1} - K) - K \log(Q_{K-1}/K)\} \times I(Q_{K-1} > K),$$

Toto asymptotické rozdelenie má v nule hodnotu rovnú $\Pr(Q_{K-1} < K)$.