

5. Úkol na předmět Mnohorozměrná analýza

Jaromír Macoun

Březen 2022

1 Simulace

V tomto úkolu jsme měli provést simulaci, která by měla porovnat dva podobné testy.

Mějme tedy náhodný výběr X_1, \dots, X_n z dvourozměrného normálního rozdělení s vektorem střední hodnoty $\mu = (1, 2)^T$ a varianční maticí $\Sigma = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 2 \end{pmatrix}$.

V našem případě nás zajímá zdali lineární kombinace $2\mu_1 - 1\mu_2$ je nulová. Označíme $\mathbb{A} = (2, -1)$.

Mějme nulovou hypotézu $H_0 : \mathbb{A}\mu = 0$ proti alternativě $H_1 : \mathbb{A}\mu \neq 0$

Dále máme dvě testové statistiky

$$T_n = n(\mathbb{A}\bar{\mathbf{X}}_n - a)^\top (\mathbb{A}\Sigma\mathbb{A}^\top)^{-1} (\mathbb{A}\bar{\mathbf{X}}_n - a) \sim \chi_1^2.$$

za předpokladu známé varianční matice *Sigma*.

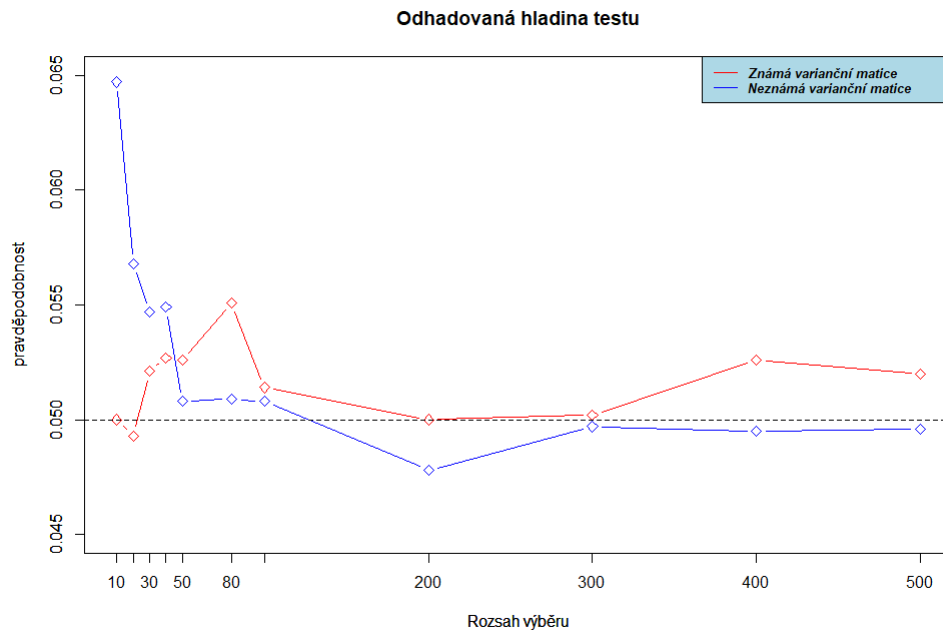
$$\tilde{T}_n = (n-1)(\mathbb{A}\bar{\mathbf{X}}_n - a)^\top (\mathbb{A}\mathcal{S}\mathbb{A}^\top)^{-1} (\mathbb{A}\bar{\mathbf{X}}_n - a) \sim T_{1, n-1}^2.$$

za předpokladu, že neznáme varianční matici, tak jí musíme odhadnout. Tento odhad je značen symbolem \mathcal{S} .

Monte-Carlo simulace o 10000 opakováních byla provedena pro zjištění, jaké hladiny dosahují jednotlivé testy. Simulace byla provedena pro rozsahy výběru 10, 20, 30, 40, 50, 80, 100, 200, 300, 400 a 500. Na obrázku č. 1 jsou shrnuty výsledky simulace. Je vidět, že pro malé rozsahy test s odhadnutou varianční maticí absolutně nedodrhuje hladinu. Od rozsahu výběru dále se zdá, že test s předpokladem známého Σ má vyšší hladinu než druhý test. Pro rozsahy výběru 300, 400, 500, lépe dodrhuje hladinu test s neznámou varianční maticí, na základě výsledků simulace.

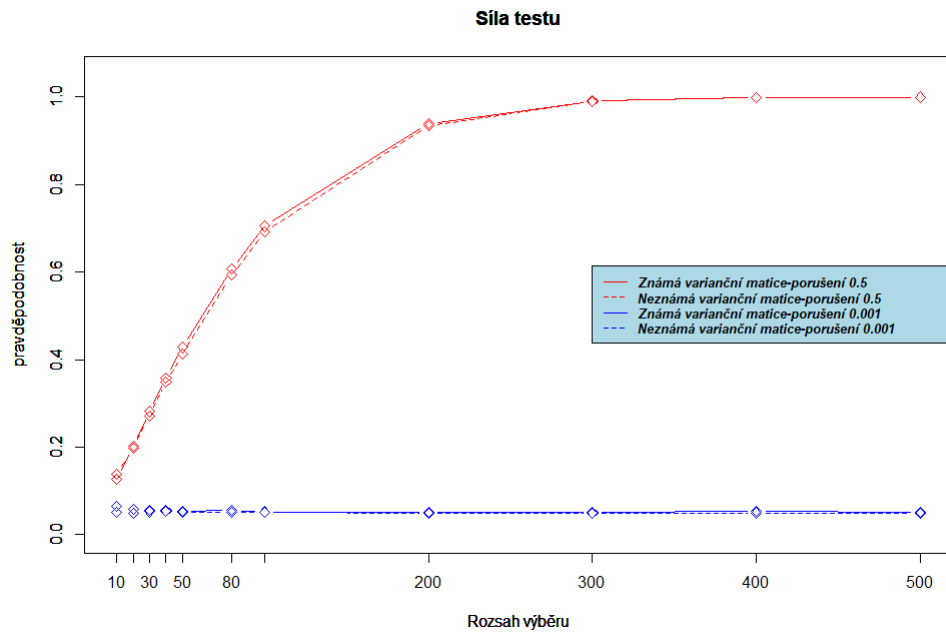
Dále jsme se zabývali dvěma porušením nulové hypotézy a to následujícím. Prvně jsme uvažovali odlišnost $\mu_2 = 1, 5$. Pro tento případ je vidět na obrázku č. 2, že již od relativně malého rozsahu výběru test spíše zamítne. (Pro rozsah výběru 80 již je simulovaná pravděpodobnost zamítnutí zhruba 0,6).

Pro porušení hypotézy takové, že položíme $\mu_2 = 1,999$ z obrázku č. 2 vyplývá, že test se drží zhruba okolo hladiny 0,05. Tedy toto porušení nulové hypotézy se zdá, že ani při rozsahu výběru 500 spíše test neodhalí.



Obrázek 1: Porovnání hladin dvou testů

Co se týče porovnání síly mezi oběma testy (známá-neznámá varianční matice), tak oba se zdají být stejné. U většího porušení hypotézy se zdá test s odhadnutou varianční maticí mírně slabší, ale opravdu jen mírně a u malých rozsahů výběru.



Obrázek 2: Síla testu v závislosti na rozsahu výběru