

**MATEMATICKO-FYZIKÁLNÍ
FAKULTA**
Univerzita Karlova

DIPLOMOVÁ PRÁCE

Bc. Jan Vávra

Bayesovská faktorová analýza

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 dne

Podpis autora

Rád bych na tomto místě věnoval poděkování vedoucímu bakalářské práce doc. RNDr. Arnoštu Komárkovi, Ph.D. za to, že si vždy našel chvíli, aby mi poskytnul nesčetné množství odborných rad a připomínek, které dopomohly k finální podobě této práce.

Název práce: Bayesovská faktorová analýza

Autor: Bc. Jan Vávra

Katedra: Katedra pravděpodobnosti a matematické statistiky

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

Abstrakt: Faktorová analýza je metoda, která umožňuje náhodný vektor o vysokém počtu měření aproximovat pomocí lineárních kombinací mnohem nižšího počtu skrytých faktorů. Klasický odhad tohoto modelu spočívá ve volbě počtu faktorů, rozkladu varianční matice tak, aby byly dodrženy identifikační podmínky, a ve vhodném zvolení rotace pro lepší interpretaci modelu. Tento model převedeme do bayesovského pojetí, které navíc oproti klasickému nabízí využití apriorní informace. Vhodnou specifikací apriorního rozdělení lze počet skrytých faktorů považovat za náhodný parametr a lze vynutit závislost každého měření na nejvýše jediném faktoru. Odhady parametrů modelu jsou pak založeny na aposterioriálním rozdělení, které je aproximováno pomocí MCMC metod. Bayesovský pohled tak naráz řeší problematiku počtu faktorů, odhad modelu, zajištění identifikovatelnosti a interpretovatelnosti. Schopnost odhadovat skutečný počet skrytých faktorů je podrobena simulační studii.

Klíčová slova: bayesovská statistika, faktorová analýza, model určujících faktorů

Title: Bayesian factor analysis

Author: Bc. Jan Vávra

Department: Department of Probability and Mathematical Statistics

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

Abstract: Factor analysis is a method which enables high-dimensional random vector of measurements to be approximated by linear combinations of much lower number of hidden factors. Classical estimation procedure of this model lies on the choice of the number of factors, the decomposition of variance matrix while keeping identification conditions satisfied and on the appropriate choice of rotation for better interpretation of the model. This model will be transferred into bayesian framework which offers the usage of prior information unlike the classical approach. The number of hidden factors can be considered as a random parameter and the dependency of each measurement on at most one factor can be forced by suitable specification of prior distribution. Estimates of model parameters are based on posterior distribution which is approximated by Monte Carlo Markov Chain methods. Bayesian approach solves the problem of selection of the number of factors, the model estimation and the ensuring of the identifiability and the interpretability at the same time. The ability to estimate the real number of hidden factors is tested in a simulation study.

Keywords: Bayesian statistics, factor analysis, dedicated factor model

Obsah

Používané značení	3
Úvod	5
1 Pravděpodobnostní rozdělení	7
1.1 Normální rozdělení	7
1.2 Wishartovo rozdělení	9
1.3 Inverzní Wishartovo rozdělení	9
1.4 Studentovo rozdělení	10
1.5 Dirichletovo rozdělení	11
2 Bayesovské metody	15
2.1 Bayesova věta	15
2.2 Volba apriorního rozdělení	16
2.2.1 Neinformativní rozdělení	17
2.2.2 Konjugované rozdělení	18
2.2.3 Volba hyperparametrů	19
2.3 Metody odhadu	21
2.3.1 MCMC metody	22
2.3.2 Gibbsův algoritmus	24
2.3.3 ICM algoritmus	25
2.4 Bayesovská mnohorozměrná regrese	27
3 Bayesovská faktorová analýza	31
3.1 Faktorová analýza	31
3.1.1 Předpoklady a interpretace	32
3.1.2 Odhad modelu faktorové analýzy	34
3.2 Bayesovská specifikace	37
3.2.1 Apriorní rozdělení	37
3.2.2 Aposteriorní rozdělení	39
3.2.3 Marginalizace aposteriorního rozdělení	41
3.3 Gibbsův algoritmus	47
3.4 ICM algoritmus	48
3.5 Další volby apriorních rozdělení	50
3.5.1 Neinformativní apriorní rozdělení	50
3.5.2 Zobecněná apriorní rozdělení	52
3.6 Zobecnění bayesovské faktorové analýzy	56
3.6.1 Model s korelovanými faktory	56
3.6.2 Model se známými regresory	57
4 Bayesovský odhad počtu skrytých faktorů	61
4.1 Model určujících faktorů	61
4.2 Věrohodnost a apriorní rozdělení	64
4.2.1 Apriorní rozdělení indikátorů	65
4.2.2 Apriorní rozdělení faktorů	67
4.2.3 Shrnutí apriorního rozdělení	68

4.3	Aposteriorní rozdělení	69
4.4	MCMC algoritmus	75
4.4.1	Metoda marginálního doplnění dat	76
4.4.2	Základní Gibbsův algoritmus	77
4.4.3	Metropolisův-Hastingsův algoritmus	78
4.5	Bayesovský odhad parametrů a počtu aktivních faktorů K	80
4.6	Zobecnění bayesovského modelu určujících faktorů	81
5	Simulační studie	83
5.1	Popis scénáře a volba parametrů	83
5.2	Výsledky studie	85
	Závěr	93
	Apendix	95
A	Tvrzení z lineární algebry	95
B	Doplnění na čtverec	97
C	Věta o transformaci	101
D	Značení parametrů v modelu určujících faktorů ve funkci <code>befa</code>	102
	Seznam použité literatury	103
	Seznam obrázků	105
	Seznam tabulek	107

Používané značení

V této práci se budeme držet následujících konvencí:

- ▷ Malými písmeny budeme označovat jednorozměrné konstanty, či funkce. Písmena i, j, k, l, m, n budou označovat přirozená čísla.
- ▷ Velkými písmeny budeme označovat náhodné veličiny (X, Y, Z), jejich realizace příslušnými malými písmeny (x, y, z). Řecká písmena (θ, μ, σ) budou víceméně vyhrazena pro (náhodné) parametry (až na výjimky - ε).
- ▷ Tučnými písmeny budeme značit vektory. Každý vektor bude chápán jako sloupcový. Pro náhodné vektory budeme užívat velká římská písmena ($\mathbf{X}, \mathbf{Y}, \mathbf{Z}$) a jejich realizace příslušnými malými písmeny ($\mathbf{x}, \mathbf{y}, \mathbf{z}$).
- ▷ Pro matice budeme používat zdvojených velkých písmen ($\mathbb{X}, \mathbb{Y}, \mathbb{Z}$) nebo velkých tučných řeckých písmen ($\mathbf{\Sigma}, \mathbf{\Psi}, \mathbf{\Phi}$). Někdy pro zdůraznění, že se jedná o realizaci náhodné matice použijeme malých zdvojených písmen ($\mathbb{x}, \mathbb{y}, \mathbb{z}$). Symboly \mathbb{N} a \mathbb{R} máme vyhrazeny pro označení množiny přirozených a reálných čísel.
- ▷ Pro míry máme vyhrazený speciální font řeckých písmen (ν, λ).
- ▷ Bezpatkové písmo je používáno pro značení pravděpodobnostních rozdělání ($\mathbf{N}, \mathbf{t}, \mathbf{W}, \dots$) nebo určitých funkcionalů či operátorů ($\mathbf{E}, \mathbf{var}, \mathbf{Tr}, \dots$).

Speciální symboly:

$\text{diag}(a_1, \dots, a_n)$	diagonální matice s danými prvky na diagonále
$\text{diag}(\mathbb{A})$	diagonála matice \mathbb{A}
$ \mathbb{A} $	determinant matice \mathbb{A}
$\text{Tr}(\mathbb{A})$	stopa matice \mathbb{A}
$\text{vec } \mathbb{A}$	po sloupcích vektorizovaná matice \mathbb{A}
$\mathbb{A} \otimes \mathbb{B}$	Kroneckerův součin matic \mathbb{A} a \mathbb{B}
$\mathbf{x}^\top, \mathbb{X}^\top$	operátor transpozice vektoru \mathbf{x} a matice \mathbb{X}
\mathbb{A}^{-1}	inverzní matice k matici \mathbb{A}
$\mathbb{A} > 0$	pozitivně definitní matice \mathbb{A}
$\mathbb{A} \geq 0$	pozitivně semidefinitní matice \mathbb{A}
$\mathbf{1}_n$	n -složkový sloupcový vektor, jehož prvky jsou jen 1
\mathbb{I}_n	jednotková matice řádu n , neboli $\text{diag}(1, \dots, 1)$
$\mathbf{0}_n$	n -složkový sloupcový nulový vektor
$\mathbf{1}_A(a) = \mathbb{1}(a \in A)$	indikátorová funkce, 1 pokud $a \in A$, 0 jinak
$\mathbb{O}_{n \times m}$	nulová matice rozměrů $n \times m$
$\stackrel{!}{=}$	výraz vlevo polož roven výrazu vpravo a vyřeš soustavu
$:=$	výrazu vlevo přiřaď hodnotu napravo
$\lfloor a \rfloor$	dolní celá část čísla a
$\lceil a \rceil$	horní celá část čísla a
Γ	gama funkce, případně označení gama rozdělení
$B(\alpha, \beta)$	beta funkce
$D_{\{x\}}$	Diracovo degenerované rozdělení v bodě x
P	pravděpodobnostní míra
$p(\mathbf{x} \mathbf{y})$	hustota rozdělení $\mathbf{X} \mathbf{Y}$
$E \mathbf{X}$	střední hodnota náhodného vektoru \mathbf{X}
$\text{var } \mathbf{X}$	rozptylová matice náhodného vektoru \mathbf{X}
$\text{corr } \mathbf{X}$	korelační matice náhodného vektoru \mathbf{X}
$\text{cov}(\mathbf{X}, \mathbf{Y})$	kovariance náhodných vektorů \mathbf{X} a \mathbf{Y}
$\text{corr}(\mathbf{X}, \mathbf{Y})$	korelace náhodných vektorů \mathbf{X} a \mathbf{Y}
$\text{mod } \mathbf{X}$	modus náhodného vektoru \mathbf{X}

Úvod

Faktorová analýza je metoda spočívající v rozkladu m -rozměrného náhodného vektoru \mathbf{Y} na součet tří jiných vektorů:

$$\underbrace{\mathbf{Y}}_{m \times 1} = \underbrace{\boldsymbol{\mu}}_{m \times 1} + \underbrace{\boldsymbol{\Lambda}}_{m \times k} \underbrace{\mathbf{F}}_{k \times 1} + \underbrace{\boldsymbol{\varepsilon}}_{m \times 1}, \quad (1)$$

kde $\boldsymbol{\mu} = \mathbf{E} \mathbf{Y}$ je vektor střední hodnoty \mathbf{Y} , $\boldsymbol{\Lambda}$ je $m \times k$ matice faktorových zátěží (tzv. „loadings“), \mathbf{F} je k -rozměrný náhodný vektor latentních proměnných (tzv. faktorů) a $\boldsymbol{\varepsilon}$ je náhodný vektor vyjadřující náhodnou odchylku vektoru \mathbf{Y} od střední hodnoty a vlivu faktorů. Podstatný je zde předpoklad $m > k$, tedy se snažíme vysvětlit náhodný vektor \mathbf{Y} pomocí menšího počtu skrytých náhodných faktorů \mathbf{F} .

K dispozici máme realizace náhodných vektorů \mathbf{Y} a úkolem je odhadnout neznámé hodnoty $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}$ i hodnoty faktorů příslušné daným realizacím. Aby bylo možné provést tento rozklad jednoznačně, je třeba zavést několikero předpokladů: budeme vyžadovat, aby vektory \mathbf{F} a $\boldsymbol{\varepsilon}$ byly centrované (pro jednoznačnost parametru $\boldsymbol{\mu}$), vzájemně nekorelované a aby splňovaly jisté požadavky na varianční strukturu (podrobněji viz kapitola 3). I po zavedení těchto omezujících podmínek lze stále vektor \mathbf{Y} rozložit vícero způsoby, což často vede k různým interpretacím výsledků. Proto použití faktorové analýzy vede k rozsáhlým diskuzím nad obdrženými výsledky.

Běžné metody pro odhad modelu faktorové analýzy spočívají ve vhodném rozkladu varianční matice náhodného vektoru \mathbf{Y} . Bayesovský přístup k tomuto problému se zaměřuje na stochastickou strukturu modelu - na samotný princip vzniku pozorovaného vektoru \mathbf{Y} . Dále nabízí využití nejen informace skryté v datech, ale také dostupné apriorní informace (v podobě historických dat nebo naší představy o struktuře latentních faktorů). V poslední době se navíc podařilo využít bayesovského přístupu k elegantnímu řešení problému správné volby počtu faktorů k , který v reálné situaci není předem znám.

Bayesovský přístup vyžaduje znalost celé řady pravděpodobnostních rozdělení. Vzhledem k časté nejednoznačnosti jejich parametrizace si v první kapitole tato rozdělení řádně zdefinujeme a vysvětlíme si význam jednotlivých parametrů. Základní principy bayesovské statistiky (zvláště pak metody odhadu parametrů) si pro přehlednost shrneme ve druhé části. Tyto metody názorně ilustrujeme na modelu mnohorozměrné lineární regrese, který je v základní struktuře velmi podobný modelu faktorové analýzy.

Posléze si podrobněji představíme model faktorové analýzy a klasické metody jeho odhadu. Znázorněním parametrů jej převedeme na model bayesovské faktorové analýzy. Ukážeme si, že ani při ideálně nastavených apriorních rozděleních nelze v marginálních aposteriorních rozděleních jednotlivých parametrů nalézt běžně používaná pravděpodobnostní rozdělení. Dokážeme však určit plně podmíněná rozdělení jednotlivých parametrů a na jejich základě zkonstruovat Gibbsův či ICM algoritmus pro odhad parametrů. Plně podmíněná rozdělení také odvodíme za různých zobrazení modelu faktorové analýzy.

Ve čtvrté kapitole si představíme model určujících faktorů (*dedicated factor model*) vycházející z modelu faktorové analýzy, kde každá složka měření je ovliv-

něna nejvýše jedním faktorem. Conti a kol. (2014) tohoto modelu používají právě k bayesovskému odhadu počtu faktorů. Představíme si lehce zjednodušený pravděpodobnostní mechanismus, na kterém tento model stojí. S využitím metod používaných v předchozích kapitolách odvodíme plně podmíněná rozdělení skupin parametrů a sestrojíme sofistikovaný algoritmus, s jehož pomocí lze odhadnout parametry modelu. Model je konstruován tak, aby rozklad \mathbf{Y} byl jednoznačný až na triviální rotace. Dostáváme tak model s řídkou maticí faktorových zátěží, který lze jednoduše interpretovat. V následné simulační studii prověříme, jak si tato metoda vede v úloze odhadu počtu faktorů ve srovnání s klasickými metodami. Podobná studie již byla provedena ve výše zmíněném článku, zde se však navíc zaměříme na situace, kdy platí obecný model faktorové analýzy a nikoliv předpokládaný model určujících faktorů.

V práci používáme celou řadu poznatků z lineární algebry jako je například metoda doplňování na čtverec. Ta hraje důležitou roli v odvozování plně podmíněných rozdělení. Věnujeme jí proto několik stránek v apendixu zařazeném na konci práce. Pro přehlednost zde přikládáme i další užitečná tvrzení pro práci s maticemi.

1. Pravděpodobnostní rozdělení

Při používání bayesovských metod jsou neznámé parametry modelu považovány za náhodné a budeme jim přisuzovat různá pravděpodobnostní rozdělení. Jelikož volby parametrizace jednotlivých rozdělení nejsou totožné ve všech matematických pramenech, tak zde pro přehlednost uvádíme přehled hustot a dalších užitečných charakteristik těch rozdělení, kterých budeme nejčastěji používat. Pro zvolené parametrizace zavedeme značení jednotlivých rozdělení, které budeme nadále v textu používat. Tato kapitola tedy čtenáři znalému různých rozdělení nepřinese nic nového, ovšem může ho však ujistit ve významech jednotlivých parametrů.

Nejprve si však zavedeme velmi užitečný pojem proporcionality dvou funkcí.

Definice 1. Řekneme, že funkce p dvou vstupních proměnných \mathbf{x} a ζ je proporcionalní funkci f (stejných proměnných) vzhledem k ζ , jestliže existuje multiplikační konstanta $c(\zeta)$ taková, že

$$p(\mathbf{x}|\zeta) = c(\zeta) \cdot f(\mathbf{x}, \zeta).$$

Budeme zkráceně zapisovat, že $p(\mathbf{x}|\zeta) \propto f(\mathbf{x}, \zeta)$

Použití této proporcionality vzhledem k parametru bude názorně ilustrováno u hustot níže vypsanych pravděpodobnostních rozdělení a bude hojně využíváno v dalších částech práce. Hustota rozdělení $\mathbb{X}|\zeta$ bude značena $p(\mathbf{x}|\zeta)$ a pro všechna uvažovaná rozdělení v této práci se jedná o hustotu vzhledem k patřičně rozměrné Lebesgueově míře λ .

1.1 Normální rozdělení

Jednorozměrné normální rozdělení

Jednorozměrné normální rozdělení veličiny X se střední hodnotou $\mu \in \mathbb{R}$ a rozptylem $0 < \sigma^2 < \infty$ definujeme pomocí hustoty

$$p(x|\mu, \sigma^2) = (2\pi)^{-\frac{1}{2}} (\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \propto \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}. \quad (1.1)$$

Onou konstantou při proporcionalitě \propto je $c_{N,1}(\sigma^2) = (2\pi\sigma^2)^{-\frac{1}{2}}$. Zkráceně zapisujeme $X|\mu, \sigma^2 \sim N(\mu, \sigma^2)$. Pro $Y \sim N(0, 1)$ platí $X = \sigma Y + \mu \sim N(\mu, \sigma^2)$. Pro generování z obecného jednorozměrného rozdělení tedy stačí umět generovat z normovaného normálního rozdělení $N(0, 1)$.

Tvrzení 1. Jsou-li U_1 a U_2 nezávislé náhodné veličiny rovnoměrně rozdělené na intervalu $(0,1)$, pak veličiny

$$X_1 = \sqrt{-2 \log U_1} \cos(2\pi U_2), \quad X_2 = \sqrt{-2 \log U_1} \sin(2\pi U_2)$$

jsou nezávislé veličiny s rozdělením $N(0, 1)$.

Důkaz. Snadno ověříme pomocí věty o transformaci C9 uvedené v apendixu. □

Vektorové normální rozdělení

Vektorové k -dimenzionální normální rozdělení \mathbf{X} se střední hodnotou $\boldsymbol{\mu} \in \mathbb{R}^k$ a rozptylovou maticí $\boldsymbol{\Sigma} > 0$ (pozitivně definitní matice typu $k \times k$) je definováno pomocí hustoty

$$\begin{aligned} p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= (2\pi)^{-\frac{k}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\} \\ &\propto \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}. \end{aligned} \quad (1.2)$$

Onou konstantou při proporcionalitě \propto je $c_{\mathbf{N},k}(\boldsymbol{\Sigma}) = (2\pi)^{-\frac{k}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}}$. Zkráceně zapisujeme $\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathbf{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Vektor $\mathbf{Y} \sim \mathbf{N}_k(\mathbf{0}_k, \mathbb{I}_k)$ má jednotlivé složky nezávislé a $\mathbf{N}(0, 1)$ rozdělené a platí $\mathbf{X} = \boldsymbol{\Sigma}^{\frac{1}{2}} \mathbf{Y} + \boldsymbol{\mu} \sim \mathbf{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Často se setkáváme s tím, že vektor \mathbf{X} se skládá ze dvou podvektorů \mathbf{X}_1 a \mathbf{X}_2 o délkách k_1 a k_2 . Rozdělme parametry $\boldsymbol{\mu}$ a $\boldsymbol{\Sigma}$ na příslušně rozměrné podvektory či podmatice tak, aby platilo

$$\begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \sim \mathbf{N}_{k_1+k_2} \left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right). \quad (1.3)$$

Potom pro podmíněné rozdělení vektoru \mathbf{X}_1 při $\mathbf{X}_2 = \mathbf{x}_2$ platí

$$\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2 \sim \mathbf{N}_{k_1} \left(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{x}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \right). \quad (1.4)$$

Maticové normální rozdělení

Dále mějme n náhodných vektorů $\mathbf{X}_i \sim \mathbf{N}_k(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$, definujme si matice $\mathbb{X}^\top = (\mathbf{X}_1, \dots, \mathbf{X}_n)$ a $\mathbb{M}^\top = (\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_n)$ a dále buď $\boldsymbol{\Phi} > 0$ matice typu $n \times n$ udávající kovarianční strukturu mezi jednotlivými vektory $\mathbf{X}_i, i = 1, \dots, n$. Potom \mathbb{X} se řídí maticovým normálním rozdělením se střední hodnotou \mathbb{M} , zapisujeme $\mathbb{X}|\mathbb{M}, \boldsymbol{\Phi}, \boldsymbol{\Sigma} \sim \mathbf{N}_{n \times k}(\mathbb{M}, \boldsymbol{\Phi} \otimes \boldsymbol{\Sigma})$. Matice $\boldsymbol{\Phi} \otimes \boldsymbol{\Sigma}$ je kovarianční maticí po řádcích vektorizované matice \mathbb{X} , tj. $\mathbf{X} = \text{vec}(\mathbb{X}^\top) \sim \mathbf{N}_{nk}(\boldsymbol{\mu}, \boldsymbol{\Phi} \otimes \boldsymbol{\Sigma})$, kde $\boldsymbol{\mu} = \text{vec}(\mathbb{M}^\top)$. Proto hustotou tohoto rozdělení je funkce

$$\begin{aligned} p(\mathbb{X}|\mathbb{M}, \boldsymbol{\Phi}, \boldsymbol{\Sigma}) &= (2\pi)^{-\frac{nk}{2}} |\boldsymbol{\Phi}|^{-\frac{k}{2}} |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top [\boldsymbol{\Phi} \otimes \boldsymbol{\Sigma}]^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \text{Tr} \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Phi}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}, \end{aligned} \quad (1.5)$$

kde multiplikatívní konstantou je zde $c_{\mathbf{N},nk}(\boldsymbol{\Phi}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{nk}{2}} |\boldsymbol{\Phi}|^{-\frac{k}{2}} |\boldsymbol{\Sigma}|^{-\frac{n}{2}}$. Díky vlastnostem Kroneckerova součinu, stopy matice a vektorizace (tvrzení A3 a A4) platí

$$\text{vec}(\mathbb{X}^\top) \sim \mathbf{N}_{nk}(\text{vec}(\mathbb{M}^\top), \boldsymbol{\Phi} \otimes \boldsymbol{\Sigma}) \iff \text{vec}(\mathbb{X}) \sim \mathbf{N}_{nk}(\text{vec}(\mathbb{M}), \boldsymbol{\Sigma} \otimes \boldsymbol{\Phi}).$$

Je-li náhodná matice \mathbb{Y} typu $n \times k$ tvořena nezávislými standardizovaně normálně rozdělenými prvky, tak $\mathbb{X} = \mathbb{A} \mathbb{Y} \mathbb{B}^\top + \mathbb{M} \sim \mathbf{N}_{n \times k}(\mathbb{M}, \mathbb{A} \mathbb{A}^\top \otimes \mathbb{B} \mathbb{B}^\top)$, kde matice \mathbb{A} a \mathbb{B} jsou čtvercové řádů n a k . Potřebné matice \mathbb{A} a \mathbb{B} získáme například Choleského rozkladem.

1.2 Wishartovo rozdělení

Náhodná veličina $G > 0$ se řídí jednorozměrným Wishartovým rozdělením s parametrem měřítka $v^2 > 0$ o $\nu_0 > 0$ stupních volnosti, píšeme $G|v^2, \nu_0 \sim \mathbf{W}_1(v^2, \nu_0)$, jestliže její hustota je

$$\begin{aligned} p(g|v^2, \nu_0) &= \frac{(2v^2)^{-\frac{\nu_0}{2}}}{\Gamma\left(\frac{\nu_0}{2}\right)} g^{\frac{\nu_0}{2}-1} \exp\left\{-\frac{g}{2v^2}\right\} \\ &\propto g^{\frac{\nu_0}{2}-1} \exp\left\{-\frac{g}{2v^2}\right\}. \end{aligned} \quad (1.6)$$

Vidíme, že se v podstatě jedná o známé gama rozdělení s parametrem tvaru $\frac{\nu_0}{2}$ a parametrem rychlosti $(2v^2)^{-1}$, tj. $\Gamma\left(\frac{\nu_0}{2}, (2v^2)^{-1}\right)$. Tedy střední hodnota je $\nu_0 v^2$, modus $(\nu_0 - 2)v^2$ (jen je-li $\nu_0 > 2$) a rozptyl je $2\nu_0 v^4$. Je-li parametr $v^2 = 1$, tak se jedná o klasické chí-kvadrát rozdělení o ν_0 stupních volnosti, tj. $\chi_{\nu_0}^2$. Jsou-li pro $\nu_0 \in \mathbb{N}$ veličiny X_1, \dots, X_{ν_0} nezávislé a $\mathbf{N}(0, v^2)$ rozdělené, pak $G = X_1^2 + \dots + X_{\nu_0}^2 \sim \mathbf{W}_1(v^2, \nu_0)$.

V tomto duchu se dá Wishartovo rozdělení zobecnit na maticovou verzi. Jsou-li $\mathbf{X}_1, \dots, \mathbf{X}_{\nu_0}$ nezávislé $\mathbf{N}_k(\boldsymbol{\mu}, \boldsymbol{\Upsilon})$ rozdělené, potom pozitivně definitní matice $\mathbb{G} = (\mathbf{X}_1 - \boldsymbol{\mu})(\mathbf{X}_1 - \boldsymbol{\mu})^\top + \dots + (\mathbf{X}_{\nu_0} - \boldsymbol{\mu})(\mathbf{X}_{\nu_0} - \boldsymbol{\mu})^\top$ má k -rozměrné maticové Wishartovo rozdělení s měřítkovou maticí $\boldsymbol{\Upsilon} > 0$ o ν_0 stupních volnosti, píšeme $\mathbf{W}_k(\boldsymbol{\Upsilon}, \nu_0)$. V řeči maticového normálního rozdělení platí, že pokud $\mathbb{X} \sim \mathbf{N}_{\nu_0 \times k}(\mathbb{M}, \mathbb{I}_{\nu_0} \otimes \boldsymbol{\Upsilon})$, tak $\mathbb{G} = (\mathbb{X} - \mathbb{M})^\top (\mathbb{X} - \mathbb{M}) \sim \mathbf{W}_k(\boldsymbol{\Upsilon}, \nu_0)$. Hustota tohoto rozdělení na množině pozitivně definitních matic řádu k je (pouze pro $\nu_0 > k - 1$) tvaru

$$\begin{aligned} p(\mathbb{g}|\boldsymbol{\Upsilon}, \nu_0) &= c_W(k, \nu_0) \cdot |\boldsymbol{\Upsilon}|^{-\frac{\nu_0}{2}} \cdot |\mathbb{g}|^{\frac{\nu_0-k-1}{2}} \exp\left\{-\frac{1}{2} \text{Tr } \boldsymbol{\Upsilon}^{-1} \mathbb{g}\right\} \\ p(\mathbb{g}|\boldsymbol{\Upsilon}) &\propto |\mathbb{g}|^{\frac{\nu_0-k-1}{2}} \exp\left\{-\frac{1}{2} \text{Tr } \boldsymbol{\Upsilon}^{-1} \mathbb{g}\right\}, \end{aligned} \quad (1.7)$$

kde $c_W(k, \nu_0)^{-1} = 2^{\frac{\nu_0 k}{2}} \pi^{\frac{k(k-1)}{4}} \prod_{j=1}^k \Gamma\left(\frac{\nu_0+1-j}{2}\right)$. Tuto konstantu často skrýváme do proporcionality (1.7), i když stále závisí na parametru ν_0 . Opět nic nebrání tomu, aby toto rozdělení bylo definováno nejen pro přirozené ν_0 , ale i obecné $\nu_0 > k - 1$. Střední hodnotou je $\nu_0 \boldsymbol{\Upsilon}$, modus je $(\nu_0 - k - 1)\boldsymbol{\Upsilon}$ (jen pokud $\nu_0 > k + 1$). Rozptylová matice $\text{vec } \mathbb{G}$ lze zapsat jako $\nu_0(\mathbb{I}_{k^2} + \mathbb{P}_{\text{vec } \mathbb{G}})(\boldsymbol{\Upsilon} \otimes \boldsymbol{\Upsilon})$, kde $\mathbb{P}_{\text{vec } \mathbb{G}}$ je transpozičně-permutační matice k matici \mathbb{G} , definice viz appendix A. Kovariance mezi diagonálními prvky je tedy $\text{cov}(G_{ii}, G_{jj}) = 2\nu_0 \Upsilon_{ij}$

1.3 Inverzní Wishartovo rozdělení

Pro naše potřeby ale bude daleko užitečnější rozdělení inverzní k Wishartovu rozdělení, tedy rozdělení, jakým se řídí inverzní matice (převrácená hodnota v jednorozměrném případě) k matici řídící se Wishartovým rozdělením. Formálně

$$\begin{aligned} \sigma^2|q, \nu &\sim \text{IW}_1(q, \nu) \iff \sigma^{-2}|q^{-1}, \nu_0 \sim \mathbf{W}_1(q^{-1}, \nu_0), \\ \Sigma|Q, \nu &\sim \text{IW}_k(Q, \nu) \iff \Sigma^{-1}|Q^{-1}, \nu_0 \sim \mathbf{W}_k(Q^{-1}, \nu_0), \end{aligned}$$

kde jsme zavedli $\nu = \nu_0 + k + 1$, se kterým se bude lépe pracovat. Počtem stupňů volnosti je v literatuře klasicky označováno ν_0 , zde však budeme nadále pracovat s ν , neboť v samotné hustotě tohoto rozdělení je význam ν přehlednější. Hustota inverzního Wishartova rozdělení pro $\nu_0 > k - 1$ (tedy $\nu > 2k$) obdržena užitím věty o transformaci C9 s jakobiánem $|\Sigma|^{-(k+1)}$ má tvar

$$\begin{aligned} p(\sigma^2 | q, \nu) &= \frac{1}{\Gamma\left(\frac{\nu-2}{2}\right)} \left(\frac{q}{2}\right)^{\frac{\nu-2}{2}} (\sigma^2)^{-\frac{\nu}{2}} \exp\left\{-\frac{q}{2\sigma^2}\right\} \\ &\propto (\sigma^2)^{-\frac{\nu}{2}} \exp\left\{-\frac{q}{2\sigma^2}\right\}, \end{aligned} \quad (1.8)$$

$$\begin{aligned} p(\Sigma | \mathbb{Q}, \nu) &= c_{IW}(k, \nu) \cdot |\mathbb{Q}|^{\frac{\nu-k-1}{2}} \cdot |\Sigma|^{-\frac{\nu}{2}} \exp\left\{-\frac{1}{2} \text{Tr} \Sigma^{-1} \mathbb{Q}\right\} \\ &\propto |\Sigma|^{-\frac{\nu}{2}} \exp\left\{-\frac{1}{2} \text{Tr} \Sigma^{-1} \mathbb{Q}\right\}, \end{aligned} \quad (1.9)$$

kde $c_{IW}(k, \nu) = c_W(k, \nu_0)$. Tato konstanta závisí na parametru ν , avšak bude praktičtější ji skrýt do proporcionalit (1.8) a (1.9). Střední hodnota je pro $\nu > 2k + 2$ rovna $\mathbb{Q}/(\nu - 2k - 2)$, modus je potom \mathbb{Q}/ν . Pro $\nu > 2k + 4$ je rozptyl i -tého diagonálního prvku matice je roven $2q_{ii}^2/(\nu - 2k - 2)^2(\nu - 2k - 4)$.

1.4 Studentovo rozdělení

Jednorozměrné Studentovo rozdělení

Buďte $X \sim \mathbf{N}(\mu, \sigma^2)$ a $G \sim W_1(1, \nu) \equiv \chi_\nu^2$ nezávislé náhodné veličiny. Potom náhodná veličina

$$T = \frac{X - \mu}{\sqrt{\frac{G}{\nu}}} + t_0$$

se řídí Studentovým t-rozdělením o ν stupních volnosti s parametrem polohy t_0 a parametrem rozptylu σ^2 , píšeme $T \sim \mathbf{t}(\nu, t_0, \sigma^2)$. Nejedná se však o tzv. *necentrální* t-rozdělení. Pomocí věty o transformaci C9 lze ukázat, že hustota tohoto rozdělení je

$$p(t | \nu, t_0, \sigma^2) = c_t(\nu) \frac{1}{\sigma} \left[1 + \frac{1}{\nu\sigma^2}(t - t_0)^2\right]^{-\frac{\nu+1}{2}} \propto \left[1 + \frac{1}{\nu\sigma^2}(t - t_0)^2\right]^{-\frac{\nu+1}{2}}, \quad (1.10)$$

kde $c_t(\nu) = \Gamma\left(\frac{\nu+1}{2}\right) / \Gamma\left(\frac{\nu}{2}\right) \sqrt{\pi\nu}$. Je-li $t_0 = 0$ a $\sigma^2 = 1$, pak se jedná o standardizované t-rozdělení s ν stupni volnosti.

Vektorové Studentovo rozdělení

Buďte $\mathbf{X} \sim \mathbf{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ a $\mathbb{G} \sim W_k(\mathbb{I}_k, \nu)$ nezávislé. Potom náhodný vektor

$$\mathbf{T} = \left(\frac{\mathbb{G}}{\nu}\right)^{-\frac{1}{2}} (\mathbf{X} - \boldsymbol{\mu}) + \mathbf{t}_0$$

se řídí k - rozměrným Studentovým t-rozdělením o ν stupních volnosti s parametrem polohy \mathbf{t}_0 a rozptylovou maticí $\boldsymbol{\Sigma}$, píšeme $\mathbf{T} \sim \mathbf{t}_k(\nu, \mathbf{t}_0, \boldsymbol{\Sigma})$. Pomocí věty

o transformaci C9 lze ukázat, že hustota tohoto rozdělení je

$$\begin{aligned} p(\mathbf{t}|\nu, \mathbf{t}_0, \Sigma) &= c_t(\nu) |\Sigma|^{-\frac{1}{2}} \left[1 + \frac{1}{\nu} (\mathbf{t} - \mathbf{t}_0)^\top \Sigma^{-1} (\mathbf{t} - \mathbf{t}_0) \right]^{-\frac{\nu+k}{2}} \\ &\propto \left[1 + \frac{1}{\nu} (\mathbf{t} - \mathbf{t}_0)^\top \Sigma^{-1} (\mathbf{t} - \mathbf{t}_0) \right]^{-\frac{\nu+k}{2}}, \end{aligned} \quad (1.11)$$

kde $c_t(\nu) = \Gamma\left(\frac{\nu+k}{2}\right) / \Gamma\left(\frac{\nu}{2}\right) (\pi\nu)^{\frac{k}{2}}$.

Maticové Studentovo rozdělení

Budte $\mathbb{X} \sim \mathbf{N}_{n \times k}(\mathbb{M}, \Phi \otimes \Sigma)$ a $\mathbb{G} \sim \mathbf{W}_k(\mathbb{I}_k, \nu)$ nezávislé. Potom náhodná matice

$$\mathbb{T} = (\mathbb{X} - \mathbb{M}) \left(\frac{\mathbb{G}}{\nu} \right)^{-\frac{1}{2}} + \mathbb{T}_0$$

se řídí $(n \times k)$ -rozměrným Studentovým \mathbf{t} -rozdělením o ν stupních volnosti s parametrem polohy \mathbb{T}_0 a rozptylovou maticí $\Phi \otimes \Sigma$, píšeme $\mathbb{T} \sim \mathbf{t}_{n \times k}(\nu, \mathbb{T}_0, \Phi \otimes \Sigma)$. Pomocí věty o transformaci C9 lze ukázat, že hustota tohoto rozdělení je

$$\begin{aligned} p(\mathbf{t}|\nu, \mathbb{T}_0, \Phi, \Sigma) &= c_{\mathbb{T}}(\nu) |\Phi|^{\frac{k}{2}} |\Sigma|^{-\frac{n}{2}} \left| \Phi + \frac{1}{\nu} (\mathbf{t} - \mathbb{T}_0) \Sigma^{-1} (\mathbf{t} - \mathbb{T}_0)^\top \right|^{-\frac{\nu+k}{2}} \\ &\propto \left| \Phi + \frac{1}{\nu} (\mathbf{t} - \mathbb{T}_0) \Sigma^{-1} (\mathbf{t} - \mathbb{T}_0)^\top \right|^{-\frac{\nu+k}{2}}, \end{aligned} \quad (1.12)$$

kde $c_{\mathbb{T}}(\nu) = \prod_{j=1}^n \Gamma\left(\frac{\nu+k+1-j}{2}\right) / \prod_{j=1}^n \Gamma\left(\frac{\nu+1-j}{2}\right) (\pi\nu)^{\frac{nk}{2}}$.

Pro Studentova \mathbf{t} -rozdělení existuje střední hodnota pouze pro $\nu > 1$ a je rovná parametru polohy, který je ze symetrie roven modu tohoto rozdělení. Rozptyl je pak konečný pro $\nu > 2$ je roven rozptylovému parametru přenásobenému $\nu/(\nu - 2)$. V případě maticového \mathbb{T} mluvíme o rozptylové matici vektoru $\text{vec}(\mathbb{T}^\top)$.

1.5 Dirichletovo rozdělení

Tohoto rozdělení se často používá k apriorní specifikaci pravděpodobností. V jednorozměrném případě se jedná o známé beta rozdělení.

Beta rozdělení

Náhodná veličina $X \in (0,1)$ se řídí beta rozdělením s parametry $\alpha, \beta > 0$, píšeme $X \sim \text{Beta}(\alpha, \beta)$, je-li její hustota tvaru

$$p(x|\alpha, \beta) = \frac{1}{\mathbf{B}(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} \propto x^{\alpha-1} (1-x)^{\beta-1}, \quad (1.13)$$

kde $\mathbf{B}(\alpha, \beta)$ je beta funkce definovaná jako $\mathbf{B}(\alpha, \beta) = \Gamma(\alpha) \cdot \Gamma(\beta) / \Gamma(\alpha + \beta)$. Střední hodnota této veličiny je $\mathbf{E} X = \alpha / (\alpha + \beta)$ a rozptyl

$$\text{var } X = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} = \frac{1}{\alpha + \beta + 1} \mathbf{E} X (1 - \mathbf{E} X).$$

Modus pro $\alpha, \beta > 1$ (nejvýše jeden z nich může být i roven 1) má hodnotu $(\alpha - 1) / (\alpha + \beta - 2)$. Příklad $\alpha = \beta = 1$ odpovídá rovnoměrnému rozdělení na intervalu $(0,1)$.

Dirichletovo rozdělení

Náhodný vektor $\mathbf{X} = (X_1, \dots, X_k)^\top$, jehož složky jsou nezáporné a jejich součet je 1, tj. $X_j \geq 0$ a $\sum_{j=1}^k X_j = 1$, se řídí k -rozměrným Dirichletovým rozdělením s kladnými parametry $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_k)^\top$, píšeme $\mathbf{X} \sim \text{Dir}_k(\boldsymbol{\alpha})$, pokud jeho hustota je tvaru

$$p(\mathbf{x}|\boldsymbol{\alpha}) = \frac{\prod_{j=1}^k \Gamma(\alpha_j)}{\Gamma\left(\sum_{j=1}^k \alpha_j\right)} \cdot \prod_{j=1}^k x_j^{\alpha_j-1} \propto \prod_{j=1}^k x_j^{\alpha_j-1}. \quad (1.14)$$

Označíme-li $\alpha_0 = \sum_{j=1}^k \alpha_j$, pak střední hodnota vektoru \mathbf{X} má tvar $\mathbf{E} \mathbf{X} = \frac{\boldsymbol{\alpha}}{\alpha_0}$.

Rozptylová matice má tvar

$$\text{var} \mathbf{X} = \frac{\text{diag}(\alpha_0 \boldsymbol{\alpha}) - \boldsymbol{\alpha} \boldsymbol{\alpha}^\top}{\alpha_0^2 (\alpha_0 + 1)},$$

tedy rozptyl X_j je roven $\alpha_j(\alpha_0 - \alpha_j)/\alpha_0^2(\alpha_0 + 1)$ a jednotlivé složky jsou mezi sebou negativně korelované: $\text{corr}(X_i, X_j) = -\sqrt{\alpha_i \alpha_j} / \sqrt{(\alpha_0 - \alpha_i)(\alpha_0 - \alpha_j)}$ pro $i \neq j$. Modus tohoto rozdělení je $(\boldsymbol{\alpha} - \mathbf{1}_k)/(\alpha_0 - k)$, pokud $\alpha_j > 1$ pro každé $j \in \{1, \dots, k\}$.

Vidíme, že beta rozdělení je vlastně rozdělení složky X_1 dvourozměrného vektoru $\mathbf{X} = (X_1, X_2)^\top \sim \text{Dir}_k(\boldsymbol{\alpha}, \beta)$. Obecně platí, že pokud $\mathbf{X} \sim \text{Dir}_k(\boldsymbol{\alpha})$, pak j -tá složka $X_j \sim \text{Beta}\left(\alpha_j, \sum_{i=1, i \neq j}^k \alpha_i\right)$.

Na závěr zde ještě uvádíme tabulku 1.1, která obsahuje přehled středních hodnot, modů a rozptylů všech představených rozdělání.

Tabulka 1.1: Základní charakteristiky pravděpodobnostních rozdělání představených v kapitole 1.

Rozdělání	E	mod	var*
$\mathbf{N}(\mu, \sigma^2)$	μ	μ	σ^2
$\mathbf{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	$\boldsymbol{\mu}$	$\boldsymbol{\mu}$	$\boldsymbol{\Sigma}$
$\mathbf{N}_{n \times k}(\mathbf{M}, \boldsymbol{\Phi} \otimes \boldsymbol{\Sigma})$	\mathbf{M}	\mathbf{M}	$\boldsymbol{\Phi} \otimes \boldsymbol{\Sigma}$
$\mathbf{W}_1(v^2, \nu_0)$	$\nu_0 v^2$	$(\nu_0 - 2)v^2$ pro $\nu_0 > 2$	$2\nu_0 v^4$
$\mathbf{W}_k(\boldsymbol{\Upsilon}, \nu_0)$	$\nu_0 \boldsymbol{\Upsilon}$	$(\nu_0 - k - 1)\boldsymbol{\Upsilon}$ pro $\nu_0 > k + 1$	$\nu_0(\mathbb{I}_{k^2} + \mathbb{P}_{\text{vec W}})(\boldsymbol{\Upsilon} \otimes \boldsymbol{\Upsilon})$
$\mathbf{IW}_1(q, \nu)$	$\frac{q}{\nu - 4}$ pro $\nu > 4$	$\frac{q}{\nu}$	$\frac{2q^2}{(\nu - 4)^2(\nu - 6)}$ pro $\nu > 6$
$\mathbf{IW}_k(\mathbb{Q}, \nu)$	$\frac{\mathbb{Q}}{\nu - 2k - 2}$ pro $\nu > 2k + 2$	$\frac{\mathbb{Q}}{\nu}$	\dagger pro $\nu > 2k + 4$
$\mathbf{t}(\nu, t_0, \sigma^2)$	t_0 pro $\nu > 1$	t_0	$\frac{\nu}{\nu - 2}\sigma^2$ pro $\nu > 2$
$\mathbf{t}_k(\nu, \mathbf{t}_0, \boldsymbol{\Sigma})$	\mathbf{t}_0 pro $\nu > 1$	\mathbf{t}_0	$\frac{\nu}{\nu - 2}\boldsymbol{\Sigma}$ pro $\nu > 2$
$\mathbf{t}_{n \times k}(\nu, \mathbb{T}_0, \boldsymbol{\Phi} \otimes \boldsymbol{\Sigma})$	\mathbb{T}_0 pro $\nu > 1$	\mathbb{T}_0	$\frac{\nu}{\nu - 2}\boldsymbol{\Phi} \otimes \boldsymbol{\Sigma}$ pro $\nu > 2$
$\mathbf{Beta}(\alpha, \beta)$	$\frac{\alpha}{\alpha + \beta}$	$\frac{\alpha - 1}{\alpha + \beta - 2}$	$\frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$
$\mathbf{Dir}_k(\boldsymbol{\alpha})$	$\frac{\boldsymbol{\alpha}}{\alpha_0}$ $\alpha_0 = \sum_{j=1}^k \alpha_j$	$\frac{\boldsymbol{\alpha} - \mathbf{1}_k}{\alpha_0 - k}$ pro $\alpha, \beta > 1$	$\frac{\text{diag}(\alpha_0 \boldsymbol{\alpha}) - \boldsymbol{\alpha} \boldsymbol{\alpha}^\top}{\alpha_0^2(\alpha_0 + 1)}$

* Pro maticová rozdělání se jedná o rozptyl vektorizovaného rozdělání s výjimkou †.

$$\dagger = \text{cov}(\sigma_{ij}, \sigma_{lm} | \mathbb{Q}, \nu) = \frac{2}{(\nu - 2k - 1)(\nu - 2k - 2)(\nu - 2k - 4)} q_{ij} q_{lm} + q_{il} q_{jm} + q_{im} q_{lj}$$

2. Bayesovské metody

V této kapitole se seznámíme s bayesovským přístupem k modelování pravděpodobnostního mechanismu, který stojí za dostupnými daty. Definujeme si základní pojmy a postupy, se kterými budeme nadále pracovat. Představíme si základní metody odhadu a následné inference o parametrech zájmu. V poslední části této kapitoly aplikujeme představené postupy na model bayesovské mnohorozměrné regrese, který má v jistém smyslu velmi blízko k modelu bayesovské faktorové analýzy, kterému se budeme věnovat později.

2.1 Bayesova věta

Nechť \mathbb{Y} reprezentuje data, která byla vygenerována nějakým pravděpodobnostním mechanismem, který závisí na hodnotě neznámého parametru $\boldsymbol{\theta} \in \Theta$, kde Θ je uvažovaný parametrický prostor. Klasicky se předpokládá, že Θ je nějaká otevřená podmnožina \mathbb{R}^d . Často se stává, že parametr $\boldsymbol{\theta}$ se skládá z J různých *podparametrů*, tj. $\boldsymbol{\theta} = \text{vec}(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_J)$, kde $\boldsymbol{\theta}_j, j \in \{1, \dots, J\}$ jsou vektory parametrů popisujících objekt uceleného významu (skalár, vektor, matice). Parametrický prostor je potom kartézským součinem jednotlivých Θ_j – otevřených množin přípustných hodnot parametru $\boldsymbol{\theta}_j$.

Předpokládejme, že pravděpodobnostní mechanismus, kterým při daném $\boldsymbol{\theta}$ vznikají naše data \mathbb{Y} je dán hustotou $p(\mathbf{y}|\boldsymbol{\theta})$ vůči nějaké σ -konečné míře μ . V klasické statistice se potom funkce $L(\boldsymbol{\theta}) = p(\mathbb{Y}|\boldsymbol{\theta})$ nazývá věrohodnostní funkce parametru $\boldsymbol{\theta}$. Nyní si však přidáme předpoklad, že samotný parametr $\boldsymbol{\theta}$ je také náhodný element. Rozdělení tohoto parametru je určeno hustotou $p(\boldsymbol{\theta})$ vůči σ -konečné míře ν , tu budeme nazývat hustotou *apriorní*.

Na apriorní hustotu se lze v jistém smyslu dívat jako na míru našeho přesvědčení, že $\boldsymbol{\theta}$ nabývá právě této hodnoty. Tato míra může být podložena historickými daty či pouze statistickým subjektivním názorem. Více o této problematice v další části 2.2.

Nyní je však čas představit si způsob, jakým je naše apriorní informace o parametru $\boldsymbol{\theta}$ a věrohodnost parametru obsažená v datech využita společně. Odpověď se skrývá ve využití podmíněného rozdělení $\mathbb{Y}|\boldsymbol{\theta}$ a apriorního rozdělení $\boldsymbol{\theta}$ pro odvození sdruženého rozdělení dvojice $(\mathbb{Y}, \boldsymbol{\theta})$. Následně z tohoto rozdělení obdržíme podmíněné rozdělení parametru $\boldsymbol{\theta}$ za znalosti našich dat \mathbb{Y} , to přesně odpovídá situaci, ve které se nacházíme. Toto takzvané *aposteriorní* rozdělení parametru $\boldsymbol{\theta}$ můžeme opět charakterizovat pomocí hustoty. Následující Bayesova věta nám dává návod, jak tuto *aposteriorní* hustotu získat.

Věta 2 (Bayesova věta). *Nechť podmíněné rozdělení $\mathbb{Y}|\boldsymbol{\theta}$ je určeno hustotou $p(\mathbf{y}|\boldsymbol{\theta})$ a apriorní hustota náhodného parametru $\boldsymbol{\theta}$ je $p(\boldsymbol{\theta})$. Potom hustota podmíněného rozdělení $\boldsymbol{\theta}|\mathbb{Y}$ má tvar*

$$p(\boldsymbol{\theta}|\mathbf{y}) = \begin{cases} \frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})}, & \text{je-li } p(\mathbf{y}) = \int_{\Theta} p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\nu(\boldsymbol{\theta}) \neq 0, \\ 0, & \text{jinak.} \end{cases} \quad (2.1)$$

Důkaz. Plyne především z rovnosti $p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) = p(\mathbf{y}, \boldsymbol{\theta}) = p(\boldsymbol{\theta}|\mathbf{y})p(\mathbf{y})$. Podrobněji viz Anděl (2007, Kapitola 3.5, Věta 3.21). □

V této práci se budeme zabývat pouze rozděleními s hustotou vůči patřičně rozměrné Lebesgueově míře λ , proto budeme nadále v integrálech psát místo $d\mathbf{v}(\boldsymbol{\theta})$ pouze $d\boldsymbol{\theta}$. Z Bayesovy věty vidíme, že pro aposteriorní hustotu parametru $\boldsymbol{\theta}$ platí

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto \underbrace{p(\mathbf{y}|\boldsymbol{\theta})}_{\text{specifikace modelu}} \cdot \underbrace{p(\boldsymbol{\theta})}_{\text{apriorní informace}}, \quad (2.2)$$

neboť nyní podmiňujeme znalostí \mathbf{y} . Stačí tedy mít jistotu, že integrál $p(\mathbf{y}) = \int p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$ je nenulový a konečný. Proporcionalita zavedená v definici 1 nám tak při zaručení tohoto předpokladu značně zjednodušuje zápis aposteriorního rozdělení.

2.2 Volba apriorního rozdělení

Nyní se budeme věnovat problému volby apriorního rozdělení parametru $\boldsymbol{\theta}$. V literatuře byla navržena celá řada způsobů jakými stanovit vhodné rozdělení. Některé z nich si zde představíme, ovšem pouze v takovém rozsahu, v jakém to budeme nadále potřebovat.

Typickým požadavkem na apriorní rozdělení je, aby ona přidávaná informace o $\boldsymbol{\theta}$ nebyla významnější než informace o $\boldsymbol{\theta}$ obsažená v datech (respektive v předpokládaném modelu pro data \mathbb{Y} při $\boldsymbol{\theta}$). Toto ne zcela jasné pravidlo bude nejlepší představit na příkladě.

Příklad 1 (Jednorozměrné normální rozdělení se známým parametrem rozptylu). Mějme náhodný výběr X_1, \dots, X_n z jednorozměrného normálního rozdělení s neznámou střední hodnotou μ a známým rozptylem σ_0^2 . Potom věrohodnost parametru μ je dána hustotou

$$\begin{aligned} p(x_1, \dots, x_n|\mu) &\propto \prod_{i=1}^n \exp\left\{-\frac{1}{2\sigma_0^2}(x_i - \mu)^2\right\} \propto \exp\left\{-\frac{1}{2\sigma_0^2} \sum_{i=1}^n (x_i - \mu)^2\right\} \\ &\propto \exp\left\{-\frac{1}{2\sigma_0^2} \sum_{i=1}^n (x_i - \bar{x}_n)^2 - \frac{n}{2\sigma_0^2}(\bar{x}_n - \mu)^2\right\}, \end{aligned}$$

kde $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$ je aritmetický průměr hodnot x_1, \dots, x_n . Považujme parametr μ za náhodnou veličinu s apriorním normálním rozdělením s pevně stanovenými parametry μ_0 a v^2 . Podle (1.1) apriorní hustota μ splňuje

$$p(\mu) \propto \exp\left\{-\frac{1}{2v^2}(\mu - \mu_0)^2\right\},$$

a tedy dle Bayesovy věty 2 je aposteriorní hustota proporcionální

$$\begin{aligned} p(\mu|x_1, \dots, x_n) &\propto \exp\left\{-\frac{n}{2\sigma_0^2}(\bar{x}_n - \mu)^2 - \frac{1}{2v^2}(\mu - \mu_0)^2\right\} \\ &\propto \exp\left\{-\frac{1}{2} \left(\frac{n}{\sigma_0^2} + \frac{1}{v^2}\right) \left(\mu - \frac{\frac{n}{\sigma_0^2}\bar{x}_n + \frac{1}{v^2}\mu_0}{\frac{n}{\sigma_0^2} + \frac{1}{v^2}}\right)^2\right\}, \end{aligned}$$

kde jsme použili metody doplnění na čtverec vzpomenuté v apendixu B.

Porovnáním výrazu obdržené hustoty s rovnicí (1.1) vidíme, že aposteriorní rozdělení parametru μ (při známých realizacích x_1, \dots, x_n našeho výběru) je tedy $\mathbf{N}\left(\frac{nv^2}{nv^2+\sigma_0^2}\bar{x}_n + \frac{\sigma_0^2}{nv^2+\sigma_0^2}\mu_0, \frac{\sigma_0^2 v^2}{nv^2+\sigma_0^2}\right)$. Aposteriorní střední hodnota parametru μ je tedy konvexní kombinací hodnot \bar{x}_n a μ_0 . Pokud jsme zvolili v^2 menší než σ_0^2/n , tak je daleko vyšší váha na apriorní střední hodnotě μ_0 než na průměru dostupných dat \bar{x}_n . Jelikož si μ_0 a v^2 volíme sami předem, tak bychom byli schopni velmi nízkou volbou v^2 a hodnotou μ_0 dle našeho přání zajistit, aby náš aposteriorní odhad střední hodnoty μ byl v podstatě přibližně roven zvolenému μ_0 .

◇

To, že se dá takhle jednoduše zmanipulovat konečný výsledek, proto bývá terčem kritiky bayesovských metod. Abychom tedy mohli používat tyto metody bez možného podezření z podvodu, musíme si být vědomi, jak silné je naše přesvědčení o skutečné hodnotě parametru θ (v podobě apriorního rozdělení) v porovnání s informací skrytou v našich datech. Snažíme se tedy proto volit apriorní rozdělení co nejméně informativní.

2.2.1 Neinformativní rozdělení

Jednou z možností, jak se dá zaručit neovlivnění výsledků, je použití zcela neinformativního rozdělení. Tento postup se dá jednoduše interpretovat tak, že když nemáme žádnou představu o hodnotě parametru θ , pak je každá možná hodnota $\theta \in \Theta$ apriori stejně pravděpodobná. Tedy apriorní rozdělení θ se volí rovnoměrné na množině Θ . Existuje-li hustota tohoto rozdělení, pak je konstantní, a tedy $p(\theta) \propto 1$. Často se však pracuje i s nevlastní hustotou, tj. specifikuje se $p(\theta) \propto 1$, i když $\int_{\Theta} d1v(\theta) = \infty$. Důležité však stále je, aby výsledná aposteriorní hustota (která je pak proporcionální pouze samotné věrohodnosti $p(y|\theta)$) stále existovala a byla vlastní.

Příklad (Pokračování příkladu 1). Jelikož $\mu \in \mathbb{R}$, tak potřebujeme specifikovat rozdělení na celém \mathbb{R} a má-li být zcela neinformativní, tak musí být nevlastní. Jeden způsob, jak se na to dívat je, že $\mu \sim \mathbf{N}(\mu_0, v^2)$, kde $v^2 \rightarrow \infty$. Čím více zvyšujeme rozptyl apriorního rozdělení, tím srovnatelnější jsou hodnoty hustoty okolo μ_0 s těmi více vzdálenými od μ_0 , až se při dostatečně velkém v^2 zdá hustota téměř konstantní, a tedy neinformativní. Lze se na to také dívat tak, že uvažujeme rovnoměrné rozdělení na intervalu $(\mu_0 - a, \mu_0 + a)$, kde $a \rightarrow \infty$. Tyto postupy lze jednoduše zobecnit i na parametr střední hodnoty vícerozměrného normálního rozdělení.

◇

Příklad 2 (Jednorozměrné normální rozdělení s neznámými parametry μ i σ^2). Nyní uvažme, že σ^2 z předchozího příkladu je také neznámým parametrem a chtěli bychom pro něj také předpokládat apriorní neinformativnost. I když se zde nabízí opět uvažovat, že σ^2 je rovnoměrně rozděleno na intervalu $(0, \infty)$, tak se nejedná o klasický postup. Pokud bychom totiž uvažovali σ^2 parametrizované pomocí parametru přesnosti $\tau = \sigma^{-2}$, tak by podle věty o transformaci C9 byla apriorní hustota $p(\tau)$ proporcionální jakobiánu τ^{-2} . Zde je již vidět, že i když jsme vyšli z žádné předem dané informace, tak v pojetí přesnosti již svým způsobem informativní je. Výsledná aposteriorní rozdělení vzniklá volbou $p(\tau) \propto \tau^{-2}$ a $p(\tau) \propto 1$

se potom od sebe liší.

Jako neinformativní apriorní rozdělení se uvažuje nevlastní konstantní hustota pro logaritmus směrodatné odchylky, tj. $p(\log(\sigma)) \propto 1$. Podle věty o transformaci C9 tato volba zpět v řeči σ^2 dává opět nevlastní hustotu $p(\sigma^2) \propto \sigma^{-2}$. Jedná se v podstatě o limitní inverzní Wishartovo rozdělení $\text{IW}_1(q, \nu)$ pro $\nu \rightarrow 2 \Leftrightarrow \nu_0 \rightarrow 0$. Tento postup lze také zobecnit na obecnou neznámou varianční matici Σ řádu k vektorového normálního rozdělení, za jejíž neinformativní apriorní hustotu se považuje $p(\Sigma) \propto |\Sigma|^{-\frac{k+1}{2}}$.

◇

Zde narážíme na problém volby apriorního rozdělení takového, aby výsledné aposteriorní rozdělení nezáviselo na volbě parametrizace. Řešením je tzv. Jeffreysova apriorní hustota (může být i nevlastní). Jelikož nebudeme tento pojem dále potřebovat, tak se zde odkážeme na učební text Bayesovské metody (viz Hušková, 1985, Kapitola 2.3).

2.2.2 Konjugované rozdělení

Na příkladu 1 výše jsme si ukázali, že pokud předpokládáme jak normalitu dat, tak samotného parametru střední hodnoty, tak výsledné aposteriorní rozdělení je opět normální. Můžeme tedy opět jednoduše pracovat s aposteriorním rozdělením (konstruovat odhady či věrohodnostní množiny). V tomto případě se tak zdařilo díky tomu, že μ vystupovalo v exponenciále v součtu dvou čtverců a ty umíme díky dodatku B znovu převést na čtverec.

Když tedy víme, že aposteriorní hustota je proporcionální součinu $p(\mathbf{y}|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta})$, tak se můžeme pokusit nalézt takovou apriorní hustotu $p(\boldsymbol{\theta})$, že lze vhodně zkombinovat s věrohodností tak, aby onen součin těchto dvou hustot jako funkce parametru $\boldsymbol{\theta}$ připomínal opět hustotu nějakého známého rozdělení. Konjugované apriorní rozdělení je takové, které nám zajistí, že výsledné aposteriorní rozdělení je ze stejné rodiny jako samotné apriorní rozdělení:

Definice 2. *Bud' \mathcal{F}_θ systém apriorních hustot náhodného parametru $\boldsymbol{\theta} \in \Theta$. Řekneme, že systém \mathcal{F}_θ je konjugovaný vzhledem k systému hustot $\{p(\mathbf{y}|\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$, jestliže pro každé $p(\boldsymbol{\theta}) \in \mathcal{F}_\theta$ aposteriorní hustota $p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \in \mathcal{F}_\theta$.*

Typický systém hustot \mathcal{F}_θ přísluší nějaké rodině rozdělení, ve které jsou jednotlivá rozdělení indexována hyperparametrem $\boldsymbol{\alpha}$, tj. $\mathcal{F}_\theta = \{q(\boldsymbol{\theta}|\boldsymbol{\alpha}), \boldsymbol{\alpha} \in A\}$. Jestliže tedy zvolíme za apriorní hustotu $p(\boldsymbol{\theta}) = q(\boldsymbol{\theta}|\boldsymbol{\alpha}_0)$, tak musí existovat takové $\boldsymbol{\alpha}_1 \in A$, že $p(\boldsymbol{\theta}|\mathbf{y}) = q(\boldsymbol{\theta}|\boldsymbol{\alpha}_1)$. Jestliže je naše apriorní rozdělení neinformativní oproti informaci ukryté v datech, tak by měla věrohodnost obsažená v datech významně změnit původní parametr $\boldsymbol{\alpha}_0$.

Existuje metoda založená na postačujících statistikách (viz Hušková, 1985), která pro daný systém $\{p(\mathbf{y}|\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$ nalezne konjugovaný systém apriorních hustot. My se zde však zaměříme spíše na využití tohoto přístupu k volbě apriorního rozdělení, zejména v případě parametru střední hodnoty a rozptylu normálního rozdělení. Již pro jednorozměrný případ z příkladu 1 víme, že při známém σ_0^2 je konjugovaným systémem pro parametr střední hodnoty μ opět obecné normální rozdělení $\text{N}(\mu_0, v^2)$, kde μ_0 a v^2 jsou dodatečné hyperparametry.

Srovnáním hustot (1.1) a (1.8) vidíme, že nehledě na znalost parametru μ je inverzní Wishartovo rozdělení pro apriorní rozdělení σ^2 ideální volbou. Jsou-li

oba dva parametry μ a σ^2 neznámé (příklad 2), tak již musíme uvažovat tzv. normální-gama rozdělení. Toto rozdělení $\text{N}\Gamma$ je definováno následovně:

$$(\mu, \sigma^2) \sim \text{N}\Gamma(\mu_0, v^2, q, \nu) \iff \mu | \sigma^2 \sim \text{N}(\mu_0, \sigma^2 v^2), \sigma^2 \sim \text{IW}_1(q, \nu).$$

Nazývá se tak proto, že se často definuje pomocí inverzního rozptylu, tzv. přesnosti, která se řídí Wishartovým rozdělením, což, jak víme, není nic jiného než gama rozdělení s příslušnými parametry. Hustota tohoto rozdělení lze snadno určit jako součin hustot $p(\mu | \sigma^2; \mu_0, v^2)$ a $p(\sigma^2 | q, \nu)$, viz (1.1) a (1.8). Dá se ukázat, že toto rozdělení je konjugované pro hustotu $p(\mathbf{y} | \mu, \sigma^2)$ z $\text{N}(\mu, \sigma^2)$, tedy že aposteriorní rozdělení μ a σ^2 je potom opět normální-gama.

Tyto poznatky se dají snadno zobecnit pro vektorové $\text{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ i maticové normální rozdělení $\text{N}_{n \times k}(\mathbb{M}, \boldsymbol{\Phi} \otimes \boldsymbol{\Sigma})$. Pro parametr střední hodnoty ($\boldsymbol{\mu}$ či \mathbb{M}) je opět vhodné předpokládat normální rozdělení s námi předem určenými hyperparametry. Pro zachování konjugovanosti je opět nutné, aby parametr popisující měřítko v apriorním rozdělení parametru střední hodnoty závisel na neznámých parametrech měřítka ($\boldsymbol{\Sigma}$ či $\boldsymbol{\Phi}$ a $\boldsymbol{\Sigma}$) původního rozdělení. Pro parametry měřítka předpokládáme inverzní Wishartovo rozdělení.

Dále lze pracovat se zobecněným konjugovaným systémem, který pro parametr střední hodnoty a parametr měřítka předpokládá apriorní nezávislost. Tedy parametr střední hodnoty se řídí normálním rozdělením s ryze vlastní měřítkovou maticí. Zde už ale ztrácíme vlastnost konjugovanosti tak, jak je zadefinována v definici 2. Nicméně i s takto zvoleným apriorním rozdělením jsme stále schopni dobře pracovat.

2.2.3 Volba hyperparametrů

Ať už využíváme konjugovaného systému hustot nebo ne, ve většině případů předpokládáme, že apriorní rozdělení parametru $\boldsymbol{\theta}$ pochází z nějakého systému hustot, který je typicky určen různými hodnotami hyperparametru $\boldsymbol{\alpha} \in A$. Hušková (1985) ve svých skriptech uvádí, že pomocí empirických bayesovských metod lze na základě historických dat stanovit apriorní rozdělení i neparametricky. Náš přístup v této práci však bude vždy čistě parametrický.

Předpokládejme, že před samotným provedením experimentu máme k dispozici nějakou minulou zkušenost v podobě historických dat \mathbf{y}^{H} pocházejících ze stejného modelu určeného $p(\mathbf{y}^{\text{H}} | \boldsymbol{\theta})$, kde $\boldsymbol{\theta}$ je náhodně rozděleno v závislosti na parametru $\boldsymbol{\alpha}$. Pak marginální rozdělení historických dat je $p(\mathbf{y}^{\text{H}} | \boldsymbol{\alpha}) = \int_{\Theta} p(\mathbf{y}^{\text{H}} | \boldsymbol{\theta}; \boldsymbol{\alpha}) \cdot p(\boldsymbol{\theta} | \boldsymbol{\alpha}) d\boldsymbol{\theta}$. Pokud jsme schopni tento integrál spočítat (např. v případě $\boldsymbol{\theta}$ parametry normálního rozdělení), tak získáme věrohodnost parametru $\boldsymbol{\alpha}$ v historických datech \mathbf{x}^{H} . Můžeme tak použít klasické nebayesovské metody pro odhad parametru $\boldsymbol{\alpha}$ (momentová metoda, metoda maximální věrohodnosti).

Příklad (Pokračování příkladu 1). Budte $X_1^{\text{H}}, \dots, X_{n_{\text{H}}}^{\text{H}}$ nezávislé z $\text{N}(\mu, \sigma_0^2)$, kde σ_0^2 je známé, a předpokládejme, že $\mu \sim \text{N}(\mu_0, v^2)$. Marginální rozdělení veličin X_i^{H} je $\text{N}(\mu_0, \sigma^2)$, kde $\sigma^2 = \sigma_0^2 + v^2$. Za odhady parametrů μ_0 a σ^2 pak vezmeme klasické statistiky

$$\hat{\mu}_0 = \frac{1}{n_{\text{H}}} \sum_{i=1}^{n_{\text{H}}} X_i^{\text{H}} \quad \text{a} \quad \hat{\sigma}^2 = \frac{1}{n_{\text{H}} - 1} \sum_{i=1}^{n_{\text{H}}} (X_i^{\text{H}} - \hat{\mu}_0)^2.$$

Ještě dopočteme $\hat{v}^2 = \hat{\sigma}^2 - \sigma_0^2$ a výsledná volba apriorního rozdělení by měla být $\mathbf{N}(\hat{\mu}_0, \hat{v}^2)$. Nastává-li, že $\hat{\sigma}^2 < \sigma_0^2$, tak bychom museli volit $\hat{v}^2 = 0$, což odpovídá degenerovanému rozdělení koncentrovanému v $\hat{\mu}_0$. Toto je silně informativní apriorní rozdělení, neboť nám zúžilo parametrický prostor $\Theta = \mathbb{R}$ na jedinou možnou hodnotu $\hat{\mu}_0$. V takovém případě bychom tedy měli volit nějaký jiný způsob volby apriorního rozdělení. ◇

Rowe (2003) ve své publikaci uvádí postup, kterým lze odvodit obecnou volbu apriorního rozdělení pro parametry skalárního, vektorového i maticového normálního rozdělení. Výsledkem je, že parametr polohy $(\mu_0, \boldsymbol{\mu}_0, \mathbb{M}_0)$ apriorního rozdělení se vždy odhadne pomocí průměru historických dat $(\frac{1}{n_H} \sum X_i^H, \frac{1}{n_H} \sum \mathbf{X}_i^H, \frac{1}{n_H} \sum \mathbb{X}_i^H)$. Za rozptylový parametr se potom při použití konjugovaného rozdělení používá $\frac{1}{n_H} \sigma^2, \frac{1}{n_H} \boldsymbol{\Sigma}$ či dvojice $\boldsymbol{\Phi}$ a $\frac{1}{n_H} \boldsymbol{\Sigma}$, tedy zde hrají roli původní náhodné parametry měřítka.

Pokud bychom zde chtěli nějaký obecný parametr měřítka, tak ztrácíme konjugovanost a nastávají problémy s určením $p(\mathbf{y}^H | \boldsymbol{\alpha}) = \int_{\Theta} p(\mathbf{y}^H | \boldsymbol{\theta}; \boldsymbol{\alpha}) \cdot p(\boldsymbol{\theta} | \boldsymbol{\alpha}) d\boldsymbol{\theta}$. Již v jednorozměrném případě nelze vyintegrovat oba neznámé a náhodné parametry $\mu \sim \mathbf{N}(\mu_0, v^2)$ a $\sigma^2 \sim \text{IW}_1(q, \nu)$ a dostat potom známé rozdělení dané hustotou $p(\mathbf{x}^H | \mu_0, v^2, q, \nu)$, ze kterého by šlo hezky odvodit nejvhodnější volbu parametrů μ_0, v^2, q a ν . Statistik tak musí zvolit nastavení hyperparametrů dle svého uvážení či může měřítkový parametr v rozdělení parametru polohy $\mu, \boldsymbol{\mu}$ či \mathbb{M} stanovit jako náhodný (pomocí inverzního Wishartova rozdělení s již pevně zvolenými parametry).

Pro náhodný parametr rozptylu normálního rozdělení předpokládáme standardně inverzní Wishartovo rozdělení $\text{IW}_k(\mathbb{Q}, \nu)$ vzhledem k jeho konjugovanosti k předpokládané normální věrohodnosti. Máme-li ve vektorovém případě historická data $\mathbf{X}_1^H, \dots, \mathbf{X}_{n_H}^H \sim \mathbf{N}_k(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$, pak Rowe (2003) doporučuje použít

$$\mathbb{Q} = \sum_{i=1}^{n_H} (\mathbf{X}_i^H - \bar{\mathbf{X}}_{n_H}^H) (\mathbf{X}_i^H - \bar{\mathbf{X}}_{n_H}^H)^\top \quad \text{a} \quad \nu = n_H, \quad \text{kde} \quad \bar{\mathbf{X}}_{n_H}^H = \frac{1}{n_H} \sum_{i=1}^{n_H} \mathbf{X}_i^H,$$

což se v jednorozměrném případě zjednodušuje na $q = \sum_{i=1}^{n_H} (X_i^H - \bar{X}_{n_H}^H)^2$. V maticovém případě $\mathbf{N}_{n \times k}(\mathbb{M}, \boldsymbol{\Phi} \otimes \boldsymbol{\Sigma})$ je třeba určit rozdělení obou parametrů $\boldsymbol{\Phi}$ a $\boldsymbol{\Sigma}$. Při dispozici historických dat $\mathbb{X}_1^H, \dots, \mathbb{X}_{n_H}^H$ se doporučuje pro parametry předpokládaných rozdělení $\boldsymbol{\Sigma} \sim \text{IW}_k(\mathbb{Q}, \nu)$ a $\boldsymbol{\Phi} \sim \text{IW}_n(\boldsymbol{\Psi}, \kappa)$ použít ideálně

$$\begin{aligned} \mathbb{Q} &= \sum_{i=1}^{n_H} (\mathbb{X}_i^H - \bar{\mathbb{X}}_{n_H}^H)^\top \boldsymbol{\Phi}^{-1} (\mathbb{X}_i^H - \bar{\mathbb{X}}_{n_H}^H) \quad \text{a} \quad \nu = n \cdot n_H, \\ \boldsymbol{\Psi} &= \sum_{i=1}^{n_H} (\mathbb{X}_i^H - \bar{\mathbb{X}}_{n_H}^H) \boldsymbol{\Sigma}^{-1} (\mathbb{X}_i^H - \bar{\mathbb{X}}_{n_H}^H)^\top \quad \text{a} \quad \kappa = k \cdot n_H. \end{aligned}$$

Jak vidíme, tyto volby jsou vzájemně provázané a nelze tak nalézt vzorce v uzavřeném tvaru. S tímto problémem si lze poradit pomocí iterování výpočtů jednotlivých matic s vhodným počátečním bodem.

Nemá-li statistik k dispozici žádná historická data, je volba hyperparametrů apriorních rozdělení ponechána čistě jeho zkušenostem a představám o rozdělení parametrů. Zde je dobré mít na paměti, jaké jsou základní charakteristiky

normálního a inverzního Wishartova rozdělení. Modus rozdělení lze chápat jako hodnota, v jejímž okolí se s největší pravděpodobností nachází skutečná hodnota parametru. Zvláště opatrní bychom měli být při volbě parametrů rozptylu, kdy musíme mít na paměti, aby výsledné apriorní rozdělení nebylo příliš informativní. Za povšimnutí zde stojí, že rozptyl inverzního Wishartova rozdělení je zhruba dvojnásobek druhé mocniny střední hodnoty (tu si typicky zvolíme) dělené navíc počtem stupňů volnosti, od kterých je odečtena konstanta závislá na dimenzi k . Tedy s vyšším počtem stupňů volnosti si jsme více jistí, že námi stanovená hodnota modu (střední hodnoty) je ta pravá.

2.3 Metody odhadu

Bayesovské metody odhadu jsou přehledně vysvětleny a předvedeny na příkladech v knize (Robert, 2001, kapitoly 4–6).

Běžně používanou metodou odhadu v klasické statistice jsou maximálně věrohodné odhady založené na věrohodnosti $p(\mathbb{Y}|\boldsymbol{\theta})$, které za jistých předpokladů mají pěkné asymptotické vlastnosti. Zde však máme navíc i předpokládané apriorní rozdělení parametru $\boldsymbol{\theta}$, které svým dílem také přispěje do výsledných bayesovských odhadů. Je intuitivně zřejmé, že pro odhadování a inferenci budeme používat aposteriorního rozdělení parametru $\boldsymbol{\theta}$, neboť hodnota $\boldsymbol{\theta}$ je nám neznámá a chceme se o ní něco dozvědět za znalosti našich naměřených dat \mathbb{Y} .

Základním odhadem samotného parametru $\boldsymbol{\theta}$ je aposteriorní střední hodnota, tedy (podmíněná) střední hodnota vzhledem k aposteriornímu rozdělení $\boldsymbol{\theta}|\mathbb{Y}$, tj.

$$\hat{\boldsymbol{\theta}} = \mathbb{E}[\boldsymbol{\theta}|\mathbb{Y}] = \int_{\Theta} \boldsymbol{\theta} p(\boldsymbol{\theta}|\mathbb{Y}) d\boldsymbol{\theta}. \quad (2.3)$$

K tomu je zapotřebí vědět, o jaké aposteriorní rozdělení se jedná (pak se lze vyhnout integrování). Jinak je zapotřebí umět dopočítat normalizační konstantu v rovnici (2.2) a umět (numericky) spočítat integrál (2.3). Jestliže je parametr $\boldsymbol{\theta} = \text{vec}(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_J)$, kde $\boldsymbol{\theta}_j$ může být skalár, vektor či (pozitivně definitní) matice, tak se pro odhad $\boldsymbol{\theta}_j, j \in \{1, \dots, J\}$ používá marginální aposteriorní rozdělení $\boldsymbol{\theta}_j|\mathbb{Y}$, které má hustotu

$$p(\boldsymbol{\theta}_j|\mathbb{Y}) = \int_{\Theta_{(-j)}} p(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_J|\mathbb{Y}) \underbrace{d\boldsymbol{\theta}_1 \cdots d\boldsymbol{\theta}_{j-1} d\boldsymbol{\theta}_{j+1} \cdots d\boldsymbol{\theta}_J}_{=: d\boldsymbol{\theta}_{(-j)}}$$

kde $\Theta_{(-j)}$ značí parametrický prostor, který vznikne zafixováním j -té složky parametru $\boldsymbol{\theta}$. K tomu, aby šlo toto marginální aposteriorní rozdělení spočítat, je třeba umět efektivně počítat integrály podle ostatních složek $\boldsymbol{\theta}$, tedy i například integrál podle pozitivně definitní matice. Následující tvrzení (zjednodušené pro pouze dvě složky) ukazuje, jak se taková integrace dá provést na základě znalosti hustoty pravděpodobnostního rozdělení.

Tvrzení 3. *Bud' $\boldsymbol{\theta} = \text{vec}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ dvousložkový parametr a necht' sdružená aposteriorní hustota je tvaru $p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\mathbb{Y}) \propto g(\boldsymbol{\theta}_1|\mathbb{Y}) \cdot h(\boldsymbol{\theta}_2|\mathbb{Y}, \boldsymbol{\theta}_1)$, kde funkce g již nezávisí na $\boldsymbol{\theta}_2$ a funkce h je v parametru $\boldsymbol{\theta}_2$ funkce proporciální hustotě nějakého známého rozdělení, tedy existuje nějaká kladná multiplikatívni konstanta $k(\boldsymbol{\theta}_1|\mathbb{Y})$, že podíl h/k je hustota. Potom $p(\boldsymbol{\theta}_1|\mathbb{Y}) \propto g(\boldsymbol{\theta}_1|\mathbb{Y}) \cdot k(\boldsymbol{\theta}_1|\mathbb{Y})$.*

Důkaz. Dle definice 1 existuje konstanta $c(\mathbf{y})$ taková, že $p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 | \mathbf{y}) = c(\mathbf{y}) \cdot g(\boldsymbol{\theta}_1, \mathbf{y}) \cdot h(\boldsymbol{\theta}_2 | \mathbf{y}, \boldsymbol{\theta}_1)$. Jelikož podíl funkcí h a k tvoří hustotu v parametru $\boldsymbol{\theta}_2$, tak platí, že

$$\begin{aligned} p(\boldsymbol{\theta}_1 | \mathbf{y}) &= \int_{\Theta} p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 | \mathbf{y}) d\boldsymbol{\theta}_2 = c(\mathbf{y}) \cdot g(\boldsymbol{\theta}_1 | \mathbf{y}) \cdot k(\boldsymbol{\theta}_1 | \mathbf{y}) \cdot \int_{\Theta} \frac{h(\boldsymbol{\theta}_2 | \mathbf{y}, \boldsymbol{\theta}_1)}{k(\boldsymbol{\theta}_1 | \mathbf{y})} d\boldsymbol{\theta}_2 \\ &= c(\mathbf{y}) \cdot g(\boldsymbol{\theta}_1 | \mathbf{y}) \cdot k(\boldsymbol{\theta}_1 | \mathbf{y}) \propto g(\boldsymbol{\theta}_1 | \mathbf{y}) \cdot k(\boldsymbol{\theta}_1 | \mathbf{y}). \end{aligned}$$

Volbou $g^*(\boldsymbol{\theta}_1 | \mathbf{y}) = c(\mathbf{y}) \cdot g(\boldsymbol{\theta}_1 | \mathbf{y})$ dostaneme dokonce rovnost namísto \propto .

□

Toto tvrzení zůstává platné i pokud podíl h/k je až na konstantu (nezávislou na $\boldsymbol{\theta}_1$ i $\boldsymbol{\theta}_2$) hustotou. Tvrzení 3 použijeme vždy tak, že $\boldsymbol{\theta}_2$ bude parametr, který se v tomto kroku snažíme vyintegrovat, a $\boldsymbol{\theta}_1$ pak představuje zbývající ještě nevyintegrované parametry. Často bývá klíčové umět zvolit správné pořadí, ve kterém provádět onu integraci.

Marginální aposteriorní střední hodnota však není jediný možný odhad $\boldsymbol{\theta}$. V jednorozměrném případě lze také použít marginálního aposteriorního mediánu $\text{med}[\boldsymbol{\theta} | \mathbb{Y}]$. Tento tzv. robustní odhad je (jako v klasické statistice) ve střední hodnotě optimální vzhledem ke ztrátové funkci vzdálenosti od samotného $\boldsymbol{\theta}$. Aposteriorní střední hodnota je ve střední hodnotě optimální vzhledem ke kvadratické ztrátové funkci. Podrobněji o tom, v jakém přesně významu jsou tyto odhady optimální, se lze dočíst ve skriptech Hušková (1985).

Zatím jsme se věnovali pouze odhadu bodovému, stejně jako v klasické statistice lze také sestavit intervalové odhady (přesněji řečeno množinové odhady). V bayesovské statistice se jim říká věrohodnostní množiny parametru $\boldsymbol{\theta}$. Jedná se jednoduše o takové množiny $M_\alpha \subset \Theta$, jejichž aposteriorní pravděpodobnost toho, že v ní $\boldsymbol{\theta}$ leží, dosahuje předem stanoveného pokrytí $1 - \alpha \in (0, 1)$, tj. $\mathbb{P}[\boldsymbol{\theta} \in M_\alpha | \mathbb{Y}] = 1 - \alpha$. Takových množin lze sestavit nekonečně mnoho, a proto je vhodné konstruovat množiny se speciální vlastností. Prvním typem jsou takzvané ET intervaly (Equal-Tailed) pro jednorozměrný parametr θ_j , které splňují to, že aposteriorní pravděpodobnost toho, že parametr θ_j nabyde hodnoty vyšší než horní mez intervalu je $\frac{\alpha}{2}$, a tedy je stejná jako pravděpodobnost toho, že parametr θ_j nabyde hodnoty nižší než dolní mez intervalu. Druhým typem je tzv. HPD množina (Highest Posterior Density), jedná se o množinu (či interval) s vlastností $M_\alpha = \{\boldsymbol{\theta} \in \Theta : p(\boldsymbol{\theta} | \mathbf{y}) \geq k_\alpha\}$, kde k_α je konstanta zajišťující $\mathbb{P}[\boldsymbol{\theta} \in M_\alpha | \mathbb{Y}] = 1 - \alpha$.

Často ve složitějších modelech (jako bude model bayesovské faktorové analýzy) nepůjde stanovit jednotlivá marginální aposteriorní rozdělení. A proto je zapotřebí jiných metod odhadu parametru $\boldsymbol{\theta}$ založených buď na generování z aposteriorního rozdělení nebo iteračních algoritmech.

2.3.1 MCMC metody

V této části práce se budeme zabývat Monte Carlo Markov Chain (MCMC) metodami. Jedná se o spojení metody odhadu založené na simulacích (Monte Carlo) a metody generování markovských řetězců s předem danými vlastnostmi. Podrobný přehled této kombinace metod je detailněji popsán například v knize Brookse, Gelmana, Jonese a Menga (Brooks, Gelman, Jones a Meng, 2011). Hned

v úvodní kapitole jsou zde definovány základní pojmy z teorie markovských řetězců, které zde nebudeme podrobně zavádět.

Monte Carlo metoda odhaduje charakteristiky rozdělení $\boldsymbol{\theta}|\mathbb{Y}$ na základě simulace náhodného výběru $\boldsymbol{\theta}^i, i \in \{1, \dots, M\}$ z tohoto rozdělení. Například jednorozměrnou funkci t parametru $\boldsymbol{\theta}$, která splňuje $\mathbf{E} [|t(\boldsymbol{\theta})| | \mathbb{Y}] < \infty$, lze odhadnout výběrovým průměrem hodnot $t(\boldsymbol{\theta}^i)$ (dle silného zákona velkých čísel). Druhá část názvu Markov Chain nám pomůže získat do jisté míry onen „náhodný výběr“. Ke generování jednotlivých $\boldsymbol{\theta}^i$ lze použít Gibbsův algoritmus (část 2.3.2).

Za generované hodnoty parametru $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_J)$ považujeme stavy, které navštíví markovský řetězec s následujícími vlastnostmi:

- jedná se o homogenní markovský řetězec,
- jeho stacionárním rozdělením je právě aposteriorní rozdělení $\boldsymbol{\theta}|\mathbb{Y}$,
- existuje limitní rozdělení tohoto markovského řetězce.

Od jistého $B > 0$ tak lze generované hodnoty $\boldsymbol{\theta}^{(B+i)}, i \in \{1, \dots, B + M\}$ považovat za reprezentanty limitního rozdělení řetězce. Z teorie markovských procesů víme, že limitní rozdělení homogenního markovského řetězce je shodné s rozdělením stacionárním. Nagenerované hodnoty lze tedy od jistého indexu B dokonce považovat za reprezentanty z aposteriorního rozdělení. Hodnoty $\boldsymbol{\theta}^i, i \in \{1, \dots, B\}$ tvoří tzv. zážehovou fázi (burn-in period) markovského řetězce a nejsou použity pro finální odhady.

Problémem však v reálné situaci bývá stanovení délky zážehové fáze B . Od jakého $B > 0$ lze opravdu tvrdit, že generovaný markovský řetězec dosáhl limitního rozdělení? Na tuto otázku neexistuje jednoznačná odpověď, případ od případu se vhodné B velmi liší, od řádů desítek po miliony. Za tímto účelem se generuje typicky velmi dlouhý řetězec a zkoumá se, zda je od nějakého okamžiku přibližně stabilní (stacionární). Dalším problémem s takto generovaným výběrem je ten, že generované hodnoty jsou velmi často pozitivně autokorelované, nemůžeme tedy generované veličiny považovat za nezávislé. Za účelem odstranění vysoké autokorelace se doporučuje ponechávat si pro finální analýzu jen každou K -tou hodnotu, kde $K > 1$ si vhodně stanovíme.

Odhadovat aposteriorní střední hodnotu průměrem generovaných hodnot (za t volíme projekci na jednorozměrnou složku parametru $\boldsymbol{\theta}$) je však možné i bez splnění předpokladu nezávislosti, zajistíme-li platnost ergodické věty, viz Robert a Casella (2004, Theorem 6.63). Platnost předpokladů ergodické věty zpravidla není nijak těžké zajistit, viz Brooks a kol. (2011, Část 2.5.4). Rozptyl $\text{var} [t(\boldsymbol{\theta}) | \mathbb{Y}] = \mathbf{E} [(t(\boldsymbol{\theta}))^2 | \mathbb{Y}] - (\mathbf{E} [t(\boldsymbol{\theta}) | \mathbb{Y}])^2$ odhadneme pomocí

$$\frac{1}{M} \sum_{i=1}^M (t(\boldsymbol{\theta}^{(B+i)}))^2 - \left(\frac{1}{M} \sum_{i=1}^M t(\boldsymbol{\theta}^{(B+i)}) \right)^2,$$

což opět plyne z ergodické věty, tentokrát pro druhý moment (za předpokladu jeho konečnosti). Tyto odhady lze pochopitelně zobecnit na vícerozměrné \boldsymbol{t} , tedy například konkrétní podsložky $\boldsymbol{t}(\boldsymbol{\theta}) = \boldsymbol{\theta}_j$. Odhadování maticového parametru $\boldsymbol{\theta}_j$ převádíme na vektorový případ pomocí vektorizace.

2.3.2 Gibbsův algoritmus

Tento algoritmus je pojmenován po fyzikovi J. W. Gibbsovi, který generoval náhodné veličiny ve statistické fyzice podobným postupem jako je dnešní podoba tohoto algoritmu. Gibbsův algoritmus tak, jak jej dnes známe a používáme, představili až 80 let po smrti samotného Gibbsa bratři Donald a Stuart Gemanovi ve svém článku Geman a Geman (1984).

Stále předpokládáme, že parametr $\boldsymbol{\theta} = \text{vec}(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_J)$ lze rozdělit do J ucelených částí. Připomínáme značení $\boldsymbol{\theta}_{(-j)} = \text{vec}(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_{j-1}, \boldsymbol{\theta}_{j+1}, \dots, \boldsymbol{\theta}_J)$. Mluvíme-li o *plně podmíněném* rozdělení části parametru $\boldsymbol{\theta}_j$, máme na mysli rozdělení $\boldsymbol{\theta}_j | \mathbb{Y}, \boldsymbol{\theta}_{(-j)}$, tedy podmiňujeme-li navíc znalostí i ostatních složek parametru $\boldsymbol{\theta}$. Hustotou tohoto rozdělení je

$$p(\boldsymbol{\theta}_j | \mathbb{Y}, \boldsymbol{\theta}_{(-j)}) = \frac{p(\boldsymbol{\theta}_j, \boldsymbol{\theta}_{(-j)} | \mathbb{Y})}{p(\boldsymbol{\theta}_{(-j)} | \mathbb{Y})} \propto p(\boldsymbol{\theta} | \mathbb{Y}), \quad (2.4)$$

tedy se stačí na aposteriori hustotu dívat jako na funkci v parametru $\boldsymbol{\theta}_j$. Model typicky konstruujeme tak, aby toto rozdělení bylo známé - např. za použití konjugovaných rozdělení, viz část 2.2.2. Důležitou podmínkou Gibbsova algoritmu je, že jsme schopni generovat ze všech plně podmíněných rozdělení jednotlivých $\boldsymbol{\theta}_j, j \in \{1, \dots, J\}$. Ke generování z rozdělení uvažovaných v této práci stačí použít vlastností uvedených v části 1.

Algoritmus 1. *Gibbsův algoritmus.*

Vstupy: požadovaná délka řetězce $B + M$, data \mathbb{y} , počáteční hodnota $\boldsymbol{\theta}^0$, metody generování z plně podmíněných rozdělení.

Výstupy: markovský řetězec $\boldsymbol{\theta}^i, i \in \{1, \dots, B + M\}$.

Postup: v $(i + 1)$ -ním kroku algoritmu generuj $\boldsymbol{\theta}^{i+1}$ po složkách následovně:

$$1) \boldsymbol{\theta}_1^{i+1} \sim \boldsymbol{\theta}_1 | \mathbb{Y} = \mathbb{y}, \boldsymbol{\theta}_{(-1)} = \boldsymbol{\theta}_{(-1)}^i,$$

$$2) \boldsymbol{\theta}_2^{i+1} \sim \boldsymbol{\theta}_2 | \mathbb{Y} = \mathbb{y}, \boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^{i+1}, \boldsymbol{\theta}_3 = \boldsymbol{\theta}_3^i, \dots, \boldsymbol{\theta}_J = \boldsymbol{\theta}_J^i,$$

⋮

$$J - 1) \boldsymbol{\theta}_{J-1}^{i+1} \sim \boldsymbol{\theta}_{J-1} | \mathbb{Y} = \mathbb{y}, \boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^{i+1}, \dots, \boldsymbol{\theta}_{J-2} = \boldsymbol{\theta}_{J-2}^{i+1}, \boldsymbol{\theta}_J = \boldsymbol{\theta}_J^i,$$

$$J) \boldsymbol{\theta}_J^{i+1} \sim \boldsymbol{\theta}_J | \mathbb{Y} = \mathbb{y}, \boldsymbol{\theta}_{(-J)} = \boldsymbol{\theta}_{(-J)}^i.$$

Konec algoritmu: jakmile $i + 1 = B + M$.

V každém kroku tohoto algoritmu generujeme z plně podmíněného rozdělení, kde jako hodnoty zbytku parametrů používáme vždy poslední nagenované hodnoty. Toto je tedy zdrojem časté vysoké autokorelovanosti našeho řetězce. Není nutné dodržovat v každém kroce stanovené pořadí od 1 po J , právě naopak pravidelná změna směru či randomizace tohoto pořadí může generovanému markovskému řetězci zajistit další pěkné vlastnosti (např. reversibilita), více viz Brooks a kol. (2011, část 1.12.7).

V téže literatuře se lze seznámit také s jinými algoritmy používanými pro generování markovského řetězce s požadovanými vlastnostmi. Jedním z nich je například Metropolisův-Hastingsův algoritmus.

2.3.3 ICM algoritmus

Analogií metody maximální věrohodnosti je v bayesovské statistice odhad sestavený na základě modu aposteriorního rozdělení, tj.

$$\tilde{\boldsymbol{\theta}} = \text{mod} [\boldsymbol{\theta}|\mathbb{Y}] = \arg \max_{\boldsymbol{\theta} \in \Theta} p(\boldsymbol{\theta}|\mathbb{Y}) = \arg \max_{\boldsymbol{\theta} \in \Theta} \log(p(\boldsymbol{\theta}|\mathbb{Y})).$$

Jsme-li schopni určit aposteriorní rozdělení jakožto nějaké známé rozdělení, lze příslušný odhad získat jednoduše dosazením příslušných hodnot do vzorce pro modus tohoto rozdělení, viz tabulka 1.1. Není-li toto rozdělení známé, je třeba použít numerických metod k nalezení onoho maxima.

Víme-li, že aposteriorní rozdělení je unimodální (a příslušnou hustotu lze derivovat), tak se můžeme uchýlit k derivování (logaritmované) aposteriorní hustoty podle $\boldsymbol{\theta}$. Následně řešením soustavy rovnic (derivace položeny rovné nule) nalezneme stacionární bod, o kterém dle našeho předpokladu unimodality můžeme tvrdit, že se jedná o maximum. Pro tyto účely je v mnohorozměrné statistice vhodné znát pravidla pro derivování dle vektorového a maticového parametru, která jsou uvedena v tvrzení A5.

Situace je poněkud komplikovanější v případě, že se $\boldsymbol{\theta}$ dá rozložit na J částí uceleného významu. Aposteriorní hustotou parametru $\boldsymbol{\theta} = \text{vec}(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_J)$ je $p(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_J|\mathbb{y})$. Maximalizaci této hustoty v celém $\boldsymbol{\theta}$ převedeme na hledání maxima v j -tých složkách $\boldsymbol{\theta}_j$, zatímco ostatní složky jsou pevně zafixované na hodnotě $\boldsymbol{\theta}_{(-j)}^0$. Základem je dívat se na aposteriorní hustotu jako na funkci v $\boldsymbol{\theta}_j$, tedy $p_j(\boldsymbol{\theta}_j) = p(\boldsymbol{\theta}_j, \boldsymbol{\theta}_{(-j)}^0|\mathbb{y})$. Maximalizace funkce p_j lze převést na maximalizaci hustoty $p(\boldsymbol{\theta}_j|\mathbb{y}, \boldsymbol{\theta}_{(-j)}^0)$ plně podmíněného rozdělení zavedeného v předchozí části. V naší situaci totiž

$$\begin{aligned} \arg \max_{\boldsymbol{\theta}_j} p(\boldsymbol{\theta}_j, \boldsymbol{\theta}_{(-j)}^0|\mathbb{y}) &= \arg \max_{\boldsymbol{\theta}_j} p(\boldsymbol{\theta}_j|\mathbb{y}, \boldsymbol{\theta}_{(-j)}^0) \overbrace{p(\boldsymbol{\theta}_{(-j)}^0|\mathbb{y})}^{\text{nezávisí na } \boldsymbol{\theta}_j} \\ &= \arg \max_{\boldsymbol{\theta}_j} p(\boldsymbol{\theta}_j|\mathbb{y}, \boldsymbol{\theta}_{(-j)}^0). \end{aligned}$$

Vidíme-li, že plně podmíněné rozdělení je známé, sáhneme po nějakém přehledu pravděpodobnostních rozdělení (např. část 1) a dosadíme příslušné hodnoty do vzorce pro modus tohoto rozdělení. Nejedná-li se o hustotu známého rozdělení, uchýlíme se k derivování. Často bývá velmi efektivní nepracovat přímo s hustotou p , ale její logaritmickou transformací. Logaritmus rozdělí známý tvar hustoty a multiplikativní konstantu na součet logaritmů a při následném derivování multiplikativní konstanta zmizí. Platí tedy

$$\frac{\partial}{\partial \boldsymbol{\theta}_j} \log p(\boldsymbol{\theta}|\mathbb{y}) = \frac{\partial}{\partial \boldsymbol{\theta}_j} \log p(\boldsymbol{\theta}_j|\mathbb{y}, \boldsymbol{\theta}_{(-j)}).$$

Z unimodality plně podmíněného rozdělení pak plyne, že řešení soustavy

$$\frac{\partial}{\partial \boldsymbol{\theta}_j} p(\boldsymbol{\theta}_j|\mathbb{y}, \boldsymbol{\theta}_{(-j)} = \boldsymbol{\theta}_{(-j)}^0) \stackrel{!}{=} \mathbf{0}, \quad \text{případně} \quad \frac{\partial}{\partial \boldsymbol{\theta}_j} \log p(\boldsymbol{\theta}_j|\mathbb{y}, \boldsymbol{\theta}_{(-j)} = \boldsymbol{\theta}_{(-j)}^0) \stackrel{!}{=} \mathbf{0},$$

musí být bodem, který maximalizuje hustotu p_j pro každé $j \in \{1, \dots, J\}$. Tyto soustavy je ale potom nutné řešit zároveň pro všechna j , což už nemusí dávat

explicitní řešení. Nelze-li řešení této soustavy nalézt v uzavřeném tvaru, je třeba se uchýlit k numerickým metodám maximalizace. Jednou z nich je právě tzv. ICM algoritmus (Iterated Conditional Modes Algorithm).

Tento ICM algoritmus již nespadá do MCMC metod. Jedná se o metodu, jak v unimodálním rozdělení o vícero složkách nalézt modus. Používá se zde deterministického postupu, nepoužívá se zde žádné generování náhodných elementů. V podstatě se jedná o klasickou úlohu maximalizace funkce více proměnných, jen přeformulované přímo pro aposteriorní hustotu.

Algoritmus 2. *ICM algoritmus.*

Vstupy: data \mathbf{y} , počáteční hodnota $\boldsymbol{\theta}^0$, aposteriorní hustota $p(\boldsymbol{\theta}|\mathbf{y})$, metody maximalizace plně podmíněných rozdělení.

Výstup: hodnota $\tilde{\boldsymbol{\theta}}$ maximalizující $p(\boldsymbol{\theta}|\mathbf{y})$.

Postup: v $(i+1)$ -ním kroku algoritmu uprav současnou nejbližší hodnotu $\tilde{\boldsymbol{\theta}}^i$ na hodnotu novou $\tilde{\boldsymbol{\theta}}^{i+1}$ po složkách následovně:

$$\begin{aligned}
1) \quad \tilde{\boldsymbol{\theta}}_1^{i+1} &= \arg \max_{\boldsymbol{\theta}_1} p\left(\boldsymbol{\theta}_1 \mid \mathbf{y}, \boldsymbol{\theta}_{(-1)} = \tilde{\boldsymbol{\theta}}_{(-1)}^i\right) = \text{mod} \left[\boldsymbol{\theta}_1 \mid \mathbb{Y} = \mathbf{y}, \boldsymbol{\theta}_{(-1)} = \tilde{\boldsymbol{\theta}}_{(-1)}^i \right], \\
2) \quad \tilde{\boldsymbol{\theta}}_2^{i+1} &= \arg \max_{\boldsymbol{\theta}_2} p\left(\boldsymbol{\theta}_2 \mid \mathbf{y}, \boldsymbol{\theta}_1 = \tilde{\boldsymbol{\theta}}_1^{i+1}, \boldsymbol{\theta}_3 = \tilde{\boldsymbol{\theta}}_3^i, \dots, \boldsymbol{\theta}_J = \tilde{\boldsymbol{\theta}}_J^i\right) \\
&= \text{mod} \left[\boldsymbol{\theta}_2 \mid \mathbb{Y} = \mathbf{y}, \boldsymbol{\theta}_1 = \tilde{\boldsymbol{\theta}}_1^{i+1}, \boldsymbol{\theta}_3 = \tilde{\boldsymbol{\theta}}_3^i, \dots, \boldsymbol{\theta}_J = \tilde{\boldsymbol{\theta}}_J^i \right], \\
&\vdots \\
J-1) \quad \tilde{\boldsymbol{\theta}}_{J-1}^{i+1} &= \arg \max_{\boldsymbol{\theta}_{J-1}} p\left(\boldsymbol{\theta}_{J-1} \mid \mathbb{Y} = \mathbf{y}, \boldsymbol{\theta}_1 = \tilde{\boldsymbol{\theta}}_1^{i+1}, \dots, \boldsymbol{\theta}_{J-2} = \tilde{\boldsymbol{\theta}}_{J-2}^{i+1}, \boldsymbol{\theta}_J = \tilde{\boldsymbol{\theta}}_J^i\right) \\
&= \text{mod} \left[\boldsymbol{\theta}_{J-1} \mid \mathbb{Y} = \mathbf{y}, \boldsymbol{\theta}_1 = \tilde{\boldsymbol{\theta}}_1^{i+1}, \dots, \boldsymbol{\theta}_{J-2} = \tilde{\boldsymbol{\theta}}_{J-2}^{i+1}, \boldsymbol{\theta}_J = \tilde{\boldsymbol{\theta}}_J^i \right], \\
J) \quad \tilde{\boldsymbol{\theta}}_J^{i+1} &= \arg \max_{\boldsymbol{\theta}_J} p\left(\boldsymbol{\theta}_J \mid \mathbf{y}, \boldsymbol{\theta}_{(-J)} = \tilde{\boldsymbol{\theta}}_{(-J)}^{i+1}\right) = \text{mod} \left[\boldsymbol{\theta}_J \mid \mathbb{Y} = \mathbf{y}, \boldsymbol{\theta}_{(-J)} = \tilde{\boldsymbol{\theta}}_{(-J)}^{i+1} \right].
\end{aligned}$$

Konec algoritmu: jakmile rozdíl (měřený nějakou vzdáleností na Θ) mezi $\tilde{\boldsymbol{\theta}}^i$ a $\tilde{\boldsymbol{\theta}}^{i+1}$ bude menší než předem stanovená hranice přesnosti. Za $\tilde{\boldsymbol{\theta}}$ zvol poslední $\tilde{\boldsymbol{\theta}}^{i+1}$.

V literatuře (viz např. O'Hagan a Forster, 2004, paragrafy 9.36–9.38) se někdy tomuto algoritmu říká také Lindleyova-Smithova iterační metoda. Byla totiž původně představena v článku Lindley a Smith (1972, závěr části 4).

Není nutné dávat algoritmu aposteriorní hustotu v přesném tvaru, stačí až na multiplikativní konstantu nezávislou na $\boldsymbol{\theta}$ – neprojeví se na maximalizaci. V tomto algoritmu jsme schopni přesně kvantifikovat míru konvergence pomocí porovnávání posledních dvou iterací. Jelikož výpočet vzdálenosti mezi $\tilde{\boldsymbol{\theta}}^i$ a $\tilde{\boldsymbol{\theta}}^{i+1}$ může být výpočetně náročný, tak se doporučuje počítat tento rozdíl jen pro každou K -tou hodnotu, kde K si zvolíme podle očekávané složitosti nalezení maxima.

Nevýhodou však zůstává, že je tento postup aplikovatelný jedině v případě, že máme zaručeno, že je aposteriorní rozdělení unimodální. V opačném případě by tento algoritmus nemusel dokonvergovat ke globálnímu maximu, ale pouze k nějakému lokálnímu. Bod, ke kterému dokonvergujeme, je vesměs určen výchozím bodem $\boldsymbol{\theta}^0$, doporučuje se tedy algoritmus spustit pro různé počáteční hodnoty a ověřit si tak, že vždy dojdeme ke stejnému výsledku.

2.4 Bayesovská mnohorozměrná regrese

Ve stěžejní části práce se budeme věnovat modelu bayesovské faktorové analýzy, který, jak se ukáže, je zobecněním modelu bayesovské mnohorozměrné regrese. To právě z tohoto jednoduššího modelu dělá ideálního adepta na předvedení bayesovských metod představených v této kapitole. Později tak budeme moct ukázat, v čem je odhad modelu bayesovské faktorové analýzy náročnější.

Model klasické lineární regrese modeluje skalární náhodnou veličinu Y při známém náhodném vektoru parametrů \mathbf{X} pomocí lineární kombinace $\boldsymbol{\beta}^\top \mathbf{X}$ a náhodné chyby modelu ε s nulovou střední hodnotou, tj. $Y = \boldsymbol{\beta}^\top \mathbf{X} + \varepsilon$. Model mnohorozměrné regrese modeluje m -rozměrný náhodný vektor \mathbf{Y} tak, že každá jeho složka Y_j je modelována opět jako nějaká lineární kombinace daného regresoru \mathbf{X} a chybového členu ε_j s nulovou střední hodnotou:

$$Y_j = \boldsymbol{\beta}_j^\top \mathbf{X} + \varepsilon_j, \quad j \in \{1, \dots, m\}.$$

Každá složka však má jiné koeficienty $\boldsymbol{\beta}_j = (\beta_{j,0}, \dots, \beta_{j,k})^\top$ a jednotlivé skalární chyby ε_j mezi sebou nemusí být nezávislé. Konkrétně vyjadřujeme

$$\underbrace{\mathbf{Y}}_{m \times 1} = \underbrace{\mathbb{B}}_{m \times (k+1)} \underbrace{\mathbf{X}}_{(k+1) \times 1} + \underbrace{\boldsymbol{\varepsilon}}_{m \times 1}, \quad \text{kde } \mathbb{B} = \begin{pmatrix} \boldsymbol{\beta}_1^\top \\ \vdots \\ \boldsymbol{\beta}_m^\top \end{pmatrix} = \begin{pmatrix} \beta_{1,0} & \cdots & \beta_{1,k} \\ \vdots & \ddots & \vdots \\ \beta_{m,0} & \cdots & \beta_{m,k} \end{pmatrix}, \quad (2.5)$$

tedy j -tá složka vektoru \mathbf{Y} je modelována jako $Y_j = \boldsymbol{\beta}_j^\top \mathbf{X} + \varepsilon_j$. Typicky je první složka vektoru \mathbf{X} rovna 1 a příslušné koeficienty $\beta_{j,0}$ tak mají význam střední hodnoty Y_j za předpokladu, že ostatní složky regresoru \mathbf{X} jsou nulové.

O vektoru chyb modelu $\boldsymbol{\varepsilon} = \mathbf{Y} - \mathbb{B}\mathbf{X}$ budeme předpokládat, že se řídí m -rozměrným normálním rozdělením s nulovou střední hodnotou a varianční maticí $\boldsymbol{\Sigma} > 0$, tedy $\boldsymbol{\varepsilon} \sim \mathbf{N}_m(\mathbf{0}_m, \boldsymbol{\Sigma})$. Potom rozdělení vektoru \mathbf{Y} při znalosti vektoru \mathbf{X} je opět mnohorozměrné normální, neboť $\mathbb{B}\mathbf{X}$ se považuje za konstantu, tedy $\mathbf{Y}|\mathbf{X} \sim \mathbf{N}_m(\mathbb{B}\mathbf{X}, \boldsymbol{\Sigma})$. Pravděpodobnostní mechanismus stojící za \mathbf{X} nás v tomto případě nezajímá, předpokládáme totiž, že jeho rozdělení na \mathbb{B} a $\boldsymbol{\Sigma}$ nezávisí. Soustředíme se tak pouze na podmíněné rozdělení $\mathbf{Y}|\mathbf{X}$.

Budeme pracovat s n nezávislými dvojicemi náhodných vektorů $(\mathbf{Y}_1, \mathbf{X}_1), \dots, (\mathbf{Y}_n, \mathbf{X}_n)$ z rozdělení stejného jako má (\mathbf{Y}, \mathbf{X}) . Zavedme si následující matice:

$$\underbrace{\mathbb{Y}}_{n \times m} = \begin{pmatrix} \mathbf{Y}_1^\top \\ \vdots \\ \mathbf{Y}_n^\top \end{pmatrix}, \quad \underbrace{\mathbb{X}}_{n \times (k+1)} = \begin{pmatrix} \mathbf{X}_1^\top \\ \vdots \\ \mathbf{X}_n^\top \end{pmatrix}, \quad \underbrace{\mathbb{E}}_{n \times m} = \begin{pmatrix} \boldsymbol{\varepsilon}_1^\top \\ \vdots \\ \boldsymbol{\varepsilon}_n^\top \end{pmatrix}. \quad (2.6)$$

Sloupec j matice \mathbb{Y} tak obsahuje všech n dostupných j -tých složek vektorů \mathbf{Y}_i . Potom dáme-li dohromady rovnice (2.5) pro každé $i \in \{1, \dots, n\}$, tak platí

$$\underbrace{\mathbb{Y}}_{n \times m} = \underbrace{\mathbb{X}}_{n \times (k+1)} \underbrace{\mathbb{B}^\top}_{(k+1) \times m} + \underbrace{\mathbb{E}}_{n \times m}. \quad (2.7)$$

Věrohodnost parametrů \mathbb{B} a Σ tak má tvar

$$\begin{aligned} p(\mathbb{Y}|\mathbb{X}, \mathbb{B}, \Sigma) &\propto \prod_{i=1}^n |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{Y}_i - \mathbb{B}\mathbf{X}_i)^\top \Sigma^{-1} (\mathbf{Y}_i - \mathbb{B}\mathbf{X}_i) \right\} \\ &\propto |\Sigma|^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \text{Tr} (\mathbf{Y}_i - \mathbb{B}\mathbf{X}_i)^\top \Sigma^{-1} (\mathbf{Y}_i - \mathbb{B}\mathbf{X}_i) \right\} \\ &\propto |\Sigma|^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \text{Tr} \Sigma^{-1} (\mathbb{Y} - \mathbb{X}\mathbb{B}^\top)^\top (\mathbb{Y} - \mathbb{X}\mathbb{B}^\top) \right\}. \end{aligned} \quad (2.8)$$

Srovnáním s hustotou (1.5) vidíme, že $\mathbb{Y}|\mathbb{X} \sim \mathbf{N}_{n \times m}(\mathbb{X}\mathbb{B}^\top, \mathbb{I}_n \otimes \Sigma)$. Jak je obecně zvykem v lineární regresi, budeme předpokládat, že \mathbb{X} je skoro jistě matice plně sloupcové hodnosti.

Nyní je na čase zapojit do tohoto modelu bayesovský přístup. Nechtě tedy parametry \mathbb{B} a Σ jsou náhodné a jejich apriorní rozdělení je konjugované k maticovému normálnímu rozdělení. To znamená volit $\mathbb{B} \sim \mathbf{N}_{m \times (k+1)}(\mathbb{B}_0, \Sigma \otimes \mathbb{D})$ a $\Sigma \sim \text{IW}_m(\mathbb{Q}, \nu)$, tedy jejich hustoty jsou

$$\begin{aligned} p(\mathbb{B}|\Sigma; \mathbb{B}_0, \mathbb{D}) &\propto |\mathbb{D}|^{-\frac{m}{2}} |\Sigma|^{-\frac{k+1}{2}} \exp \left\{ -\frac{1}{2} \text{Tr} \Sigma^{-1} (\mathbb{B} - \mathbb{B}_0) \mathbb{D}^{-1} (\mathbb{B} - \mathbb{B}_0)^\top \right\}, \\ p(\Sigma|\mathbb{Q}, \nu) &\propto |\mathbb{Q}|^{-\frac{\nu-m-1}{2}} |\Sigma|^{-\frac{\nu}{2}} \exp \left\{ -\frac{1}{2} \text{Tr} \Sigma^{-1} \mathbb{Q} \right\}. \end{aligned} \quad (2.9)$$

O hyperparametrech $\boldsymbol{\alpha} = \text{vec}(\mathbb{B}_0, \mathbb{D}, \mathbb{Q}, \nu)$ budeme předpokládat, že již jsou pevně stanoveny předem, a budeme na ně nahlížet jako na konstanty. Dále je již pro zjednodušení zápisu nebudeme vypisovat do příslušných hustot.

Dle Bayesovy věty 2 vznikne aposteriorní hustota parametrů \mathbb{B} a Σ součinem jejich věrohodnosti (2.8) a apriorní hustoty (2.9):

$$\begin{aligned} p(\mathbb{B}, \Sigma|\mathbb{Y}, \mathbb{X}) &\propto p(\mathbb{Y}|\mathbb{X}, \mathbb{B}, \Sigma) \cdot p(\mathbb{B}|\Sigma) \cdot p(\Sigma) \propto |\Sigma|^{-\frac{n+\nu+k+1}{2}} \\ &\cdot \exp \left\{ -\frac{1}{2} \text{Tr} \Sigma^{-1} \left[(\mathbb{Y} - \mathbb{X}\mathbb{B}^\top)^\top (\mathbb{Y} - \mathbb{X}\mathbb{B}^\top) + (\mathbb{B} - \mathbb{B}_0) \mathbb{D}^{-1} (\mathbb{B} - \mathbb{B}_0)^\top + \mathbb{Q} \right] \right\}. \end{aligned} \quad (2.10)$$

Vzhledem k tomu, že používáme konjugovaného rozdělení, tak aposteriorní rozdělení je ze stejné rodiny pravděpodobnostních rozdělení jako rozdělení apriorní.

Tvrzení 4. *V modelu bayesovské mnohorozměrné regrese nastíněném výše platí*

$$\begin{aligned} \mathbb{B}|\mathbb{Y}, \mathbb{X}, \Sigma &\sim \mathbf{N}_{m \times (k+1)} \left(\hat{\mathbb{B}}, \Sigma \otimes [\mathbb{X}^\top \mathbb{X} + \mathbb{D}^{-1}]^{-1} \right) \quad a \quad \Sigma|\mathbb{Y}, \mathbb{X} \sim \text{IW}_m(\Psi, n + \nu), \\ \text{kde } \hat{\mathbb{B}} &= (\mathbb{Y}^\top \mathbb{X} + \mathbb{B}_0 \mathbb{D}^{-1}) [\mathbb{X}^\top \mathbb{X} + \mathbb{D}^{-1}]^{-1}, \\ \Psi &= \mathbb{Q} + \mathbb{Y}^\top \mathbb{Y} - \hat{\mathbb{B}} [\mathbb{X}^\top \mathbb{X} + \mathbb{D}^{-1}] \hat{\mathbb{B}}^\top + \mathbb{B}_0 \mathbb{D}^{-1} \mathbb{B}_0^\top. \end{aligned}$$

Důkaz. Podle důsledku tvrzení B7 uvedeného v appendixu platí:

$$\begin{aligned} &(\mathbb{Y} - \mathbb{X}\mathbb{B}^\top)^\top (\mathbb{Y} - \mathbb{X}\mathbb{B}^\top) + (\mathbb{B} - \mathbb{B}_0) \mathbb{D}^{-1} (\mathbb{B} - \mathbb{B}_0)^\top = \\ &= (\mathbb{B} - \hat{\mathbb{B}}) [\mathbb{X}^\top \mathbb{X} + \mathbb{D}^{-1}] (\mathbb{B} - \hat{\mathbb{B}})^\top - \hat{\mathbb{B}} [\mathbb{X}^\top \mathbb{X} + \mathbb{D}^{-1}] \hat{\mathbb{B}}^\top + \mathbb{B}_0 \mathbb{D}^{-1} \mathbb{B}_0^\top + \mathbb{Y}^\top \mathbb{Y}, \end{aligned}$$

kde $\widehat{\mathbb{B}}$ je definováno jako výše ve znění tvrzení. Proto lze (2.10) rozdělit na

$$p(\mathbb{B}, \Sigma | \mathbb{Y}, \mathbb{X}) \propto |\Sigma|^{-\frac{k+1}{2}} \exp \left\{ -\frac{1}{2} \text{Tr} \Sigma^{-1} (\mathbb{B} - \widehat{\mathbb{B}}) [\mathbb{X}^\top \mathbb{X} + \mathbb{D}^{-1}] (\mathbb{B} - \widehat{\mathbb{B}})^\top \right\} \cdot |\Sigma|^{-\frac{n+\nu}{2}} \exp \left\{ \frac{1}{2} \text{Tr} \Sigma^{-1} \underbrace{[\mathbb{Q} + \mathbb{Y}^\top \mathbb{Y} - \widehat{\mathbb{B}} [\mathbb{X}^\top \mathbb{X} + \mathbb{D}^{-1}] \widehat{\mathbb{B}}^\top + \mathbb{B}_0 \mathbb{D}^{-1} \mathbb{B}_0]}_{\Psi} \right\}. \quad (2.11)$$

První řádek odpovídá hustotě rozdělení $\mathbb{B} | \mathbb{Y}, \mathbb{X}, \Sigma$, která ve srovnání s (1.5) je hustotou maticového normálního rozdělení s parametry uvedenými ve znění tvrzení. Druhý řádek se potom věnuje aposteriornímu rozdělení $\Sigma | \mathbb{Y}, \mathbb{X}$, kde rozpoznáváme hustotu (1.9) inverzního Wishartova rozdělení s Ψ a stupni volnosti $n + \nu$. \square

Za předpokladu plné sloupcové hodnosti matice \mathbb{X} můžeme použít také odhad $\widehat{\mathbb{B}}^* = \mathbb{Y}^\top \mathbb{X} (\mathbb{X}^\top \mathbb{X})^{-1}$ z klasické mnohorozměrné regrese a vidíme, že

$$\widehat{\mathbb{B}} = \left(\widehat{\mathbb{B}}^* (\mathbb{X}^\top \mathbb{X}) + \mathbb{B}_0 \mathbb{D}^{-1} \right) [\mathbb{X}^\top \mathbb{X} + \mathbb{D}^{-1}]^{-1}.$$

Tento bayesovský odhad je tedy konvexní kombinací odhadu $\widehat{\mathbb{B}}^*$ a apriorního předpokladu střední hodnoty \mathbb{B}_0 , podobně jako tomu bylo v příkladě 2.

Na základě tohoto tvrzení již můžeme odhadovat a provádět inferenci o parametru Σ , neboť již známe jeho marginální aposteriorní rozdělení. Pro inferenci o parametru \mathbb{B} potřebujeme znát aposteriorní rozdělení $\mathbb{B} | \mathbb{Y}, \mathbb{X}$ již nepodmíněné Σ .

Tvrzení 5. *Za značení zavedeného v tvrzení 4 je marginální aposteriorní rozdělení parametru \mathbb{B} v bayesovském modelu mnohorozměrné regrese dáno hustotou*

$$p(\mathbb{B} | \mathbb{Y}, \mathbb{X}) \propto \left| \Psi + (\mathbb{B} - \widehat{\mathbb{B}}) [\mathbb{X}^\top \mathbb{X} + \mathbb{D}^{-1}] (\mathbb{B} - \widehat{\mathbb{B}})^\top \right|^{-\frac{n+\nu-m-1+(k+1)}{2}},$$

$$\text{tedy } \mathbb{B} | \mathbb{Y}, \mathbb{X} \sim \mathbf{t}_{m \times (k+1)} \left(n + \nu - m - 1, \widehat{\mathbb{B}}, \frac{1}{n+\nu-m-1} \Psi \otimes [\mathbb{X}^\top \mathbb{X} + \mathbb{D}^{-1}]^{-1} \right).$$

Důkaz. Použijeme tvrzení 3, kde volíme za $\theta_1 = \mathbb{B}$ a $\theta_2 = \Sigma$. Podíváme-li se na sdruženou hustotu (2.10), tak vidíme, že v proměnné Σ připomíná tvaru hustoty inverzního Wishartova rozdělení, srovnej s hustotou (1.9). Jak snadno určíme, jedná se o plně podmíněné rozdělení

$$\Sigma | \mathbb{B}, \mathbb{Y}, \mathbb{X} \sim \text{IW}_m \left(\underbrace{\Psi + (\mathbb{B} - \widehat{\mathbb{B}}) [\mathbb{X}^\top \mathbb{X} + \mathbb{D}^{-1}] (\mathbb{B} - \widehat{\mathbb{B}})^\top}_{\Psi(\mathbb{B} | \mathbb{Y}, \mathbb{X})}, n + \nu + k + 1 \right).$$

Vidíme tedy, že onen maticový parametr $\Psi(\mathbb{B} | \mathbb{Y}, \mathbb{X})$ stále závisí na parametru \mathbb{B} . Proto je třeba použít za funkci k ve tvrzení 3 následující funkci

$$k(\mathbb{B} | \mathbb{Y}, \mathbb{X}) = |\Psi(\mathbb{B} | \mathbb{Y}, \mathbb{X})|^{-\frac{(n+\nu+k+1)-m-1}{2}} = |\Psi(\mathbb{B} | \mathbb{Y}, \mathbb{X})|^{-\frac{(n+\nu-m-1)+(k+1)}{2}}.$$

Jelikož funkcí g je v tomto případě libovolná konstanta, tak dle tvrzení 3 je $p(\mathbb{B}|\mathbb{Y}, \mathbb{X}) \propto k(\mathbb{B}|\mathbb{Y}, \mathbb{X})$. Nyní již srovnáme tento výsledek s hustotou (1.12) a uvědomíme si, že v tomto případě jsou rozměry m a $k+1$ a že počet stupňů volnosti je $n + \nu - m - 1$, který se musí vyskytnout ve jmenovateli u kvadratické formy s \mathbb{B} . Počet stupňů volnosti se stále počítá do proporcionalní konstanty, ovšem je třeba jej vytknout od Ψ , proto ve znění tvrzení používáme $\frac{1}{n+\nu-m-1}\Psi$. \square

Jelikož jsme schopni odvodit jednotlivá marginální aposteriorní rozdělení, tak není zapotřebí používat MCMC metody. Kdyby však někdo i přesto potřeboval sestrotit Gibbsův algoritmus, využil by plně podmíněných rozdělení

$$\begin{aligned} \mathbb{B}|\mathbb{Y}, \mathbb{X}, \Sigma &\sim N_{m \times (k+1)} \left(\hat{\mathbb{B}}, \Sigma \otimes [\mathbb{X}^\top \mathbb{X} + \mathbb{D}^{-1}]^{-1} \right), \\ \Sigma|\mathbb{Y}, \mathbb{X}, \mathbb{B} &\sim IW_m \left(\Psi + (\mathbb{B} - \hat{\mathbb{B}}) [\mathbb{X}^\top \mathbb{X} + \mathbb{D}^{-1}] (\mathbb{B} - \hat{\mathbb{B}})^\top, n + \nu + k + 1 \right), \end{aligned} \quad (2.12)$$

jejichž tvar plyne z tvrzení 4 a důkazu tvrzení 5.

Ze symetrie normálního a Studentova rozdělení plyne, že odhad \mathbb{B} založený na aposteriorním modu musí být totožný s tím založeným na aposteriorní střední hodnotě. Z tvaru aposteriorní hustoty (2.11) lze také okamžitě vidět, že volbou $\mathbb{B} = \hat{\mathbb{B}}$ dosáhneme maxima první části aposteriorní hustoty (vynulováním exponentu) nehledě na to, jaké je Σ (lze ověřit i derivováním, viz tvrzení A5 bod vii). Ovšem u parametru Σ se již tyto dva druhy odhadu musí lišit, viz např. tabulka 1.1 ohledně rozdílu mezi střední hodnotou a modem inverzního Wishartova rozdělení. Jelikož se snažíme maximalizovat aposteriorní hustotu a přitom už víme, že $\hat{\mathbb{B}}$ maximalizuje v \mathbb{B} pro každé Σ , tak stačí maximalizovat hustotu plně podmíněného rozdělení $\Sigma|\mathbb{B}, \mathbb{Y}, \mathbb{X} \sim IW_m(\Psi(\mathbb{B}|\mathbb{Y}, \mathbb{X}), n + \nu + k + 1)$. Proto odhadem Σ na základě aposteriorního modu je

$$\tilde{\Sigma} = \frac{\Psi(\hat{\mathbb{B}}|\mathbb{Y}, \mathbb{X})}{n + \nu + k + 1} = \frac{\Psi}{n + \nu + k + 1}, \quad \text{kdežto } \hat{\Sigma} = \frac{\Psi}{n + \nu - 2m - 2}$$

je odhad na základě marginální aposteriorní střední hodnoty, viz tvrzení 4.

Tyto heuristické myšlenky lze podpořit derivací logaritmu aposteriorní hustoty (logaritmus zachová bod maxima, multiplikativní konstanta z proporcionality derivováním zmizí). Použijeme tvrzení A5 body iv) a vi) pro symetrickou pozitivně definitní matici Σ následovně

$$\begin{aligned} \left. \frac{\partial \log p(\mathbb{B}, \Sigma|\mathbb{Y}, \mathbb{X})}{\partial \Sigma} \right|_{\mathbb{B}=\hat{\mathbb{B}}} &= \frac{\partial}{\partial \Sigma} \left(-\frac{n + \nu + k + 1}{2} \log |\Sigma| - \frac{1}{2} \text{Tr} \Sigma^{-1} \Psi(\hat{\mathbb{B}}|\mathbb{Y}, \mathbb{X}) \right) \\ &= -\frac{n + \nu + k + 1}{2} (2\Sigma^{-1} - \text{diag} \Sigma^{-1}) + \frac{1}{2} [2\Sigma^{-1} \Psi \Sigma^{-1} - \text{diag} (\Sigma^{-1} \Psi \Sigma^{-1})]. \end{aligned}$$

Chceme najít $\tilde{\Sigma}$ jako takovou hodnotu, která nuluje tuto derivaci. Lze si jednoduše uvědomit, že jak mimo diagonálu, tak na ní je třeba splnit rovnici

$$(n + \nu + k + 1)\tilde{\Sigma}^{-1} = \tilde{\Sigma}^{-1} \Psi \tilde{\Sigma}^{-1},$$

z čehož už jasně plyne, že musí $\tilde{\Sigma} = \Psi / (n + \nu + k + 1)$.

3. Bayesovská faktorová analýza

Nyní již máme vše připraveno a můžeme začít s modelem bayesovské faktorové analýzy. V této kapitole se podrobněji podíváme na model klasické faktorové analýzy, který byl představen v samotném úvodu. Postupně aplikujeme metody uvedené v předchozí kapitole a získáme tak bayesovské metody odhadu tohoto modelu, a to za různých předpokladů apriorního rozdělení. Ukážeme si také jednoduchá zobecnění tohoto modelu.

3.1 Faktorová analýza

Zaměříme se nejprve na klasický model faktorové analýzy, který byl představen v úvodu. O samotné faktorové analýze byla napsána celá řada odborných knih a článků (viz Harmann, 1976), zde se však omezíme na definici modelu a hlavní problémy, které je třeba při jeho odhadu řešit.

Snažíme se rozložit m -rozměrný vektor \mathbf{Y} na součet tří jiných vektorů dle rovnice (1). První složkou je vektor $\boldsymbol{\mu}$ popisující střední hodnotu $\boldsymbol{\mu} = \mathbf{E}\mathbf{Y}$. Další část tvoří m lineárních kombinací k latentních (nám skrytých) faktorů $\mathbf{F} = (F_1, \dots, F_k)^\top$, kde k je mnohem menší než počet složek m pozorovaného vektoru \mathbf{Y} . Tyto lineární kombinace jsou plně popsány maticí faktorových zátěží $\boldsymbol{\Lambda}$. Zbylou částí tohoto rozkladu je vektor náhodné chyby $\boldsymbol{\varepsilon}$, který vyjadřuje odchylku \mathbf{Y} od našeho předpokládaného modelu.

Pro j -tou složku vektoru \mathbf{Y} platí $Y_j = \mu_j + \boldsymbol{\lambda}_j^\top \mathbf{F} + \varepsilon_j$, kde $\boldsymbol{\Lambda}^\top = (\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_m)$, tedy řádek $\boldsymbol{\lambda}_j$ matice zátěží udává lineární kombinaci latentních faktorů, která spolu se střední hodnotou μ_j tvoří základ veličiny Y_j . Jednotlivé zátěže λ_{jl} pak udávají, jakým způsobem a jak moc přispívá l -tý faktor F_l do j -té složky Y_j .

Vidíme zde tak velkou příbuznost s modelem bayesovské regrese, viz část 2.4. Pokud zvolíme vektor regresorů $\mathbf{X} = (1, \mathbf{F}^\top)^\top$ a sloučíme neznámé parametry $\boldsymbol{\mu}$ a $\boldsymbol{\Lambda}$ do matice $\mathbb{B} = (\boldsymbol{\mu}, \boldsymbol{\Lambda})$, tak lze rovnici rozkladu faktorové analýzy přepsat následovně

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{F} + \boldsymbol{\varepsilon} = \mathbb{B}\mathbf{X} + \boldsymbol{\varepsilon}.$$

Zásadním rozdílem oproti modelu bayesovské regrese je ten, že *neznáme* hodnotu regresoru \mathbf{X} , neznáme vlastně ani jeden ze tří vektorů, na které \mathbf{Y} rozkládáme. Snažíme se odhadnout jak parametry $\boldsymbol{\mu}$ a $\boldsymbol{\Lambda}$, tak i samotné faktory \mathbf{F} . Cílem je zredukovat informaci obsaženou v m složkách vektoru \mathbf{Y} na k faktorů, na nichž lze posléze stavět další modely a provádět hlubší analýzy. Tato metodika je hojně využívána v psychologii, kde slouží k definování abstraktních (ne přímo měřitelných) veličin.

Příklad 3 (Politický kompas). V předvolebním období bývá velkým hitem takzvaná *politická kalkulačka* či *politický kompas*. Častým modelem pro politické rozhodování voliče je model faktorové analýzy založený na dvou abstraktních veličinách (latentních faktorech): míra ekonomické svobody (dělení na levici a pravici) a míra osobní svobody (dělení na konzervativitu a liberalitu). Tyto míry však nelze na člověku přímo změřit, lze je pouze otestovat. Nerozhodnutý volič dostane sadu m (řádů desítek) politických prohlášení či názorů a jeho úkolem je na zvolené škále odpovědět, do jaké míry s tímto výrokem souhlasí. Škála může

být zavedená od elementární úrovně (nesouhlasím, nevím či souhlasím), až po číselnou (např. mezi 0 a 100). Poté, co volič vyplní dotazník, jsou jeho odpovědi zpracovány a kalkulačka mu spočte jeho míry ekonomické a osobní svobody, které zakreslí jako souřadnice do obrázku. Může svoje výsledky porovnat s tím, jak by v tomto testu dopadly jednotlivé politické strany, a může se tak lépe rozhodnout, komu dát svůj hlas. Ukázkou takového výstupu je obrázek 3.1, kde můžeme vidět hodnoty obou skrytých faktorů nejen nejsilnějších politických stran v roce 2017, ale také prvních tří prezidentů České republiky.

Můžeme tak tento problém napasovat přímo na model faktorové analýzy se dvěma skrytými faktory $\mathbf{F} = (F_1, F_2)$. Odpověď na j -tou otázku Y_j dle modelu faktorové analýzy vzniká z obecné střední hodnoty odpovědi μ_j , dále především konkrétními faktory, jejichž vztah k Y_j je popsán pomocí faktorových zátěží $\boldsymbol{\lambda}_j = (\lambda_{j1}, \lambda_{j2})$, a také chybou modelu ε_j . Ta vzniká jedinečností názorů a specifikami každého člověka, které nemusí ve všech směrech dodržovat předepsanou závislost na mírách svobod. Kalkulačka musí na základě historických dat spočítat odhady parametrů $\boldsymbol{\mu}$ a $\boldsymbol{\Lambda}$ a s jejich pomocí určit odhady \hat{F}_1 a \hat{F}_2 skutečných měř svobody F_1 a F_2 .

◇

3.1.1 Předpoklady a interpretace

Ujasněme si základní předpoklady modelu faktorové analýzy, které již byly naznačeny v úvodu. Aby parametr $\boldsymbol{\mu}$ byl jednoznačně určen, tak musíme předpokládat nejen nulovou střední hodnotu chyby modelu ($\mathbf{E} \boldsymbol{\varepsilon} = \mathbf{0}_m$), ale i nulovou střední hodnotu faktorů ($\mathbf{E} \mathbf{F} = \mathbf{0}_k$). Latentní faktory F_1, \dots, F_k budou normované na jednotkový rozptyl a navzájem nekorelované, tj. $\text{var} \mathbf{F} = \mathbb{I}_k$. Typicky se předpokládá, že náhodné chyby modelu $\varepsilon_1, \dots, \varepsilon_m$ jsou mezi sebou také nekorelované (nyní už však ne nutně normované), tedy $\text{var} \boldsymbol{\varepsilon} = \boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \dots, \sigma_m^2)$. Později si ukážeme, že předpoklady vnitřní nekorelovanosti lze uvolnit. Nekorelovanost faktorů \mathbf{F} a chyby $\boldsymbol{\varepsilon}$ však budeme předpokládat vždy, tj. $\text{cov}(\mathbf{F}, \boldsymbol{\varepsilon}) = \mathbf{0}_{k \times m}$.



Obrázek 3.1: Míry ekonomické a osobní svobody prvních tří prezidentů České republiky dle politického kompasu pro volby na podzim roku 2017 (zdroj internet).

Předpokládáme-li navíc i normalitu $\mathbf{F} \sim \mathbf{N}_k(\mathbf{0}_k, \mathbb{I}_k)$ a $\boldsymbol{\varepsilon} \sim \mathbf{N}_m(\mathbf{0}_m, \boldsymbol{\Sigma})$, tak jsou \mathbf{F} a $\boldsymbol{\varepsilon}$ dokonce nezávislé a výsledný vektor \mathbf{Y} je též normálně rozdělený.

Z těchto předpokladů lze odvodit varianční strukturu samotného vektoru \mathbf{Y} :

$$\boldsymbol{\Upsilon} := \text{var } \mathbf{Y} = \text{var}(\boldsymbol{\Lambda}\mathbf{F} + \boldsymbol{\varepsilon}) = \boldsymbol{\Lambda}\mathbb{I}_k\boldsymbol{\Lambda}^\top + \boldsymbol{\Sigma} + \boldsymbol{\Lambda}\mathbb{O}_{k \times m} + \mathbb{O}_{m \times k}\boldsymbol{\Lambda}^\top = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Sigma}. \quad (3.1)$$

Při předpokladu normality tak dostáváme $\mathbf{Y} \sim \mathbf{N}_m(\boldsymbol{\mu}, \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Sigma})$, neboť normální rozdělení je plně specifikováno svými parametry střední hodnoty a rozptylu. Vidíme, že při diagonální $\boldsymbol{\Sigma}$ dostaneme nekorelované složky Y_j a Y_l , kde $j \neq l$, pouze tehdy, když $\boldsymbol{\lambda}_j^\top \boldsymbol{\lambda}_l = 0$, tedy pokud jsou na sebe příslušné faktorové zátěže kolmé. Rozptyl j -té složky pak má tvar $v_{j,j} = \text{var } Y_j = \boldsymbol{\lambda}_j^\top \boldsymbol{\lambda}_j + \sigma_j^2$. Části variability vysvětlené pomocí faktorů: $h_j^2 = \boldsymbol{\lambda}_j^\top \boldsymbol{\lambda}_j$ se říká *komunalita*, zbylé nevysvětlené části σ_j^2 se říká specifický rozptyl.

Podobně se lze podívat i na vzájemnou kovarianci mezi vektory \mathbf{Y} a \mathbf{F} :

$$\text{cov}(\mathbf{Y}, \mathbf{F}) = \text{cov}(\boldsymbol{\Lambda}\mathbf{F} + \boldsymbol{\varepsilon}, \mathbf{F}) = \boldsymbol{\Lambda}\mathbb{I}_k + \mathbb{O}_{m \times k} = \boldsymbol{\Lambda}. \quad (3.2)$$

Faktorová zátěž $\lambda_{j,l}$ tak dostává význam kovariance mezi měřenou veličinou Y_j a skrytým faktorem F_l . Pokud bychom zajistili, že pro každé $j \in \{1, \dots, m\}$ by $\text{var } Y_j = v_{j,j} = 1$, pak by dokonce $\lambda_{j,l}$ mělo význam korelace mezi Y_j a F_l , tedy $\lambda_{j,l} = \text{corr}(Y_j, F_l)$. Obecně však platí

$$\text{corr}(\mathbf{Y}, \mathbf{F}) = \begin{pmatrix} \boldsymbol{\lambda}_1^\top / \sqrt{\text{var } Y_1} \\ \vdots \\ \boldsymbol{\lambda}_m^\top / \sqrt{\text{var } Y_m} \end{pmatrix} = \left[\text{diag}(\boldsymbol{\lambda}_1^\top \boldsymbol{\lambda}_1 + \sigma_1^2, \dots, \boldsymbol{\lambda}_m^\top \boldsymbol{\lambda}_m + \sigma_m^2) \right]^{-\frac{1}{2}} \boldsymbol{\Lambda} \\ = [\text{diag } \boldsymbol{\Upsilon}]^{-\frac{1}{2}} \boldsymbol{\Lambda}.$$

Za předpokladu normality potom platí, že

$$\begin{pmatrix} \mathbf{Y} - \boldsymbol{\mu} \\ \mathbf{F} \end{pmatrix} \sim \mathbf{N}_{m+k} \left(\mathbf{0}_{m+k}, \begin{pmatrix} \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Sigma} & \boldsymbol{\Lambda} \\ \boldsymbol{\Lambda}^\top & \mathbb{I}_k \end{pmatrix} \right).$$

Z vlastnosti (1.4) normálního rozdělení pak můžeme vyjádřit podmíněné rozdělení

$$\mathbf{F} | \mathbf{Y} - \boldsymbol{\mu} \sim \mathbf{N}_k \left(\boldsymbol{\Lambda}^\top [\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Sigma}]^{-1} (\mathbf{Y} - \boldsymbol{\mu}), \mathbb{I}_k - \boldsymbol{\Lambda}^\top [\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Sigma}]^{-1} \boldsymbol{\Lambda} \right). \quad (3.3)$$

Této vlastnosti se potom používá k odhadu skryté hodnoty faktorů \mathbf{F} pomocí podmíněné střední hodnoty \mathbf{F} při znalosti \mathbf{Y} , kam za $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\Sigma}$ dosadíme spočtené odhady.

Uvědomme si, že pokud u náhodných veličin Y_j změním měřítko na $c_j Y_j$, kde c_j jsou známé konstanty, které tvoří diagonální matici $\mathbb{C} = \text{diag}(c_1, \dots, c_m)$, tak tyto veličiny opět splňují předpoklady modelu faktorové analýzy:

$$\mathbf{Y}_{\mathbb{C}} = \mathbb{C}\mathbf{Y} = \underbrace{\mathbb{C}\boldsymbol{\mu}}_{\boldsymbol{\mu}_{\mathbb{C}}} + \underbrace{\mathbb{C}\boldsymbol{\Lambda}}_{\boldsymbol{\Lambda}_{\mathbb{C}}} \mathbf{F} + \underbrace{\mathbb{C}\boldsymbol{\varepsilon}}_{\boldsymbol{\varepsilon}_{\mathbb{C}}}.$$

Chyba $\boldsymbol{\varepsilon}_{\mathbb{C}}$ má opět nulovou střední hodnotu a diagonální rozptylovou matici $\text{var } \boldsymbol{\varepsilon}_{\mathbb{C}} = \mathbb{C}\boldsymbol{\Sigma}\mathbb{C}$. Faktory \mathbf{F} zůstávají stejné a jsou stále nekorelované s $\boldsymbol{\varepsilon}_{\mathbb{C}}$.

Jak již bylo naznačeno v úvodu, rozklad (1) není jednoznačný. Uvažme libovolnou ortonormální matici $\mathbb{G} \in \mathbb{R}^{k \times k}$, tedy matici s vlastností $\mathbb{G}\mathbb{G}^\top = \mathbb{I}_k$. Potom lze také zapsat, že

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbb{I}_k\mathbf{F} + \boldsymbol{\varepsilon} = \boldsymbol{\mu} + \underbrace{\boldsymbol{\Lambda}\mathbb{G}}_{\boldsymbol{\Lambda}^*} \underbrace{\mathbb{G}^\top\mathbf{F}}_{\mathbf{F}^*} + \boldsymbol{\varepsilon}.$$

Jednoduše se lze přesvědčit, že \mathbf{F}^* má nulovou střední hodnotu, jeho rozptylová matice je rovna $\mathbb{G}\mathbb{I}_k\mathbb{G}^\top = \mathbb{I}_k$ a je stále nekorelovaný s $\boldsymbol{\varepsilon}$. Tedy jakákoliv volba ortonormální \mathbb{G} je matematicky správná, ovšem každá pracuje s jiným vektorem faktorů \mathbf{F} a také jinými vztahy mezi \mathbf{F} a \mathbf{Y} . Aplikuje-li tak faktorovou analýzu vícero lidí, tak každý může dojít k jiné interpretaci výsledků.

Matice \mathbb{G} reprezentuje nějakou rotaci v k -rozměrném prostoru. Typicky se model odhadne pro nějakou konkrétní rotaci a pro interpretaci se následně zvolí taková rotace \mathbb{G} , která splňuje nějakou předem stanovenou vlastnost. Hojně používanou rotací je rotace typu *varimax*. Jedná se o ortonormální matici \mathbb{G} maximalizující součet přes všechny sloupce $\boldsymbol{\Lambda}\mathbb{G}$ „odhadů rozptylů“ druhých mocnin nově vzniklých faktorových zátěží v rámci uvažovaného sloupce. Tento postup je lépe zřejmý z matematického zápisu

$$\mathbb{G}^{\text{varimax}} = \arg \max_{\mathbb{G}: \mathbb{G}\mathbb{G}^\top = \mathbb{I}_k} \sum_{l=1}^k \left[\frac{1}{m} \sum_{j=1}^m (\boldsymbol{\Lambda}\mathbb{G})_{j,l}^4 - \left(\frac{1}{m} \sum_{j=1}^m (\boldsymbol{\Lambda}\mathbb{G})_{j,l}^2 \right) \right].$$

Výsledkem této volby je taková rotace, kdy výsledné faktorové zátěže $\lambda_{j,l}^{\text{varimax}}$ jsou buď velmi blízké nule (veličina Y_j je téměř nekorelovaná s faktorem F_l), nebo naopak velmi vysoké (je zde silná korelace mezi Y_j a F_l). Veličina Y_j je pak vysoce korelována s co nejméně faktory a faktory jsou významně korelovány s vybranými veličinami Y_j a přitom minimálně s ostatními. Nejčastější interpretací tak často bývá rozdělení složek vektoru \mathbf{Y} na skupiny podle toho, s jakým faktorem jsou nejvíce korelovány.

Varimaxová rotace ale není jedinou používanou možností. Dále například rotace *quartimax* je dána ortonormální maticí takovou, která minimalizuje počet faktorů nutných k vysvětlení jednotlivých složek vektoru \mathbf{Y} . Rotace *equimax* je pak kompromisem mezi rotacemi varimax a quartimax.

Dále existují i *křivé* rotace, které mají opět hezké interpretační vlastnosti. Nicméně se již nejedná o ortonormální matice, a tedy je třeba mít na paměti, že se zde může vytratit nekorelovanost jednotlivých složek \mathbf{F} či jejich jednotkový rozptyl. Příkladem jsou rotace *oblimin* či *promax*. Více o těchto metodách viz (Harmann, 1976).

3.1.2 Odhad modelu faktorové analýzy

Model faktorové analýzy budeme odhadovat za pomoci náhodného výběru $\mathbf{Y}_1, \dots, \mathbf{Y}_n$, tedy n nezávislých náhodných vektorů ze stejného rozdělení daného předpokládaným modelem. Za těmito vektory stojí latentní náhodné vektory $\mathbf{F}_1, \dots, \mathbf{F}_n$ a $\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n$. Zavedme si následující značení:

$$\mathbb{Y} = \begin{pmatrix} \mathbf{Y}_1^\top \\ \vdots \\ \mathbf{Y}_n^\top \end{pmatrix}, \quad \mathbb{F} = \begin{pmatrix} \mathbf{F}_1^\top \\ \vdots \\ \mathbf{F}_n^\top \end{pmatrix}, \quad \mathbb{E} = \begin{pmatrix} \boldsymbol{\varepsilon}_1^\top \\ \vdots \\ \boldsymbol{\varepsilon}_n^\top \end{pmatrix}.$$

Potom lze model faktorové analýzy zapsat v maticovém tvaru

$$\underbrace{\mathbb{Y}}_{n \times m} = \underbrace{\mathbf{1}_n}_{n \times 1} \underbrace{\boldsymbol{\mu}^\top}_{1 \times m} + \underbrace{\mathbb{F}}_{n \times k} \underbrace{\boldsymbol{\Lambda}^\top}_{k \times m} + \underbrace{\mathbb{E}}_{n \times m}. \quad (3.4)$$

Použijeme-li značení $\mathbb{B} = (\boldsymbol{\mu}, \boldsymbol{\Lambda})$, $\mathbf{X}_i = (\mathbf{1}, \mathbf{F}^\top)^\top$ a $\mathbb{X} = (\mathbf{1}_n, \mathbb{F})$, tak lze přepsat rovnici (3.4) do tvaru modelu mnohorozměrné regrese (2.7). Ovšem připomínáme, že nyní je zde rozdíl v tom, že matici \mathbb{F} *neznáme*.

Za předpokladu normality po vzoru rovnice (2.8) platí, že

$$p(\mathbb{Y}|\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \text{Tr} \boldsymbol{\Sigma}^{-1} \cdot \left(\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top - \mathbb{F} \boldsymbol{\Lambda}^\top \right)^\top \left(\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top - \mathbb{F} \boldsymbol{\Lambda}^\top \right) \right\}, \quad (3.5)$$

$$p(\mathbb{Y}|\mathbb{X}, \mathbb{B}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \text{Tr} \boldsymbol{\Sigma}^{-1} \left(\mathbb{Y} - \mathbb{X} \mathbb{B}^\top \right)^\top \left(\mathbb{Y} - \mathbb{X} \mathbb{B}^\top \right) \right\}. \quad (3.6)$$

Jelikož za předpokladu normality také platí, že rozdělení \mathbf{Y} nepodmíněně znalostí \mathbf{F} je $\mathbb{N}_m(\boldsymbol{\mu}, \boldsymbol{\Upsilon})$, tak lze také psát, že

$$p(\mathbb{Y}|\boldsymbol{\mu}, \boldsymbol{\Upsilon}) \propto |\boldsymbol{\Upsilon}|^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \text{Tr} \boldsymbol{\Upsilon}^{-1} \left(\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top \right)^\top \left(\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top \right) \right\}. \quad (3.7)$$

Použijeme-li dodatečný rozklad uvedený v důkazu důsledku tvrzení B7 pro $\mathbb{B} = \boldsymbol{\mu}$ a $\mathbb{X} = \mathbf{1}_n$, tak dostaneme, že $\hat{\boldsymbol{\mu}}^\top = (\mathbf{1}_n^\top \mathbf{1}_n)^{-1} \mathbf{1}_n^\top \mathbb{Y} = \frac{1}{n} \mathbf{1}_n^\top \mathbb{Y}$ (tedy průměry jednotlivých sloupců matice \mathbb{Y}), a také dostaneme, že věrohodnost $p(\mathbb{Y}|\boldsymbol{\mu}, \boldsymbol{\Upsilon})$ je proporcionální výrazu

$$|\boldsymbol{\Upsilon}|^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \text{Tr} \boldsymbol{\Upsilon}^{-1} \left[\mathbb{Y}^\top \left(\mathbb{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right) \mathbb{Y} + n (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^\top \right] \right\}.$$

Odtud je vidět, že maximálně věrohodný odhad $\boldsymbol{\mu}$ je tvořen právě aritmetickými průměry $\hat{\boldsymbol{\mu}} =: \bar{\mathbf{Y}}_n$ nehledě na to, jaký je náš odhad $\boldsymbol{\Upsilon}$. Pro maximalizaci tohoto výrazu v proměnné $\boldsymbol{\Upsilon}$ nejprve logaritmuje:

$$\log p(\mathbb{Y}|\boldsymbol{\mu} = \hat{\boldsymbol{\mu}}, \boldsymbol{\Upsilon}) \propto -\frac{n}{2} \log |\boldsymbol{\Upsilon}| - \frac{1}{2} \text{Tr} \boldsymbol{\Upsilon}^{-1} \mathbb{Y}^\top \left(\mathbb{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right) \mathbb{Y}.$$

Posléze za pomoci tvrzení A5 derivujeme podle pozitivně definitní matice:

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\Upsilon}} \log p(\mathbb{Y}|\boldsymbol{\mu} = \hat{\boldsymbol{\mu}}, \boldsymbol{\Upsilon}) &= -\frac{n}{2} \left[2\boldsymbol{\Upsilon}^{-1} - \text{diag} \left(\boldsymbol{\Upsilon}^{-1} \right) \right] + \\ &+ \frac{1}{2} \left[2\boldsymbol{\Upsilon}^{-1} \mathbb{Y}^\top \left(\mathbb{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right) \mathbb{Y} \boldsymbol{\Upsilon}^{-1} - \text{diag} \left(\boldsymbol{\Upsilon}^{-1} \mathbb{Y}^\top \left(\mathbb{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right) \mathbb{Y} \boldsymbol{\Upsilon}^{-1} \right) \right]. \end{aligned}$$

Aby tento výraz byl roven nulové matici, je třeba zajistit, aby jak mimo diagonálu, tak na ní byla splněna rovnice

$$n\boldsymbol{\Upsilon}^{-1} = \boldsymbol{\Upsilon}^{-1} \mathbb{Y}^\top \left(\mathbb{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right) \mathbb{Y} \boldsymbol{\Upsilon}^{-1},$$

z čehož už je zřejmý klasický odhad varianční matice

$$\widehat{\boldsymbol{\Upsilon}} = \frac{1}{n} \mathbb{Y}^\top \left(\mathbb{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right) \mathbb{Y} = \frac{1}{n} \sum_{i=1}^n (\mathbf{Y}_i - \bar{\mathbf{Y}}_n) (\mathbf{Y}_i - \bar{\mathbf{Y}}_n)^\top,$$

Tabulka 3.1: Hodnoty Ledermannovy meze $\phi(m)$ pro počet faktorů k pro různá m .

m	3	6	15	21	36	45	66	78	105	120	153	210	276	300
$\phi(m)$	1	3	10	15	28	36	55	66	91	105	136	190	253	276
$m/3$	1	2	5	7	12	15	22	26	35	40	51	70	92	100

který se používá i bez nutnosti předpokladu normálního rozdělení.

Chceme-li tedy nalézt maximálně věrohodné odhady matic $\mathbf{\Lambda}$ a $\mathbf{\Sigma}$, tak musíme najít takový rozklad odhadu $\widehat{\mathbf{Y}}$, aby platilo

$$\widehat{\mathbf{Y}} = \widehat{\mathbf{\Lambda}}\widehat{\mathbf{\Lambda}}^\top + \widehat{\mathbf{\Sigma}}. \quad (3.8)$$

To znamená splnit $\frac{1}{2}m(m+1)$ rovnic ($\widehat{\mathbf{Y}}$ je symetrická matice, tedy se stačí soustředit na její horní trojúhelníkovou část). Počet neznámých parametrů je při předpokladu diagonální matice $\mathbf{\Sigma}$ roven $m(k+1)$ (mk za $\mathbf{\Lambda}$ a m za $\mathbf{\Sigma}$). Bekker a ten Berge (1997) ukázali, že pokud je počet faktorů k nižší než tzv. Ledermannova mez $\phi(m)$, tj.

$$k < \phi(m) = \frac{2m+1 - \sqrt{8m+1}}{2} \quad \Leftrightarrow \quad m+k < (m-k)^2, \quad (3.9)$$

pak pro skoro všechny dvojice $\mathbf{\Lambda}$ a $\mathbf{\Sigma}$ je $\mathbf{\Sigma}$ globálně identifikovatelná, tj. existuje jediné řešení takové, že $0 \leq \mathbf{\Sigma} \leq \mathbf{Y}$ a $\text{rank}(\mathbf{Y} - \mathbf{\Sigma}) \leq k$.

K tomuto omezení se lze také dopracovat za dodatečného předpokladu, že $\mathbf{\Lambda}^\top \mathbf{\Sigma} \mathbf{\Lambda}$ je diagonální matice, což dává dalších $\frac{1}{2}k(k-1)$ podmínek, které v jistém smyslu zajistí jednoznačnost rozkladu. Aby existovalo přesné řešení, je třeba, aby počet podmínek byl roven počtu parametrů, tj.

$$\begin{aligned} m(k+1) &= \frac{1}{2}m(m+1) + \frac{1}{2}k(k-1) \\ 2mk + 2m &= m^2 + m + k^2 - k \\ m+k &= (m-k)^2 \\ 0 &= k^2 - k(2m+1) + m^2 - m. \end{aligned}$$

Odtud už je vidět, že vzhledem k předpokladu $k < m$ musí být $k = \phi(m)$. Ovšem toto numerické řešení nemusí být v souladu se statistickou interpretací, neboť může nastat, že v nalezeném rozkladu budou záporné odhady diagonálních prvků matice $\mathbf{\Sigma}$. Bude-li existovat vícero řešení, zvolíme nějaké přípustné z nich. Neexistuje-li, je třeba se uchýlit k aproximativním metodám. V tabulce 3.1 můžeme vidět, že Ledermannova mez pro $m = 6$ je $\phi(6) = 3$ (tedy maximálně polovina z původního m), pro $m = 120$ již je $\phi(120) = 105$ (tedy maximálně 7/8 z původního m).

Obvykle se nejprve vstupní matice \mathbf{Y} centralizuje a standardizuje, tedy se pracuje s maticí $\mathbf{Y}^* = \left(\mathbb{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top\right)\mathbf{Y}\left(\text{diag}\widehat{\mathbf{Y}}\right)^{-\frac{1}{2}}$. Tím je zaručeno, že $\widehat{\mathbf{Y}}^* = \left(\text{diag}\widehat{\mathbf{Y}}\right)^{-\frac{1}{2}}\widehat{\mathbf{Y}}\left(\text{diag}\widehat{\mathbf{Y}}\right)^{-\frac{1}{2}}$ je odhadnutá korelační matice vektoru \mathbf{Y} a odhadnuté $\widehat{\lambda}_{j,l}$ pak mají význam odhadu korelace veličiny Y_j a faktoru F_l . Lze využít metody

hlavních komponent k rozložení $\widehat{\mathbf{Y}}$ na dvě pozitivně definitní matice, z nichž jedna nese podstatnou část informace obsažené v datech, ta druhá pouze představuje zanedbatelnou odchylku. Za onu podstatnou část se považuje za $\widehat{\mathbf{\Lambda}}\widehat{\mathbf{\Lambda}}^\top$ a za diagonální $\widehat{\mathbf{\Sigma}}$ se považuje dopočet do 1, co chybí na diagonále $\widehat{\mathbf{\Lambda}}\widehat{\mathbf{\Lambda}}^\top$. Existuje celá řada dalších metod, pomocí kterých lze nalézt onen rozklad (3.8) (alespoň aproximativně), ovšem nemáme zde prostor pro jejich rigorózní provedení, chceme se totiž především věnovat bayesovským metodám odhadu.

Jedním z hlavních problémů je dále určení počtu faktorů. Před spuštěním těchto klasických procedur se totiž musíme sami rozhodnout, jaký je počet faktorů k . Jsou případy, kdy můžeme předpokládat daný počet faktorů (jako například v příkladě 3). V takovém případě si ještě před sběrem dat můžeme zajistit, aby pro toto k počet měřených veličin m splňoval Ledermannovu mez (3.9). V příkladě 3 to znamená do dotazníku zvolit alespoň $5 = \lceil \phi^{-1}(2) \rceil$ otázek, ovšem čím více otázek, tím lépe dokážeme podchytit skryté latentní faktory. Pokud však nemáme předem jasnou představu o počtu faktorů a jde nám například jen o redukci dat, tak můžeme zvolit k dle některého z celé řady dostupných kritérií. Ty jsou vesměs založené na stejném principu jako doporučení počtu hlavních komponent. Několik takových metod je představeno v poslední kapitole, kde jsme některé tyto metody použili v simulačních studiích. Prakticky se vyplatí zkusit odhadnout model pro vícero počtů faktorů a pak si vybrat ten počet, který se dá podložit rozumnou interpretací. V kapitole 4 si představíme bayesovskou metodu, která tento problém řeší považováním počtu faktorů k za náhodný parametr modelu, který pak dále odhadne na základě MCMC metod jako ostatní parametry.

3.2 Bayesovská specifikace

Nyní opustíme klasické metody faktorové analýzy a pustíme se do bayesovské specifikace modelu. V předchozí části jsme se již potkali s věrohodností (3.5), která bude tvořit základ aposteriorní hustoty. Nyní však musíme přikročit k specifikaci apriorního rozdělení. To lze provést různými způsoby, my se zde zaměříme především na jednu specifikaci. V části 3.5 se seznámíme s dalšími možnými volbami a vysvětlíme si rozdíly.

3.2.1 Apriorní rozdělení

Zde vyjdeme z teorie konjugovaných rozdělení (část 2.2.2), ovšem nebude se jednat přímo o rozdělení v souladu s definicí 2. Využijeme jen poznatků o vhodném rozdělení pro parametr polohy a parametr rozptylu maticového normálního rozdělení. Apriorní rozdělení parametrů \mathbb{B} a $\mathbf{\Sigma}$ zavedeme jako v modelu bayesovské regrese 2.4, viz rovnice (2.9), tj.

$$\mathbb{B}|\mathbf{\Sigma} \sim \mathbf{N}_{m \times (1+k)}(\mathbb{B}_0, \mathbf{\Sigma} \otimes \mathbb{D}) \quad \text{a} \quad \mathbf{\Sigma} \sim \text{IW}_m(\mathbb{Q}, \nu),$$

kde \mathbb{B}_0 , \mathbb{D} , \mathbb{Q} a ν jsou hyperparametry.

Dále se již na \mathbb{B} nebudeme dívat jako na ucelený parametr, ale budeme pra-

covat zvlášt s $\boldsymbol{\mu}$ a $\boldsymbol{\Lambda}$. Necht hyperparametr rozptylu \mathbb{D} lze blokově rozdělit na

$$\mathbb{D} = \begin{pmatrix} d_{0,0} & \mathbf{0}_k^\top \\ \mathbf{0}_k & \mathbb{D}_{11} \end{pmatrix}.$$

Pro parametry $\boldsymbol{\mu}$ a $\boldsymbol{\Lambda}$ tedy předpokládáme apriorní nezávislost. Konkrétně tedy předpokládáme, že $\boldsymbol{\mu}|\boldsymbol{\Sigma} \sim \mathbf{N}_m(\boldsymbol{\mu}_0, d_{0,0}\boldsymbol{\Sigma})$ a $\boldsymbol{\Lambda}|\boldsymbol{\Sigma} \sim \mathbf{N}_{m \times k}(\boldsymbol{\Lambda}_0, \boldsymbol{\Sigma} \otimes \mathbb{D}_{11})$ jsou nezávislé, tomu odpovídají následující hustoty:

$$\begin{aligned} p(\boldsymbol{\mu}, \boldsymbol{\Lambda}|\boldsymbol{\Sigma}; \boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0, \mathbb{D}) &= p(\boldsymbol{\mu}|\boldsymbol{\Sigma}; \boldsymbol{\mu}_0, d_{0,0}) \cdot p(\boldsymbol{\Lambda}|\boldsymbol{\Sigma}; \boldsymbol{\Lambda}_0, \mathbb{D}_{11}), \\ p(\boldsymbol{\mu}|\boldsymbol{\Sigma}; \boldsymbol{\mu}_0, d_{0,0}) &\propto |d_{0,0}|^{-\frac{m}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\boldsymbol{\mu} - \boldsymbol{\mu}_0)^\top [d_{0,0}\boldsymbol{\Sigma}]^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0) \right\}, \\ p(\boldsymbol{\Lambda}|\boldsymbol{\Sigma}; \boldsymbol{\Lambda}_0, \mathbb{D}_{11}) &\propto |\mathbb{D}_{11}|^{-\frac{m}{2}} |\boldsymbol{\Sigma}|^{-\frac{k}{2}} \exp \left\{ -\frac{1}{2} \text{Tr} \boldsymbol{\Sigma}^{-1} (\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0) \mathbb{D}_{11}^{-1} (\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0)^\top \right\}. \end{aligned} \quad (3.10)$$

Pro parametr $\boldsymbol{\Sigma}$ budeme stejně jako v bayesovské mnohorozměrné regresi předpokládat, že $\boldsymbol{\Sigma} \sim \text{IW}_m(\mathbb{Q}, \nu)$, to znamená apriorní hustotu

$$p(\boldsymbol{\Sigma}|\mathbb{Q}, \nu) \propto |\mathbb{Q}|^{-\frac{\nu-m-1}{2}} |\boldsymbol{\Sigma}|^{-\frac{\nu}{2}} \exp \left\{ -\frac{1}{2} \text{Tr} \boldsymbol{\Sigma}^{-1} \mathbb{Q} \right\}. \quad (3.11)$$

Tedy dále již vše odvozujeme pro obecnou $\boldsymbol{\Sigma} > 0$. Pokud bychom se drželi klasického předpokladu diagonality matice $\boldsymbol{\Sigma}$, pak bychom uvažovali, že diagonální prvky $\sigma_{i,i} \sim \text{IW}_1(q_{i,i}, \nu)$ pro každé $i \in \{1, \dots, m\}$.

V souladu s předpoklady klasického modelu faktorové analýzy budeme předpokládat, že náhodný vektor faktorů \mathbf{F} pevně daného rozměru k pochází z normovaného normálního rozdělení, tj. $\mathbf{F} \sim \mathbf{N}_k(\mathbf{0}_k, \mathbb{I}_k)$. Apriorní hustotu všech n skrytých vektorů \mathbf{F}_i tak lze zapsat ve tvaru

$$p(\mathbb{F}) \propto \exp \left\{ -\frac{1}{2} \text{Tr} \mathbb{F} \mathbb{F}^\top \right\}. \quad (3.12)$$

Budeme předpokládat, že tyto faktory jsou rozděleny nezávisle na parametrech $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}$ a $\boldsymbol{\Sigma}$. Potom lze zapsat apriorní hustotu všech parametrů ve tvaru

$$\begin{aligned} p(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma}) &= p(\mathbb{F}) \cdot p(\boldsymbol{\mu}|\boldsymbol{\Sigma}) \cdot p(\boldsymbol{\Lambda}|\boldsymbol{\Sigma}) \cdot p(\boldsymbol{\Sigma}) \\ &\propto |\boldsymbol{\Sigma}|^{-\frac{k+1+\nu}{2}} \cdot \exp \left\{ -\frac{1}{2} \text{Tr} \mathbb{F} \mathbb{F}^\top \right\} \\ &\quad \cdot \exp \left\{ -\frac{1}{2} \text{Tr} \boldsymbol{\Sigma}^{-1} \left[(\boldsymbol{\mu} - \boldsymbol{\mu}_0) d_{0,0}^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0)^\top + (\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0) \mathbb{D}_{11}^{-1} (\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0)^\top + \mathbb{Q} \right] \right\}. \end{aligned} \quad (3.13)$$

Konkrétní hyperparametry $\boldsymbol{\mu}_0$, $\boldsymbol{\Lambda}_0$, $d_{0,0}$, $\mathbb{D}_{1,1}$, \mathbb{Q} a ν volíme na základě metod představených v části 2.2.3. Jelikož na ně dále nahlížíme jako na pevně volené konstanty, tak je pro přehlednost zápisu již nebudeme vypisovat v používaných označeních hustot. Volbu $d_{0,0}^{-1} = 0$ a $\mathbb{D}_{11}^{-1} = \mathbb{O}_{k \times k}$ lze považovat za neinformativní apriorní rozdělení (část 2.2.1) parametru $\boldsymbol{\mu}$ či $\boldsymbol{\Lambda}$. Ovšem pro další výpočty budeme předpokládat informativní rozdělení parametrů $\boldsymbol{\mu}$ a $\boldsymbol{\Lambda}$, tj. $d_{0,0}^{-1} > 0$ a $\mathbb{D}_{11}^{-1} > 0$.

3.2.2 Aposteriorní rozdělení

Pomocí Bayesovy věty 2 dáme dohromady věrohodnost (3.5) a apriorní hustotu (3.13) a získáme tak aposteriorní hustotu parametrů modelu. Platí tedy

$$\begin{aligned}
p(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma} | \mathbb{Y}) &\propto p(\mathbb{Y} | \mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma}) \cdot p(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma}) \\
&\propto |\boldsymbol{\Sigma}|^{-\frac{n+k+1+\nu}{2}} \exp \left\{ -\frac{1}{2} \left(\text{Tr } \mathbb{F} \mathbb{F}^\top + \right. \right. \\
&\quad \left. \left. + \text{Tr } \boldsymbol{\Sigma}^{-1} \left[\left(\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top - \mathbb{F} \boldsymbol{\Lambda}^\top \right)^\top \left(\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top - \mathbb{F} \boldsymbol{\Lambda}^\top \right) + \right. \right. \\
&\quad \left. \left. + (\boldsymbol{\mu} - \boldsymbol{\mu}_0) d_{0,0}^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0)^\top + (\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0) \mathbb{D}_{11}^{-1} (\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0)^\top + \mathbb{Q} \right] \right\}. \tag{3.14}
\end{aligned}$$

Nejprve se na tuto hustotu podíváme jako na funkci jednotlivých neznámých parametrů, abychom mohli určit jejich plně podmíněná rozdělení.

Tvrzení 6. *V modelu bayesovské faktorové analýzy nastíněném výše pro plně podmíněná rozdělení parametrů \mathbb{F} , $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}$ a $\boldsymbol{\Sigma}$ platí*

$$\begin{aligned}
\mathbb{F} | \mathbb{Y}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma} &\sim \mathbb{N}_{n \times k} \left(\widehat{\mathbb{F}}(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma}), \mathbb{I}_n \otimes \left[\boldsymbol{\Lambda}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda} + \mathbb{I}_k \right]^{-1} \right), \\
\boldsymbol{\mu} | \mathbb{Y}, \mathbb{F}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma} &\sim \mathbb{N}_m \left(\widehat{\boldsymbol{\mu}}(\mathbb{F}, \boldsymbol{\Lambda}), \boldsymbol{\Sigma} \left[n + d_{0,0}^{-1} \right]^{-1} \right), \\
\boldsymbol{\Lambda} | \mathbb{Y}, \mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Sigma} &\sim \mathbb{N}_{m \times k} \left(\widehat{\boldsymbol{\Lambda}}(\mathbb{F}, \boldsymbol{\mu}), \boldsymbol{\Sigma} \otimes \left[\mathbb{F}^\top \mathbb{F} + \mathbb{D}_{11}^{-1} \right]^{-1} \right), \\
\boldsymbol{\Sigma} | \mathbb{Y}, \mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda} &\sim \text{IW}_m \left(\boldsymbol{\Psi}(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda}), n + \nu + k + 1 \right),
\end{aligned}$$

kde

$$\begin{aligned}
\widehat{\mathbb{F}}(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma}) &= \left(\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top \right) \boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda} \left[\boldsymbol{\Lambda}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda} + \mathbb{I}_k \right]^{-1}, \\
\widehat{\boldsymbol{\mu}}(\mathbb{F}, \boldsymbol{\Lambda}) &= \frac{n}{n + d_{0,0}^{-1}} \left(\overline{\mathbb{Y}} - \boldsymbol{\Lambda} \overline{\mathbb{F}} \right) + \frac{d_{0,0}^{-1}}{n + d_{0,0}^{-1}} \boldsymbol{\mu}_0, \quad \overline{\mathbb{Y}} = \frac{1}{n} \mathbb{Y}^\top \mathbf{1}_n, \quad \overline{\mathbb{F}} = \frac{1}{n} \mathbb{F}^\top \mathbf{1}_n, \\
\widehat{\boldsymbol{\Lambda}}(\mathbb{F}, \boldsymbol{\mu}) &= \left[\left(\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top \right)^\top \mathbb{F} + \boldsymbol{\Lambda}_0 \mathbb{D}_{11}^{-1} \right] \left[\mathbb{F}^\top \mathbb{F} + \mathbb{D}_{11}^{-1} \right]^{-1} \\
&= \left[\widehat{\boldsymbol{\Lambda}}_\Lambda \mathbb{F}^\top \mathbb{F} + \boldsymbol{\Lambda}_0 \mathbb{D}_{11}^{-1} \right] \left[\mathbb{F}^\top \mathbb{F} + \mathbb{D}_{11}^{-1} \right]^{-1}, \\
\widehat{\boldsymbol{\Lambda}}_\Lambda &= \left(\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top \right)^\top \mathbb{F} \left(\mathbb{F}^\top \mathbb{F} \right)^{-1}, \quad \text{pokud } \mathbb{F}^\top \mathbb{F} > 0, \\
\boldsymbol{\Psi}(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) &= \left(\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top - \mathbb{F} \boldsymbol{\Lambda}^\top \right)^\top \left(\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top - \mathbb{F} \boldsymbol{\Lambda}^\top \right) + \\
&\quad + (\boldsymbol{\mu} - \boldsymbol{\mu}_0) d_{0,0}^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0)^\top + (\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0) \mathbb{D}_{11}^{-1} (\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0)^\top + \mathbb{Q} \\
&= (\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}(\mathbb{F}, \boldsymbol{\Lambda})) \left(n + d_{0,0}^{-1} \right) (\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}(\mathbb{F}, \boldsymbol{\Lambda}))^\top + \\
&\quad + (\boldsymbol{\Lambda} - \overline{\boldsymbol{\Lambda}}(\mathbb{F})) \left[\mathbb{F}^\top \mathbb{A}_d \mathbb{F} + \mathbb{D}_{11}^{-1} \right] (\boldsymbol{\Lambda} - \overline{\boldsymbol{\Lambda}}(\mathbb{F}))^\top - \\
&\quad - \overline{\boldsymbol{\Lambda}}(\mathbb{F}) \left[\mathbb{F}^\top \mathbb{A}_d \mathbb{F} + \mathbb{D}_{11}^{-1} \right] \overline{\boldsymbol{\Lambda}}(\mathbb{F})^\top - \\
&\quad - \left(n \overline{\mathbb{Y}} - \boldsymbol{\mu}_0 d_{0,0}^{-1} \right) \left(n + d_{0,0}^{-1} \right)^{-1} \left(n \overline{\mathbb{Y}} - \boldsymbol{\mu}_0 d_{0,0}^{-1} \right)^\top + \\
&\quad + \mathbb{Y}^\top \mathbb{Y} + \boldsymbol{\mu}_0 d_{0,0}^{-1} \boldsymbol{\mu}_0^\top + \boldsymbol{\Lambda}_0 \mathbb{D}_{11}^{-1} \boldsymbol{\Lambda}_0^\top + \mathbb{Q}, \\
\overline{\boldsymbol{\Lambda}}(\mathbb{F}) &= \left(\mathbb{Y}^\top \mathbb{A}_d \mathbb{F} + \boldsymbol{\Lambda}_0 \mathbb{D}_{11}^{-1} + \boldsymbol{\mu}_0 d_{0,0}^{-1} \left(n + d_{0,0}^{-1} \right)^{-1} n \overline{\mathbb{F}}^\top \right) \left[\mathbb{F}^\top \mathbb{A}_d \mathbb{F} + \mathbb{D}_{11}^{-1} \right]^{-1}, \\
\mathbb{A}_d &= \mathbb{I}_n - \frac{1}{n + d_{0,0}^{-1}} \mathbf{1}_n \mathbf{1}_n^\top.
\end{aligned}$$

Důkaz. Základem zde bude dívat se na aposteriorní hustotu (3.14), kterou lze zkrátit na tvar

$$p(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma} | \mathbb{Y}) \propto |\boldsymbol{\Sigma}|^{-\frac{n+\nu+k+1}{2}} \exp \left\{ -\frac{1}{2} \left(\text{Tr } \mathbb{F} \mathbb{F}^\top + \text{Tr } \boldsymbol{\Sigma}^{-1} \boldsymbol{\Psi}(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) \right) \right\},$$

jako na funkci právě zkoumaného parametru, viz rovnice (2.4).

Začneme nejprve s parametrem $\boldsymbol{\Sigma}$. Jelikož $\text{Tr } \mathbb{F} \mathbb{F}^\top$ nezávisí na $\boldsymbol{\Sigma}$, můžeme exponenciálu rozdělit a tuto část skrýt do proporcionality. Plně podmíněná hustota parametru $\boldsymbol{\Sigma}$ pak má tvar

$$p(\boldsymbol{\Sigma} | \mathbb{Y}, \mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) \propto |\boldsymbol{\Sigma}|^{-\frac{n+\nu+k+1}{2}} \exp \left\{ -\frac{1}{2} \text{Tr } \boldsymbol{\Sigma}^{-1} \boldsymbol{\Psi}(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) \right\}, \quad (3.15)$$

který nám připomíná tvar hustoty inverzního Wishartova rozdělení, viz (1.9), kde maticovým parametrem je $\boldsymbol{\Psi}(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda})$ a počet stupňů volnosti je roven $n+\nu+k+1$.

Pro obdržení plně podmíněných rozdělení parametrů $\boldsymbol{\mu}$ a $\boldsymbol{\Lambda}$ je třeba vhodně rozložit matici $\boldsymbol{\Psi}(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda})$. Rozklad uvedený ve znění tvrzení obdržíme použitím tvrzení B8 pro $\boldsymbol{\mu}$ následovně:

$$\mathbb{X} = \mathbf{1}_n, \mathbb{B} = \boldsymbol{\mu}, \mathbb{B}_0 = \boldsymbol{\mu}_0, \boldsymbol{\Phi} = d_{0,0}, \mathbb{Z} = \mathbb{F}, \mathbb{C} = \boldsymbol{\Lambda}, \mathbb{C}_0 = \boldsymbol{\Lambda}_0, \boldsymbol{\Psi} = \mathbb{D}_{11}.$$

Dostaneme tak rozklad $\boldsymbol{\Psi}(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda})$, kde $(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) (n + d_{0,0}^{-1}) (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^\top$ tvoří jeho stěžejní část. Pro zkrácení zápisu zde používáme $\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\mu}}(\mathbb{F}, \boldsymbol{\Lambda})$ definovaný výše. Všimněme si, že zbytek rozkladu $\boldsymbol{\Psi}(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda})$ již nezávisí na $\boldsymbol{\mu}$. Plně podmíněná hustota parametru $\boldsymbol{\mu}$ tak má tvar

$$\begin{aligned} p(\boldsymbol{\mu} | \mathbb{Y}, \mathbb{F}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma}) &\propto \exp \left\{ -\frac{1}{2} \text{Tr } \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) (n + d_{0,0}^{-1}) (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^\top \right\} \\ &\propto \exp \left\{ -\frac{1}{2} (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^\top \left[\frac{1}{n + d_{0,0}^{-1}} \boldsymbol{\Sigma} \right]^{-1} (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) \right\}, \end{aligned} \quad (3.16)$$

což připomíná tvar hustoty (1.2) vektorového normálního rozdělení s parametrem střední hodnoty $\hat{\boldsymbol{\mu}}$ a rozptylu $\boldsymbol{\Sigma} [n + d_{0,0}^{-1}]^{-1}$.

Prohodíme-li při použití tvrzení B8 významy \mathbb{B} a \mathbb{C} následovně:

$$\mathbb{X} = \mathbb{F}, \mathbb{B} = \boldsymbol{\Lambda}, \mathbb{B}_0 = \boldsymbol{\Lambda}_0, \boldsymbol{\Phi} = \mathbb{D}_{11}, \mathbb{Z} = \mathbf{1}_n, \mathbb{C} = \boldsymbol{\mu}, \mathbb{C}_0 = \boldsymbol{\mu}_0, \boldsymbol{\Psi} = d_{0,0},$$

tak dostaneme rozklad $\boldsymbol{\Psi}(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda})$, jehož stěžejní částí je

$$(\boldsymbol{\Lambda} - \hat{\boldsymbol{\Lambda}}) \left[\mathbb{F}^\top \mathbb{F} + \mathbb{D}_{11}^{-1} \right] (\boldsymbol{\Lambda} - \hat{\boldsymbol{\Lambda}})^\top,$$

kde $\hat{\boldsymbol{\Lambda}}$ je zkratkou pro $\hat{\boldsymbol{\Lambda}}(\mathbb{F}, \boldsymbol{\mu})$ definovaném ve znění tvrzení. Podobně jako tomu bylo u $\boldsymbol{\mu}$, zbytek rozkladu by nezávisel na $\boldsymbol{\Lambda}$, a proto ho lze při získávání plně podmíněného rozdělení zanedbat. Dostáváme tak plně podmíněnou hustotu parametru $\boldsymbol{\Lambda}$

$$p(\boldsymbol{\Lambda} | \mathbb{Y}, \mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto \exp \left\{ -\frac{1}{2} \text{Tr } \boldsymbol{\Sigma}^{-1} (\boldsymbol{\Lambda} - \hat{\boldsymbol{\Lambda}}) \left[\mathbb{F}^\top \mathbb{F} + \mathbb{D}_{11}^{-1} \right] (\boldsymbol{\Lambda} - \hat{\boldsymbol{\Lambda}})^\top \right\}, \quad (3.17)$$

kde rozpoznáváme hustotu (1.5) maticového normálního rozdělení s parametry $\hat{\boldsymbol{\Lambda}}$ a $\boldsymbol{\Sigma} \otimes \left[\mathbb{F}^\top \mathbb{F} + \mathbb{D}_{11}^{-1} \right]^{-1}$.

Určení plně podmíněného rozdělení parametru \mathbb{F} začněme tak, že z matice $\Psi(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda})$ zanedbejme sčítance nezávisající na \mathbb{F} (skryjí se v proporcionalitě). V exponenciále aposteriorní hustoty tak zbyde pouze stopa z matice

$$\left[\mathbb{F} \boldsymbol{\Lambda}^\top - (\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top) \right] \boldsymbol{\Sigma}^{-1} \left[\mathbb{F} \boldsymbol{\Lambda}^\top - (\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top) \right]^\top + \mathbb{F} \mathbb{F}^\top.$$

Použitím tvrzení B7, kde $\mathbb{X} = \mathbb{F}$, $\mathbb{C} = \boldsymbol{\Lambda}^\top$, $\mathbb{A} = \mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top$, $\boldsymbol{\Psi} = \boldsymbol{\Sigma}^{-1}$, $\mathbb{B} = \mathbb{O}_{n \times k}$, $\mathbb{D} = \mathbb{I}_k$, dostaneme, že

$$\begin{aligned} & \left[\mathbb{F} \boldsymbol{\Lambda}^\top - (\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top) \right] \boldsymbol{\Sigma}^{-1} \left[\mathbb{F} \boldsymbol{\Lambda}^\top - (\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top) \right]^\top + \mathbb{F} \mathbb{F}^\top = \\ & = (\mathbb{F} - \widehat{\mathbb{F}}) \left[\boldsymbol{\Lambda}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda} + \mathbb{I}_k \right] (\mathbb{F} - \widehat{\mathbb{F}})^\top - \widehat{\mathbb{F}} \left[\boldsymbol{\Lambda}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda} + \mathbb{I}_k \right] \widehat{\mathbb{F}}^\top + \mathbb{Y} \boldsymbol{\Sigma}^{-1} \mathbb{Y}^\top, \end{aligned} \quad (3.18)$$

kde $\widehat{\mathbb{F}}$ je zkratkou pro $\widehat{\mathbb{F}}(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma})$ definovaném ve znění tvrzení. Jelikož v rozkladu (3.18) se vyskytuje \mathbb{F} pouze v té první části, můžeme plně podmíněnou hustotu přepsat do tvaru

$$p(\mathbb{F} | \mathbb{Y}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma}) \propto \exp \left\{ -\frac{1}{2} \text{Tr} \left((\mathbb{F} - \widehat{\mathbb{F}}) \left[\boldsymbol{\Lambda}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda} + \mathbb{I}_k \right] (\mathbb{F} - \widehat{\mathbb{F}})^\top \right) \right\}, \quad (3.19)$$

který odpovídá tvaru hustoty (1.5) maticového normálního rozdělení s parametry $\widehat{\mathbb{F}}$ a $\mathbb{I}_n \otimes \left[\boldsymbol{\Lambda}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda} + \mathbb{I}_k \right]^{-1}$. □

Vidíme, že $\widehat{\boldsymbol{\mu}}$ a $\widehat{\boldsymbol{\Lambda}}$ jsou opět konvexní kombinací apriorní střední hodnoty $\boldsymbol{\mu}_0$ a $\boldsymbol{\Lambda}_0$ a odhadů z klasické statistiky, které předpokládají, že druhý z parametrů je pevný a známý. Ovšem nejedná se marginální aposteriorní střední hodnoty, nýbrž o střední hodnoty plně podmíněného rozdělení. Proto je nelze považovat za bayesovské odhady příslušných parametrů.

Toto tvrzení 6 později (část 3.3) použijeme ke konstrukci Gibbsova algoritmu, pomocí kterého můžeme dosáhnout generování z aposteriorního rozdělení. Odhady pak založíme na základě takto provedené simulace. Ukažme si ale nejprve, proč v tomto modelu nemůžeme použít odhadu parametrů pomocí marginální aposteriorní střední hodnoty.

3.2.3 Marginalizace aposteriorního rozdělení

V části 2.3 jsme si ukázali, že základním bayesovským odhadem jednotlivých částí $\boldsymbol{\theta}_j$ parametru $\boldsymbol{\theta}$ jsou marginální aposteriorní střední hodnoty $\mathbf{E}[\boldsymbol{\theta}_j | \mathbb{Y}]$. K tomu je zapotřebí umět jednoduše spočítat onen integrál $\int_{\Theta_j} \boldsymbol{\theta}_j p(\boldsymbol{\theta}_j | \mathbb{Y}) d\boldsymbol{\theta}_j$. Tomu se lze vyhnout například tím, že marginální aposteriorní rozdělení je jedno ze známých rozdělení. Ukažme si, že v modelu bayesovské faktorové analýzy žádná hezká známá marginální aposteriorní rozdělení nevystupují. Budeme se držet postupu představeného v části 2.3, který spočívá v postupném integrování jednotlivých parametrů z aposteriorní hustoty, dokud nezůstane marginální aposteriorní hustota jediného parametru. Nejprve budeme integrovat podle $\boldsymbol{\Sigma}$, potom odstraníme parametr střední hodnoty $\boldsymbol{\mu}$ a nakonec i $\boldsymbol{\Lambda}$. Dostaneme tak marginální aposteriorní hustotu \mathbb{F} , kterou však již nebudeme schopni klasifikovat do žádné rodiny běžných rozdělení.

Integrujme tedy nejprve aposteriorní hustotu (3.14) dle Σ . Dle tvrzení 6 je plně podmíněné rozdělení tohoto parametru inverzní Wishartovo s hustotou (3.15). Podle tvrzení 3 můžeme tuto část z aposteriorní hustoty vyintegrovat, je ale třeba zachovat či přidat multiplikatívni konstanty závislé na zbylých parametrech μ , Λ a \mathbb{F} . V hustotě inverzního Wishartova rozdělení (1.7) vystupuje oproti (3.15) ještě navíc determinant příslušného maticového parametru, kterým je v tomto případě $\Psi(\mathbb{F}, \mu, \Lambda)$, jenž kvůli závislosti na zbylých parametrech nesmí být vynechán. Proto dle tvrzení 3 je hustota rozdělení $\mathbb{F}, \mu, \Lambda | \mathbb{Y}$ rovna

$$p(\mathbb{F}, \mu, \Lambda | \mathbb{Y}) \propto \exp \left\{ -\frac{1}{2} \text{Tr } \mathbb{F} \mathbb{F}^\top \right\} \cdot |\Psi(\mathbb{F}, \mu, \Lambda)|^{-\frac{(n+\nu+k+1)-(m+1)}{2}}. \quad (3.20)$$

Nyní použijme rozklad matice $\Psi(\mathbb{F}, \mu, \Lambda)$ uvedený v tvrzení 6. Zkrajme tento zápis pomocí zbytkové matice $\mathbb{U}(\mathbb{F}, \Lambda)$ následovně:

$$\Psi(\mathbb{F}, \mu, \Lambda) = (\mu - \hat{\mu}) \left(n + d_{0,0}^{-1} \right) (\mu - \hat{\mu})^\top + \mathbb{U}(\mathbb{F}, \Lambda). \quad (3.21)$$

Potom lze pro výpočet determinantu použít Sylvestrovo pravidlo A1 ii) a dostaneme tak, že

$$|\Psi(\mathbb{F}, \mu, \Lambda)| = |\mathbb{U}(\mathbb{F}, \Lambda)| \left(1 + (n + d_{0,0}^{-1}) (\mu - \hat{\mu})^\top [\mathbb{U}(\mathbb{F}, \Lambda)]^{-1} (\mu - \hat{\mu}) \right).$$

Tento tvar již začíná velmi připomínat základ hustoty vektorového Studentova rozdělení (1.11). Uvědomme si, že μ je m -rozměrný vektor, tedy počet stupňů volnosti bude $\nu^* = n + \nu + k - 2m$, aby byl determinant umocněn na zlomek s čitatelem $\nu^* + m = n + \nu + k - m$. Tedy

$$|\Psi(\mathbb{F}, \mu, \Lambda)|^{-\frac{n+\nu+k-m}{2}} = |\mathbb{U}(\mathbb{F}, \Lambda)|^{-\frac{n+\nu+k-m}{2}} \cdot \left(1 + \frac{1}{\nu^*} (\mu - \hat{\mu})^\top \left[\frac{1}{\nu^*(n + d_{0,0}^{-1})} \mathbb{U}(\mathbb{F}, \Lambda) \right]^{-1} (\mu - \hat{\mu}) \right)^{-\frac{\nu^*+m}{2}}. \quad (3.22)$$

Vidíme, že druhá část výrazu (3.22) tvoří základ hustoty rozdělení

$$\mathfrak{t}_m \left(\nu^*, \hat{\mu}(\mathbb{F}, \Lambda), \frac{1}{\nu^*(n + d_{0,0}^{-1})} \mathbb{U}(\mathbb{F}, \Lambda) \right).$$

K úplnému doplnění hustoty (1.11) je zapotřebí dodat

$$\left| \frac{1}{\nu^*(n + d_{0,0}^{-1})} \mathbb{U}(\mathbb{F}, \Lambda) \right|^{-\frac{1}{2}},$$

na ν^* , n a $d_{0,0}^{-1}$ však nahlížíme jako na konstanty, proto stačí pouze determinant matice $\mathbb{U}(\mathbb{F}, \Lambda)$. Použijeme opět tvrzení 3 na hustotu (3.20), kam dosadíme (3.22) a kam přidáme $|\mathbb{U}(\mathbb{F}, \Lambda)|^{+\frac{1}{2}}$, dostaneme tedy hustotu rozdělení $\mathbb{F}, \Lambda | \mathbb{Y}$:

$$p(\mathbb{F}, \Lambda | \mathbb{Y}) \propto \exp \left\{ -\frac{1}{2} \text{Tr } \mathbb{F} \mathbb{F}^\top \right\} \cdot |\mathbb{U}(\mathbb{F}, \Lambda)|^{-\frac{n+\nu+k-m-1}{2}}. \quad (3.23)$$

Nyní zbývá integrovat už jen parametr Λ . Rozklad matice $\Psi(\mathbb{F}, \mu, \Lambda)$ z tvrzení 6 dále naznačuje, že zbytková matice $\mathbb{U}(\mathbb{F}, \Lambda)$ definovaná rovnicí (3.21) lze rozeepsat na

$$\mathbb{U}(\mathbb{F}, \Lambda) = \left(\Lambda - \bar{\Lambda}(\mathbb{F}) \right) \left[\mathbb{F}^\top \mathbb{A}_d \mathbb{F} + \mathbb{D}_{11}^{-1} \right] \left(\Lambda - \bar{\Lambda}(\mathbb{F}) \right)^\top + \Omega(\mathbb{F}), \quad (3.24)$$

kde $\mathbf{\Omega}(\mathbb{F})$ obsahuje zbytek rozkladu matice $\mathbf{\Psi}(\mathbb{F}, \boldsymbol{\mu}, \mathbf{\Lambda})$, který po několika úpravách nabývá následujícího tvaru:

$$\begin{aligned} \mathbf{\Omega}(\mathbb{F}) = & \mathbb{Q} + \mathbb{Y}^\top \left(\mathbb{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right) \mathbb{Y} + \frac{nd_{0,0}^{-1}}{n + d_{0,0}^{-1}} (\boldsymbol{\mu}_0 - \bar{\mathbf{Y}}) (\boldsymbol{\mu}_0 - \bar{\mathbf{Y}})^\top + \mathbf{\Lambda}_0 \mathbb{D}_{11}^{-1} \mathbf{\Lambda}_0^\top - \\ & - (\mathbb{V}\mathbb{F} + \mathbf{\Lambda}_0 \mathbb{D}_{11}^{-1}) \left[\mathbb{F}^\top \mathbb{A}_d \mathbb{F} + \mathbb{D}_{11}^{-1} \right]^{-1} (\mathbb{V}\mathbb{F} + \mathbf{\Lambda}_0 \mathbb{D}_{11}^{-1})^\top, \end{aligned} \quad (3.25)$$

kde $\mathbb{V} = \frac{d_{0,0}^{-1} \boldsymbol{\mu}_0 - n \bar{\mathbf{Y}}}{n + d_{0,0}^{-1}} \mathbf{1}_n^\top + \mathbb{Y}^\top = \mathbb{Y}^\top \mathbb{A}_d + \frac{d_{0,0}^{-1}}{n + d_{0,0}^{-1}} \boldsymbol{\mu}_0 \mathbf{1}_n^\top$.

Proto lze hustotu (3.23) rozdělení $\mathbb{F}, \mathbf{\Lambda} | \mathbb{Y}$ upravit do tvaru

$$\begin{aligned} p(\mathbb{F}, \mathbf{\Lambda} | \mathbb{Y}) \propto & \exp \left\{ -\frac{1}{2} \text{Tr} \mathbb{F} \mathbb{F}^\top \right\} \cdot \\ & \cdot \left| (\mathbf{\Lambda} - \bar{\mathbf{\Lambda}}(\mathbb{F})) \left[\mathbb{F}^\top \mathbb{A}_d \mathbb{F} + \mathbb{D}_{11}^{-1} \right] (\mathbf{\Lambda} - \bar{\mathbf{\Lambda}}(\mathbb{F}))^\top + \mathbf{\Omega}(\mathbb{F}) \right|^{-\frac{n+\nu+k-m-1}{2}}. \end{aligned} \quad (3.26)$$

Nyní už zbývá jen porovnat druhou část hustoty (3.26) s hustotou maticového Studentova rozdělení (1.12). Roli $\mathbf{\Phi}$ zde hraje $\mathbf{\Omega}(\mathbb{F})$ a $\boldsymbol{\Sigma}^{-1}$ v tomto případě $\bar{\nu} \left[\mathbb{F}^\top \mathbb{A}_d \mathbb{F} + \mathbb{D}_{11}^{-1} \right]$, kde počet stupňů volnosti je tentokrát $\bar{\nu} = n + \nu - m - 1$. K hustotě rozdělení

$$\mathbf{t}_{m \times k} \left(\bar{\nu}, \bar{\mathbf{\Lambda}}(\mathbb{F}), \mathbf{\Omega}(\mathbb{F}) \otimes \bar{\nu}^{-1} \left[\mathbb{F}^\top \mathbb{A}_d \mathbb{F} + \mathbb{D}_{11}^{-1} \right]^{-1} \right)$$

je dle tvrzení 3 nutno přidat také determinanty těchto matic umocněné po řadě na $\frac{-\bar{\nu}}{2}$ a $\frac{-m}{2}$. Tímto jsme ukázali, že marginální aposteriorní hustota \mathbb{F} , tedy hustota rozdělení $\mathbb{F} | \mathbb{Y}$, je rovna

$$p(\mathbb{F} | \mathbb{Y}) \propto \exp \left\{ -\frac{1}{2} \text{Tr} \mathbb{F} \mathbb{F}^\top \right\} \cdot |\mathbf{\Omega}(\mathbb{F})|^{-\frac{n+\nu-m-1}{2}} \cdot \left| \mathbb{F}^\top \mathbb{A}_d \mathbb{F} + \mathbb{D}_{11}^{-1} \right|^{-\frac{m}{2}}. \quad (3.27)$$

Pokusme se ještě tento výraz lépe upravit, abychom viděli, jak přesně na \mathbb{F} závisí. Zavedme si matici \mathbb{W} , která dle (3.25) splňuje

$$\mathbf{\Omega}(\mathbb{F}) = \mathbb{W} - (\mathbb{V}\mathbb{F} + \mathbf{\Lambda}_0 \mathbb{D}_{11}^{-1}) \left[\mathbb{F}^\top \mathbb{A}_d \mathbb{F} + \mathbb{D}_{11}^{-1} \right]^{-1} (\mathbb{V}\mathbb{F} + \mathbf{\Lambda}_0 \mathbb{D}_{11}^{-1})^\top.$$

Podle Sylvestрова pravidla A1 platí

$$\begin{aligned} |\mathbf{\Omega}(\mathbb{F})| &= |\mathbb{W}| \left| \mathbb{I}_n - \mathbb{W}^{-1} (\mathbb{V}\mathbb{F} + \mathbf{\Lambda}_0 \mathbb{D}_{11}^{-1}) \left[\mathbb{F}^\top \mathbb{A}_d \mathbb{F} + \mathbb{D}_{11}^{-1} \right]^{-1} (\mathbb{V}\mathbb{F} + \mathbf{\Lambda}_0 \mathbb{D}_{11}^{-1})^\top \right| \\ &= |\mathbb{W}| \left| \mathbb{F}^\top \mathbb{A}_d \mathbb{F} + \mathbb{D}_{11}^{-1} \right|^{-1} \cdot \\ &\quad \cdot \left| \mathbb{F}^\top \mathbb{A}_d \mathbb{F} + \mathbb{D}_{11}^{-1} - (\mathbb{V}\mathbb{F} + \mathbf{\Lambda}_0 \mathbb{D}_{11}^{-1})^\top \mathbb{W}^{-1} (\mathbb{V}\mathbb{F} + \mathbf{\Lambda}_0 \mathbb{D}_{11}^{-1}) \right|. \end{aligned}$$

Jestliže je matice $\mathbb{A}_d - \mathbb{V}^\top \mathbb{W}^{-1} \mathbb{V}$ invertovatelná, tak lze použít tvrzení B7, dle kterého lze upravit

$$\begin{aligned} \mathbb{F}^\top \mathbb{A}_d \mathbb{F} + \mathbb{D}_{11}^{-1} - (\mathbb{V}\mathbb{F} + \mathbf{\Lambda}_0 \mathbb{D}_{11}^{-1})^\top \mathbb{W}^{-1} (\mathbb{V}\mathbb{F} + \mathbf{\Lambda}_0 \mathbb{D}_{11}^{-1}) &= \\ &= (\mathbb{F} - \tilde{\mathbb{F}})^\top \left[\mathbb{A}_d - \mathbb{V}^\top \mathbb{W}^{-1} \mathbb{V} \right] (\mathbb{F} - \tilde{\mathbb{F}}) + \mathbb{Z}, \end{aligned}$$

$$\begin{aligned} \text{kde } \mathbf{Z} &= \mathbb{D}_{11}^{-1} - \mathbb{D}_{11}^{-1} \mathbf{\Lambda}_0^\top \mathbb{W}^{-1} \mathbf{\Lambda}_0 \mathbb{D}_{11}^{-1} - \tilde{\mathbf{F}}^\top [\mathbf{A}_d - \mathbb{V}^\top \mathbb{W}^{-1} \mathbb{V}] \tilde{\mathbf{F}} \\ \text{a } \tilde{\mathbf{F}} &= [\mathbf{A}_d - \mathbb{V}^\top \mathbb{W}^{-1} \mathbb{V}]^{-1} \mathbb{V}^\top \mathbb{W}^{-1} \mathbf{\Lambda}_0 \mathbb{D}_{11}^{-1}. \end{aligned}$$

Nyní už stačí jen tento výraz dosadit do vzorce pro determinant $\Omega(\mathbb{F})$, který dosadíme do hustoty (3.27) marginálního rozdělení matice faktorů \mathbb{F} . Výsledek je shrnut v následujícím tvrzení.

Tvrzení 7. *V modelu bayesovské faktorové analýzy s apriorním rozdělením definovaným v části 3.2.1 lze zapsat hustotu marginálního posteriorního rozdělení $\mathbb{F}|\mathbb{Y}$ ve tvaru*

$$\begin{aligned} p(\mathbb{F}|\mathbb{Y}) \propto \exp \left\{ -\frac{1}{2} \text{Tr} \mathbb{F} \mathbb{F}^\top \right\} \cdot \left| \mathbb{F}^\top \mathbf{A}_d \mathbb{F} + \mathbb{D}_{11}^{-1} \right|^{\frac{n+\nu-2m-1}{2}} \cdot \\ \cdot \left| (\mathbb{F} - \tilde{\mathbf{F}})^\top [\mathbf{A}_d - \mathbb{V}^\top \mathbb{W}^{-1} \mathbb{V}] (\mathbb{F} - \tilde{\mathbf{F}}) + \mathbf{Z} \right|^{-\frac{n+\nu-m-1}{2}}, \end{aligned}$$

pokud $\mathbf{A}_d - \mathbb{V}^\top \mathbb{W}^{-1} \mathbb{V}$ je invertibilní. Zde jsou tvary pomocných matic:

$$\begin{aligned} \mathbb{V} &= \frac{d_{0,0}^{-1} \boldsymbol{\mu}_0 - n \bar{\mathbf{Y}}}{n + d_{0,0}^{-1}} \mathbf{1}_n^\top + \mathbb{Y}^\top = \mathbb{Y}^\top \mathbf{A}_d + \frac{d_{0,0}^{-1}}{n + d_{0,0}^{-1}} \boldsymbol{\mu}_0 \mathbf{1}_n^\top, \\ \mathbb{W} &= \mathbb{Q} + \mathbb{Y}^\top \left(\mathbb{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right) \mathbb{Y} + \frac{nd_{0,0}^{-1}}{n + d_{0,0}^{-1}} (\boldsymbol{\mu}_0 - \bar{\mathbf{Y}}) (\boldsymbol{\mu}_0 - \bar{\mathbf{Y}})^\top + \mathbf{\Lambda}_0 \mathbb{D}_{11}^{-1} \mathbf{\Lambda}_0^\top, \\ \mathbf{Z} &= \mathbb{D}_{11}^{-1} - \mathbb{D}_{11}^{-1} \mathbf{\Lambda}_0^\top \mathbb{W}^{-1} \mathbf{\Lambda}_0 \mathbb{D}_{11}^{-1} - \tilde{\mathbf{F}}^\top [\mathbf{A}_d - \mathbb{V}^\top \mathbb{W}^{-1} \mathbb{V}] \tilde{\mathbf{F}}, \\ \tilde{\mathbf{F}} &= [\mathbf{A}_d - \mathbb{V}^\top \mathbb{W}^{-1} \mathbb{V}]^{-1} \mathbb{V}^\top \mathbb{W}^{-1} \mathbf{\Lambda}_0 \mathbb{D}_{11}^{-1}, \\ \mathbf{A}_d &= \mathbb{I}_n - \frac{1}{n + d_{0,0}^{-1}} \mathbf{1}_n \mathbf{1}_n^\top. \end{aligned}$$

Z tohoto tvaru již nejsme schopni vyčíst žádné běžně používané rozdělení. Pozitivní ale je, že máme základní tvar marginální posteriorní hustoty \mathbb{F} . Pro bayesovský odhad jednotlivých faktorů \mathbf{F}_i obsažených v \mathbb{F} tak už jen stačí nalézt metodu, která by uměla provést integraci

$$\mathbb{E}[\mathbb{F}|\mathbb{Y}] = \int_{\mathbb{R}^{n \times k}} \mathbb{F} p(\mathbb{F}|\mathbb{Y}) d\mathbb{F}.$$

Toto však již nemusí být snadný úkol a dále se mu věnovat nebudeme. Pro odhad marginálních posteriorních středních hodnot použijeme MCMC metod.

To by bylo marginální posteriorní rozdělení faktorů \mathbb{F} , dá se však k něčemu rozumnému dojít u ostatních parametrů? Tuto otázku se pokusíme zodpovědět tím, že se vrátíme k posteriorní hustotě (3.14), a pokusíme se nejprve vyintegrovat neznámé faktory \mathbb{F} , neboť v přehlednějším tvaru než zde již nevystupuje.

Dle tvrzení 6 víme, že plně podmíněné rozdělení \mathbb{F} je maticové normální, a proto by mělo jít tyto faktory vyintegrovat. Použijeme k tomu opět tvrzení 3, kde známou hustotou bude (3.19), ovšem musíme zde ponechat všechny činitele závislé na $\boldsymbol{\mu}$, $\mathbf{\Lambda}$ a $\boldsymbol{\Sigma}$, které jsme v důkazu tvrzení 6 při přechodu od (3.14) k (3.19)

zanedbali (přesněji řečeno skryli do proporcionality). Chceme tedy integrovat hustotu

$$\begin{aligned} p(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma} | \mathbb{Y}) &\propto |\boldsymbol{\Sigma}|^{-\frac{n+\nu+k+1}{2}} \exp \left\{ -\frac{1}{2} \text{Tr} \left(\mathbb{F} - \widehat{\mathbb{F}} \right) \left[\boldsymbol{\Lambda}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda} + \mathbb{I}_k \right] \left(\mathbb{F} - \widehat{\mathbb{F}} \right)^\top \right\} \\ &\cdot \left| \boldsymbol{\Lambda}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda} + \mathbb{I}_k \right|^{\frac{n}{2} - \frac{n}{2}} \exp \left\{ -\frac{1}{2} \text{Tr} \left[\mathbb{Y} \boldsymbol{\Sigma}^{-1} \mathbb{Y}^\top - \widehat{\mathbb{F}} \left[\boldsymbol{\Lambda}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda} + \mathbb{I}_k \right] \widehat{\mathbb{F}}^\top \right] \right\} \\ &\cdot \exp \left\{ -\frac{1}{2} \text{Tr} \boldsymbol{\Sigma}^{-1} \left[\mathbb{Q} + (\boldsymbol{\mu} - \boldsymbol{\mu}_0) d_{0,0}^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0)^\top + (\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0) \mathbb{D}_{11}^{-1} (\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0)^\top \right] \right\}, \end{aligned}$$

kde $\widehat{\mathbb{F}}$ je zkratkou pro $\widehat{\mathbb{F}}(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma})$, které je definované ve znění tvrzení 6. Použitím tvrzení 3 dostáváme hustotu rozdělení $\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma} | \mathbb{Y}$:

$$\begin{aligned} p(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma} | \mathbb{Y}) &\propto |\boldsymbol{\Sigma}|^{-\frac{n+\nu+k+1}{2}} \cdot \left| \boldsymbol{\Lambda}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda} + \mathbb{I}_k \right|^{-\frac{n}{2}} \\ &\cdot \exp \left\{ -\frac{1}{2} \text{Tr} \boldsymbol{\Sigma}^{-1} \left[\mathbb{J}(\boldsymbol{\Lambda}) + (\boldsymbol{\mu} - \boldsymbol{\mu}_0) d_{0,0}^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0)^\top \right] \right\} \\ &\cdot \exp \left\{ -\frac{1}{2} \text{Tr} \left[-(\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top) \underbrace{\boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda} \left[\boldsymbol{\Lambda}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda} + \mathbb{I}_k \right]^{-1} \boldsymbol{\Lambda}^\top \boldsymbol{\Sigma}^{-1}}_{=: \mathbb{K}(\boldsymbol{\Lambda}, \boldsymbol{\Sigma})} (\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top)^\top \right] \right\}, \end{aligned} \quad (3.28)$$

kde $\mathbb{J}(\boldsymbol{\Lambda}) = (\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0) \mathbb{D}_{11}^{-1} (\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0)^\top + \mathbb{Q} + \mathbb{Y}^\top \mathbb{Y}$. Vidíme, že na parametrech $\boldsymbol{\Lambda}$ a $\boldsymbol{\Sigma}$ tato funkce závisí velmi složitě. Jediný parametr $\boldsymbol{\mu}$, který není přímo v definici modelu spjatý s maticí \mathbb{F} , se zde vyskytuje v tvaru, který opět lze doplnit na čtverec a následně vyintegrovat. Začneme tím, že si nejprve pomocí tvrzení A2 upravíme výraz

$$\begin{aligned} \text{Tr} \left(\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top \right) \mathbb{K}(\boldsymbol{\Lambda}, \boldsymbol{\Sigma}) \left(\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top \right)^\top &= \\ &= \text{Tr} \left[\mathbb{Y} \mathbb{K}(\boldsymbol{\Lambda}, \boldsymbol{\Sigma}) \mathbb{Y}^\top - \mathbb{Y} \mathbb{K}(\boldsymbol{\Lambda}, \boldsymbol{\Sigma}) \boldsymbol{\mu} \mathbf{1}_n^\top - \mathbf{1}_n \boldsymbol{\mu}^\top \mathbb{K}(\boldsymbol{\Lambda}, \boldsymbol{\Sigma}) \mathbb{Y}^\top + \mathbf{1}_n \boldsymbol{\mu}^\top \mathbb{K}(\boldsymbol{\Lambda}, \boldsymbol{\Sigma}) \boldsymbol{\mu} \mathbf{1}_n^\top \right] \\ &= \text{Tr} \left[\mathbb{Y} \mathbb{K}(\boldsymbol{\Lambda}, \boldsymbol{\Sigma}) \mathbb{Y}^\top \right] - \mathbf{1}_n^\top \mathbb{Y} \mathbb{K}(\boldsymbol{\Lambda}, \boldsymbol{\Sigma}) \boldsymbol{\mu} - \boldsymbol{\mu}^\top \mathbb{K}(\boldsymbol{\Lambda}, \boldsymbol{\Sigma}) \mathbb{Y}^\top \mathbf{1}_n + n \boldsymbol{\mu}^\top \mathbb{K}(\boldsymbol{\Lambda}, \boldsymbol{\Sigma}) \boldsymbol{\mu} \\ &= \text{Tr} \left[\mathbb{Y} \mathbb{K}(\boldsymbol{\Lambda}, \boldsymbol{\Sigma}) \mathbb{Y}^\top \right] + n \left(\boldsymbol{\mu} - \bar{\boldsymbol{Y}} \right)^\top \mathbb{K}(\boldsymbol{\Lambda}, \boldsymbol{\Sigma}) \left(\boldsymbol{\mu} - \bar{\boldsymbol{Y}} \right) - n \bar{\boldsymbol{Y}}^\top \mathbb{K}(\boldsymbol{\Lambda}, \boldsymbol{\Sigma}) \bar{\boldsymbol{Y}}. \end{aligned} \quad (3.29)$$

Nyní vidíme, že ze vzorce (3.28) v exponenciále závisí na $\boldsymbol{\mu}$ pouze výraz

$$\begin{aligned} (\boldsymbol{\mu} - \boldsymbol{\mu}_0)^\top [d_{0,0} \boldsymbol{\Sigma}]^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0) + (\boldsymbol{\mu} - \bar{\boldsymbol{Y}})^\top [(-n) \mathbb{K}(\boldsymbol{\Lambda}, \boldsymbol{\Sigma})] (\boldsymbol{\mu} - \bar{\boldsymbol{Y}}) &= \\ &= (\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}})^\top \left[d_{0,0}^{-1} \boldsymbol{\Sigma}^{-1} - n \mathbb{K}(\boldsymbol{\Lambda}, \boldsymbol{\Sigma}) \right] (\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}) - \\ &- \tilde{\boldsymbol{\mu}}^\top \left[d_{0,0}^{-1} \boldsymbol{\Sigma}^{-1} - n \mathbb{K}(\boldsymbol{\Lambda}, \boldsymbol{\Sigma}) \right] \tilde{\boldsymbol{\mu}} + \boldsymbol{\mu}_0^\top [d_{0,0} \boldsymbol{\Sigma}]^{-1} \boldsymbol{\mu}_0 - n \bar{\boldsymbol{Y}}^\top \bar{\boldsymbol{Y}}^\top \mathbb{K}(\boldsymbol{\Lambda}, \boldsymbol{\Sigma}) \bar{\boldsymbol{Y}}, \end{aligned} \quad (3.30)$$

který jsme doplnili na čtverec pomocí důsledku tvrzení B6, kde

$$\tilde{\boldsymbol{\mu}} = \left[d_{0,0}^{-1} \boldsymbol{\Sigma}^{-1} - n \mathbb{K}(\boldsymbol{\Lambda}, \boldsymbol{\Sigma}) \right]^{-1} \left(d_{0,0}^{-1} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_0 - n \mathbb{K}(\boldsymbol{\Lambda}, \boldsymbol{\Sigma}) \bar{\boldsymbol{Y}} \right).$$

To lze ovšem použít pouze v případě, pokud je zaručeno, že matice

$$\mathbb{L}(\boldsymbol{\Lambda}, \boldsymbol{\Sigma}) = d_{0,0}^{-1} \boldsymbol{\Sigma}^{-1} - n \mathbb{K}(\boldsymbol{\Lambda}, \boldsymbol{\Sigma}) \quad (3.31)$$

je invertibilní. Podívejme se na determinant této matice

$$\begin{aligned}
|\mathbb{L}(\mathbf{\Lambda}, \mathbf{\Sigma})| &= |d_{0,0}\mathbf{\Sigma}|^{-1} \left| \mathbb{I}_m - nd_{0,0}\mathbf{\Lambda} \left[\mathbb{I}_k + \mathbf{\Lambda}^\top \mathbf{\Sigma}^{-1} \mathbf{\Lambda} \right]^{-1} \mathbf{\Lambda}^\top \mathbf{\Sigma}^{-1} \right| \\
&= |d_{0,0}\mathbf{\Sigma}|^{-1} \left| \mathbb{I}_k + \mathbf{\Lambda}^\top \mathbf{\Sigma}^{-1} \mathbf{\Lambda} \right|^{-1} \left| \mathbb{I}_k + \mathbf{\Lambda}^\top \mathbf{\Sigma}^{-1} \mathbf{\Lambda} - nd_{0,0}\mathbf{\Lambda}^\top \mathbf{\Sigma}^{-1} \mathbf{\Lambda} \right| \\
&= \left| \mathbf{\Lambda} \mathbf{\Lambda}^\top + \mathbf{\Sigma} \right|^{-1} \left| d_{0,0}^{-1} \mathbb{I}_k + (d_{0,0}^{-1} - n) \mathbf{\Lambda}^\top \mathbf{\Sigma}^{-1} \mathbf{\Lambda} \right| \\
&= \underbrace{\left| \mathbf{\Lambda} \mathbf{\Lambda}^\top + \mathbf{\Sigma} \right|^{-1}}_{>0} \underbrace{|\mathbf{\Sigma}|^{-1}}_{>0} \underbrace{\left| (d_{0,0}^{-1} - n) \mathbf{\Lambda} \mathbf{\Lambda}^\top + d_{0,0}^{-1} \mathbf{\Sigma} \right|}_{\star},
\end{aligned} \tag{3.32}$$

kde jsme opakovaně používali základní vlastnosti determinantu a Sylvestrovo pravidlo A1. Aby byla zajištěna invertibilita matice $\mathbb{L}(\mathbf{\Lambda}, \mathbf{\Sigma})$, je potřeba, aby její determinant byl nenulový. K tomu stačí předpokládat, že determinant \star je nenulový. Obecně však není jednoduché stanovit, kdy toto bude splněno.

Celkově lze užitím (3.29), (3.30) a Sylvestrova pravidla přepsat hustotu (3.28) do tvaru

$$\begin{aligned}
p(\boldsymbol{\mu}, \mathbf{\Lambda}, \mathbf{\Sigma} | \mathbb{Y}) &\propto |\mathbf{\Sigma}|^{-\frac{\nu+k+1}{2}} \cdot \left| \mathbf{\Lambda} \mathbf{\Lambda}^\top + \mathbf{\Sigma} \right|^{-\frac{n}{2}} \cdot \\
&\cdot \exp \left\{ -\frac{1}{2} \text{Tr} \mathbf{\Sigma}^{-1} \left[\mathbb{J}(\mathbf{\Lambda}) + \boldsymbol{\mu}_0 d_{0,0}^{-1} \boldsymbol{\mu}_0^\top \right] + \frac{1}{2} \text{Tr} \mathbb{Y} \mathbb{K}(\mathbf{\Lambda}, \mathbf{\Sigma}) \mathbb{Y}^\top + \tilde{\boldsymbol{\mu}}^\top \mathbb{L}(\mathbf{\Lambda}, \mathbf{\Sigma}) \tilde{\boldsymbol{\mu}} \right\} \cdot \\
&\cdot |\mathbb{L}(\mathbf{\Lambda}, \mathbf{\Sigma})|^{\frac{1}{2} - \frac{1}{2}} \cdot \exp \left\{ -\frac{1}{2} (\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}})^\top \mathbb{L}(\mathbf{\Lambda}, \mathbf{\Sigma}) (\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}) \right\},
\end{aligned} \tag{3.33}$$

ze kterého již můžeme jednoduše dle tvrzení 3 vyintegrovat parametr $\boldsymbol{\mu}$, neboť zde na posledním řádku spatřujeme hustotu (1.2) vektorového normálního rozdělení. Proto hustota rozdělení $\mathbf{\Lambda}, \mathbf{\Sigma} | \mathbb{Y}$ musí být tvaru

$$\begin{aligned}
p(\mathbf{\Lambda}, \mathbf{\Sigma} | \mathbb{Y}) &\propto |\mathbf{\Sigma}|^{-\frac{\nu+k+1}{2}} \cdot \left| \mathbf{\Lambda} \mathbf{\Lambda}^\top + \mathbf{\Sigma} \right|^{-\frac{n}{2}} |\mathbb{L}(\mathbf{\Lambda}, \mathbf{\Sigma})|^{-\frac{1}{2}} \cdot \\
&\cdot \exp \left\{ -\frac{1}{2} \text{Tr} \mathbf{\Sigma}^{-1} \left[\mathbb{J}(\mathbf{\Lambda}) + \boldsymbol{\mu}_0 d_{0,0}^{-1} \boldsymbol{\mu}_0^\top \right] + \frac{1}{2} \text{Tr} \mathbb{Y} \mathbb{K}(\mathbf{\Lambda}, \mathbf{\Sigma}) \mathbb{Y}^\top + \tilde{\boldsymbol{\mu}}^\top \mathbb{L}(\mathbf{\Lambda}, \mathbf{\Sigma}) \tilde{\boldsymbol{\mu}} \right\}.
\end{aligned} \tag{3.34}$$

Dosazením determinantu matice $\mathbb{L}(\mathbf{\Lambda}, \mathbf{\Sigma})$ (3.32) do této rovnice lze dojít ke konečnému tvaru, který je shrnut v následujícím tvrzení.

Tvrzení 8. *V modelu bayesovské faktorové analýzy s apriorním rozdělením definovaným v části 3.2.1 lze zapsat hustotu rozdělení $\mathbf{\Lambda}, \mathbf{\Sigma} | \mathbb{Y}$ ve tvaru*

$$\begin{aligned}
p(\mathbf{\Lambda}, \mathbf{\Sigma} | \mathbb{Y}) &\propto |\mathbf{\Sigma}|^{-\frac{\nu+k}{2}} \cdot \left| \mathbf{\Lambda} \mathbf{\Lambda}^\top + \mathbf{\Sigma} \right|^{-\frac{n-1}{2}} \cdot \left| (d_{0,0}^{-1} - n) \mathbf{\Lambda} \mathbf{\Lambda}^\top + d_{0,0}^{-1} \mathbf{\Sigma} \right|^{-\frac{1}{2}} \cdot \\
&\cdot \exp \left\{ -\frac{1}{2} \text{Tr} \mathbf{\Sigma}^{-1} \left[(\mathbf{\Lambda} - \mathbf{\Lambda}_0) \mathbb{D}_{11}^{-1} (\mathbf{\Lambda} - \mathbf{\Lambda}_0)^\top + \mathbb{Q} + \boldsymbol{\mu}_0 d_{0,0}^{-1} \boldsymbol{\mu}_0^\top \right] \right\} \cdot \\
&\cdot \exp \left\{ \frac{1}{2} \tilde{\boldsymbol{\mu}}^\top \mathbb{L}(\mathbf{\Lambda}, \mathbf{\Sigma}) \tilde{\boldsymbol{\mu}} - \frac{1}{2} \text{Tr} \mathbb{Y} \left[\mathbf{\Sigma}^{-1} - \mathbb{K}(\mathbf{\Lambda}, \mathbf{\Sigma}) \right] \mathbb{Y}^\top \right\},
\end{aligned}$$

pokud matice $\mathbb{L}(\mathbf{\Lambda}, \mathbf{\Sigma})$ je invertibilní. Zde jsou tvary pomocných matic:

$$\begin{aligned}
\mathbb{K}(\mathbf{\Lambda}, \mathbf{\Sigma}) &= \mathbf{\Sigma}^{-1} \mathbf{\Lambda} \left[\mathbf{\Lambda}^\top \mathbf{\Sigma}^{-1} \mathbf{\Lambda} + \mathbb{I}_k \right]^{-1} \mathbf{\Lambda}^\top \mathbf{\Sigma}^{-1}, \\
\mathbb{L}(\mathbf{\Lambda}, \mathbf{\Sigma}) &= d_{0,0}^{-1} \mathbf{\Sigma}^{-1} - n \mathbb{K}(\mathbf{\Lambda}, \mathbf{\Sigma}), \\
\tilde{\boldsymbol{\mu}} &= [\mathbb{L}(\mathbf{\Lambda}, \mathbf{\Sigma})]^{-1} \left(d_{0,0}^{-1} \mathbf{\Sigma}^{-1} \boldsymbol{\mu}_0 - n \mathbb{K}(\mathbf{\Lambda}, \mathbf{\Sigma}) \bar{\mathbb{Y}} \right).
\end{aligned}$$

V hustotě uvedené v tomto tvrzení vystupuje determinant z matice $\Upsilon = \Lambda\Lambda^\top + \Sigma$, o které jsme v části 3.1.1 (rovnice (3.1)) ukázali, že je rozptylová matice náhodného vektoru \mathbf{Y} nepodmíněného znalostí \mathbf{F} . Bohužel tato hustota tak komplikovaně závisí na Λ a Σ , že nejen, že nejsme schopni vyintegrovat jeden z těchto parametrů, ale dokonce tato hustota nepřipomíná žádné běžně používané rozdělení. Zřejmě nebude jednoduché navrhnout metody integrace dle této hustoty.

V této části jsme se tedy dozvěděli, že jednotlivá marginální aposteriorní rozdělení v modelu bayesovské regrese nejsou žádná známá a běžně používaná. Jsme proto nuceni použít jiných bayesovských metod odhadu jako například MCMC metody.

3.3 Gibbsův algoritmus

Jak jsme si ukázali v předchozí části, v modelu bayesovské faktorové analýzy nejsme schopni efektivně pracovat s marginálními aposteriorními rozděleními jednotlivých parametrů. Ovšem v tvrzení 6 jsme ukázali, že plně podmíněná rozdělení jednotlivých parametrů jsou běžně používaná rozdělení. Proto jsme schopni pro odhad parametrů použít Gibbsova algoritmu 1 představeného v části 2.3.2.

Uvedme si, jak tento algoritmus konkrétně použít v situaci bayesovské faktorové analýzy. Jediné, co si budeme muset stanovit, je pořadí, v jakém budeme tyto parametry generovat. Parametr, který se bude generovat jako první, má tu výhodu, že nepotřebuje počáteční hodnotu. Typicky se tedy rozhodujeme podle toho, co je pro nás obtížnější stanovit. My začneme s maticí faktorů \mathbb{F} , potom budeme pokračovat parametrem střední hodnoty $\boldsymbol{\mu}$, dále faktorovými zátěžemi Λ a poslední budeme generovat rozptylovou matici chyb modelu Σ .

Algoritmus 3. *Gibbsův algoritmus pro model bayesovské faktorové analýzy.*

Vstupy: požadovaná délka řetězce $B + M$, data \mathbf{y} , počáteční hodnoty parametrů $\boldsymbol{\mu}^0, \Lambda^0, \Sigma^0$.

Výstupy: markovský řetězec $(\mathbb{F}^i, \boldsymbol{\mu}^i, \Lambda^i, \Sigma^i), i \in \{1, \dots, B + M\}$.

Postup: v $(i + 1)$ -ním kroku algoritmu generuj z plně podmíněných rozdělení parametrů $\mathbb{F}, \boldsymbol{\mu}, \Lambda$ a Σ následovně:

- 1) $\mathbb{F}^{i+1} \sim \mathbb{F} \mid \mathbf{Y} = \mathbf{y}, \boldsymbol{\mu} = \boldsymbol{\mu}^i, \Lambda = \Lambda^i, \Sigma = \Sigma^i$
 $\sim \mathcal{N}_{n \times k} \left(\hat{\mathbb{F}}(\boldsymbol{\mu}^i, \Lambda^i, \Sigma^i), \mathbb{I}_n \otimes \left[(\Lambda^i)^\top [\Sigma^i]^{-1} \Lambda^i + \mathbb{I}_k \right]^{-1} \right),$
- 2) $\boldsymbol{\mu}^{i+1} \sim \boldsymbol{\mu} \mid \mathbf{Y} = \mathbf{y}, \mathbb{F} = \mathbb{F}^{i+1}, \Lambda = \Lambda^i, \Sigma = \Sigma^i$
 $\sim \mathcal{N}_m \left(\hat{\boldsymbol{\mu}}(\mathbb{F}^{i+1}, \Lambda^i), \Sigma^i [n + d_{0,0}^{-1}]^{-1} \right),$
- 3) $\Lambda^{i+1} \sim \Lambda \mid \mathbf{Y} = \mathbf{y}, \mathbb{F} = \mathbb{F}^{i+1}, \boldsymbol{\mu} = \boldsymbol{\mu}^{i+1}, \Sigma = \Sigma^i$
 $\sim \mathcal{N}_{m \times k} \left(\hat{\Lambda}(\mathbb{F}^{i+1}, \boldsymbol{\mu}^{i+1}), \Sigma^i \otimes \left[(\mathbb{F}^{i+1})^\top \mathbb{F}^{i+1} + \mathbb{D}_{11}^{-1} \right]^{-1} \right),$
- 4) $\Sigma^{i+1} \sim \Sigma \mid \mathbf{Y} = \mathbf{y}, \mathbb{F} = \mathbb{F}^{i+1}, \boldsymbol{\mu} = \boldsymbol{\mu}^{i+1}, \Lambda = \Lambda^{i+1}$
 $\sim \text{IW}_m \left(\Psi(\mathbb{F}^{i+1}, \boldsymbol{\mu}^{i+1}, \Lambda^{i+1}), n + \nu + k + 1 \right).$

Konec algoritmu: jakmile $i + 1 = B + M$.

Jestliže uvažujeme matici Σ pouze diagonální, tak musíme generovat jednotlivé prvky diagonály $\sigma_j^{2^{i+1}}$ z jednorozměrného Wishartova rozdělení se stejným počtem stupňů volnosti, ovšem s parametrem $\psi_{j,j}(\mathbb{F}^{i+1}, \boldsymbol{\mu}^{i+1}, \Lambda^{i+1})$, což odpovídá j -tému diagonálnímu prvku matice $\Psi(\mathbb{F}^{i+1}, \boldsymbol{\mu}^{i+1}, \Lambda^{i+1})$.

Pomocí vlastností uvedených v části 1.1 lze převést generování z mnohorozměrného normálního rozdělení pouze na generování z normovaného normálního rozdělení $\mathbf{N}(0, 1)$. Generování z inverzního Wishartova rozdělení s počtem stupňů volnosti, které je přirozené číslo, lze inverzí matice převést na generování z Wishartova rozdělení za pomoci součtu čtverců vektorů z normálního rozdělení, podrobněji viz vlastnosti v částech 1.2 a 1.3.

Dle úvah v části 2.3.1 pro odhad parametrů využijeme hodnot \mathbb{F}^i , $\boldsymbol{\mu}^i$, Λ^i a Σ^i , kde i je pouze v rozmezí $\{B + 1, \dots, B + M\}$. Pomocí ergodické věty tak můžeme průměry těchto hodnot odhadnout samotné parametry. Můžeme odhadovat i jiné charakteristiky marginálních aposteriorních rozdělení jednotlivých parametrů, ovšem musíme mít na paměti, že tyto generované veličiny netvoří nezávislé reprezentanty kvůli značné autokorelovanosti plynoucí ze samotného postupu generování.

3.4 ICM algoritmus

Alternativní metodou odhadu matice skrytých faktorů \mathbb{F} , parametru střední hodnoty $\boldsymbol{\mu}$, faktorových zátěží Λ a rozptylovou matici chyb modelu Σ je modus aposteriorního rozdělení. Vidíme, že aposteriorní hustota (3.14) je hladkou funkcí v těchto parametrech. Zkusme tedy nejprve nalézt maximum aposteriorní hustoty pomocí derivování a vyřešení soustavy, kde tyto první derivace položíme rovny nulovému elementu.

Držme se postupu nastíněném v části 2.3.3. Budeme maximalizovat logaritmovanou aposteriorní hustotu, kterou půjde snáze derivovat. Derivace této logaritmované aposteriorní hustoty dle jednotlivých parametrů převedeme na derivování logaritmovaných hustot plně podmíněných rozdělení. Všechny multiplikativní členy, které jsou obsaženy v proporcionalitě značené \propto , díky logaritmu přejdou v součet a následně jsou derivovány na nulu, neboť nezávisí na zkoumaném parametru. Podle pravidel pro derivování dle vícesložkových parametrů uvedených v tvrzení A5 dostáváme soustavu

$$\begin{aligned}
\frac{\partial}{\partial \mathbb{F}} \log(\mathbb{F} | \mathbb{Y}, \boldsymbol{\mu}, \Lambda, \Sigma) &= -(\mathbb{F} - \hat{\mathbb{F}}) \left[\Lambda^\top \Sigma^{-1} \Lambda + \mathbb{I}_k \right] \stackrel{!}{=} \mathbb{O}_{n \times k}, \\
\frac{\partial}{\partial \boldsymbol{\mu}} \log(\boldsymbol{\mu} | \mathbb{Y}, \mathbb{F}, \Lambda, \Sigma) &= -\Sigma^{-1} (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) \cdot (n + d_{0,0}^{-1}) \stackrel{!}{=} \mathbf{0}_m, \\
\frac{\partial}{\partial \Lambda} \log(\Lambda | \mathbb{Y}, \mathbb{F}, \boldsymbol{\mu}, \Sigma) &= -\Sigma^{-1} (\Lambda - \hat{\Lambda}) \left[\mathbb{F}^\top \mathbb{F} + \mathbb{D}_{11}^{-1} \right] \stackrel{!}{=} \mathbb{O}_{m \times k}, \\
\frac{\partial}{\partial \Sigma} \log(\Sigma | \mathbb{Y}, \mathbb{F}, \boldsymbol{\mu}, \Lambda) &= -\frac{n + \nu + k + 1}{2} \left[2\Sigma^{-1} - \text{diag} \Sigma^{-1} \right] + \\
&\quad + \frac{1}{2} \left[2\Sigma^{-1} \Psi \Sigma^{-1} - \text{diag} (\Sigma^{-1} \Psi \Sigma^{-1}) \right] \stackrel{!}{=} \mathbb{O}_{m \times m}.
\end{aligned} \tag{3.35}$$

Poslední rovnici pro parametr Σ řešíme analogicky, jako jsme řešili v závěru části 2.4 o modelu bayesovské mnohorozměrné regrese. Řešení soustavy (3.35)

vede na soustavu:

$$\begin{aligned}
\mathbb{F} &= \widehat{\mathbb{F}}(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma}), \\
\boldsymbol{\mu} &= \widehat{\boldsymbol{\mu}}(\mathbb{F}, \boldsymbol{\Lambda}), \\
\boldsymbol{\Lambda} &= \widehat{\boldsymbol{\Lambda}}(\mathbb{F}, \boldsymbol{\mu}), \\
\boldsymbol{\Sigma} &= \frac{1}{n + \nu + k + 1} \boldsymbol{\Psi}(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda}),
\end{aligned} \tag{3.36}$$

kde hodnoty na pravé straně jsou mody (srovnej s hodnotami v tabulce 1.1) jednotlivých plně podmíněných rozdělení z tvrzení 6, kde jsou také uvedeny definice $\widehat{\mathbb{F}}$, $\widehat{\boldsymbol{\mu}}$, $\widehat{\boldsymbol{\Lambda}}$ a $\boldsymbol{\Psi}$. Tato soustava již nemá explicitní tvar řešení, a proto je zapotřebí použít ICM algoritmu (představeného v části 2.3.3) pro iterativní nalezení výsledných odhadů.

I tento algoritmus lze provést v různém pořadí parametrů, přičemž pro první z nich nebudeme potřebovat jeho počáteční hodnotu. V našem případě budeme začínat parametrem \mathbb{F} , pro nějž potřebujeme počáteční hodnoty $\boldsymbol{\mu}^0$, $\boldsymbol{\Lambda}^0$ a $\boldsymbol{\Sigma}^0$ parametrů $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}$ a $\boldsymbol{\Sigma}$. Kdybychom začínali parametrem $\boldsymbol{\mu}$, tak by stačilo znát výchozí hodnoty pouze u \mathbb{F} a $\boldsymbol{\Lambda}$.

Algoritmus 4. *ICM algoritmus pro model bayesovské faktorové analýzy.*

Vstupy: data \mathbf{y} , počáteční hodnoty $\boldsymbol{\mu}^0$, $\boldsymbol{\Lambda}^0$ a $\boldsymbol{\Sigma}^0$.

Výstup: $\widetilde{\mathbb{F}}$, $\widetilde{\boldsymbol{\mu}}$, $\widetilde{\boldsymbol{\Lambda}}$ a $\widetilde{\boldsymbol{\Sigma}}$ maximalizující a posteriorní hustotu $p(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma} | \mathbf{Y} = \mathbf{y})$.

Postup: v $(i + 1)$ -ním kroku algoritmu uprav poslední známé hodnoty \mathbb{F}^i , $\boldsymbol{\mu}^i$, $\boldsymbol{\Lambda}^i$ a $\boldsymbol{\Sigma}^i$ na hodnoty nové \mathbb{F}^{i+1} , $\boldsymbol{\mu}^{i+1}$, $\boldsymbol{\Lambda}^{i+1}$ a $\boldsymbol{\Sigma}^{i+1}$ po složkách následovně:

$$\begin{aligned}
1) \mathbb{F}^{i+1} &= \left(\mathbf{y} - \mathbf{1}_n (\boldsymbol{\mu}^i)^\top \right) \left[\boldsymbol{\Sigma}^i \right]^{-1} \boldsymbol{\Lambda}^i \left[(\boldsymbol{\Lambda}^i)^\top \left[\boldsymbol{\Sigma}^i \right]^{-1} \boldsymbol{\Lambda}^i + \mathbb{I}_k \right]^{-1}, \\
2) \boldsymbol{\mu}^{i+1} &= \frac{n}{n + d_{0,0}^{-1}} \left(\bar{\mathbf{y}} - \boldsymbol{\Lambda}^i \bar{\mathbb{F}}^{i+1} \right) + \frac{d_{0,0}^{-1}}{n + d_{0,0}^{-1}} \boldsymbol{\mu}_0, \\
3) \boldsymbol{\Lambda}^{i+1} &= \left[\left(\mathbf{y} - \mathbf{1}_n (\boldsymbol{\mu}^{i+1})^\top \right)^\top \mathbb{F}^{i+1} + \boldsymbol{\Lambda}_0 \mathbb{D}_{11}^{-1} \right] \left[(\mathbb{F}^{i+1})^\top \mathbb{F}^{i+1} + \mathbb{D}_{11}^{-1} \right]^{-1}, \\
4) \boldsymbol{\Sigma}^{i+1} &= \frac{\left(\mathbf{y} - \mathbf{1}_n (\boldsymbol{\mu}^{i+1})^\top - \mathbb{F}^{i+1} (\boldsymbol{\Lambda}^{i+1})^\top \right)^\top \left(\mathbf{y} - \mathbf{1}_n (\boldsymbol{\mu}^{i+1})^\top - \mathbb{F}^{i+1} (\boldsymbol{\Lambda}^{i+1})^\top \right)}{n + \nu + k + 1} + \\
&\quad + \frac{(\boldsymbol{\mu}^{i+1} - \boldsymbol{\mu}_0) d_{0,0}^{-1} (\boldsymbol{\mu}^{i+1} - \boldsymbol{\mu}_0)^\top + (\boldsymbol{\Lambda}^{i+1} - \boldsymbol{\Lambda}_0) \mathbb{D}_{11}^{-1} (\boldsymbol{\Lambda}^{i+1} - \boldsymbol{\Lambda}_0)^\top + \mathbb{Q}}{n + \nu + k + 1}.
\end{aligned}$$

Konec algoritmu: jakmile vzdálenost mezi původním vektor $(\mathbb{F}^i, \boldsymbol{\mu}^i, \boldsymbol{\Lambda}^i, \boldsymbol{\Sigma}^i)$ a nově spočteným vektor $(\mathbb{F}^{i+1}, \boldsymbol{\mu}^{i+1}, \boldsymbol{\Lambda}^{i+1}, \boldsymbol{\Sigma}^{i+1})$ bude menší než předem stanovená hranice přesnosti. Za $\widetilde{\mathbb{F}}$, $\widetilde{\boldsymbol{\mu}}$, $\widetilde{\boldsymbol{\Lambda}}$ a $\widetilde{\boldsymbol{\Sigma}}$ zvol poslední \mathbb{F}^{i+1} , $\boldsymbol{\mu}^{i+1}$, $\boldsymbol{\Lambda}^{i+1}$, $\boldsymbol{\Sigma}^{i+1}$.

Jako vzdálenost mezi vektor $(\mathbb{F}^i, \boldsymbol{\mu}^i, \boldsymbol{\Lambda}^i, \boldsymbol{\Sigma}^i)$ a vektor $(\mathbb{F}^{i+1}, \boldsymbol{\mu}^{i+1}, \boldsymbol{\Lambda}^{i+1}, \boldsymbol{\Sigma}^{i+1})$ lze uvažovat vážený součet zvolených vzdáleností pro jednotlivé parametry, např.

$$\alpha_{\mathbb{F}} \left\| \mathbb{F}^{i+1} - \mathbb{F}^i \right\| + \alpha_{\boldsymbol{\mu}} \left\| \boldsymbol{\mu}^{i+1} - \boldsymbol{\mu}^i \right\| + \alpha_{\boldsymbol{\Lambda}} \left\| \boldsymbol{\Lambda}^{i+1} - \boldsymbol{\Lambda}^i \right\| + \alpha_{\boldsymbol{\Sigma}} \left\| \boldsymbol{\Sigma}^{i+1} - \boldsymbol{\Sigma}^i \right\|,$$

kde $\alpha_{\mathbb{F}} + \alpha_{\boldsymbol{\mu}} + \alpha_{\boldsymbol{\Lambda}} + \alpha_{\boldsymbol{\Sigma}} = 1$. Vyšší váha $\alpha_{\bullet} \geq 0$ znamená větší důraz na přesnost odhadu tohoto parametru.

Tento algoritmus nám tedy dává přibližné řešení $\widetilde{\mathbb{F}}$, $\widetilde{\boldsymbol{\mu}}$, $\widetilde{\boldsymbol{\Lambda}}$ a $\widetilde{\boldsymbol{\Sigma}}$ soustavy (3.36). Lze je tak považovat za bayesovské odhady parametrů \mathbb{F} , $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}$ a $\boldsymbol{\Sigma}$ založené na modu a posteriorního rozdělení.

3.5 Další volby apriorních rozdělání

V části 3.2.1 jsme si zavedli bayesovskou specifikaci modelu faktorové analýzy. Ovšem používaná apriorní rozdělání nejsou jediná, která lze v této situaci použít. Nyní si ukážeme, jaké máme jiné možnosti volby a jaká jsou potom výsledná aposteriorní rozdělání. Důkazy tvrzení v této části již nebudeme dělat tak podrobně, jako jsme činili doposud. Bude se totiž jednat o velmi podobné úpravy, často jen s drobnou změnou.

3.5.1 Neinformativní apriorní rozdělání

V části 2.2.1 jsme si ukázali, že nechceme-li o parametru modelu apriori přidávat žádnou informaci, tak pro tento parametr volíme neinformativní rozdělání. Ukážeme si zde, jak se zjednoduší plně podmíněná rozdělání v tvrzení 6, pokud nebudeme do modelu dodávat žádnou informaci o jednom z parametrů $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\Sigma}$.

Začněme nejprve s parametrem střední hodnoty $\boldsymbol{\mu}$. V části 3.2.1 jsme o něm apriori předpokládali, že se řídí normálním rozděláním $\mathbf{N}_m(\boldsymbol{\mu}_0, d_{0,0}\boldsymbol{\Sigma})$. Nyní o $\boldsymbol{\mu}$ nebudeme nic předpokládat, což vede na nevlastní apriorní hustotu $p(\boldsymbol{\mu}) \propto 1$. Lze se na to dívat také tak, že parametr apriorního rozptylu $d_{0,0}$ z (3.10) zvětšujeme nade všechny meze, až dosáhneme toho, že $d_{0,0}^{-1} = 0$. A jak si ukážeme v následujícím tvrzení, dostaneme výsledky obdobné tvrzení 6, ovšem právě pro $d_{0,0}^{-1} = 0$.

Tvrzení 9. *V modelu bayesovské faktorové analýzy, kde se apriorní specifikace od té zavedené v části 3.2.1 liší v apriorní neinformativnosti parametru $\boldsymbol{\mu}$, tj. $p(\boldsymbol{\mu}) \propto 1$, je aposteriorní hustota parametrů \mathbb{F} , $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}$ a $\boldsymbol{\Sigma}$ rovna*

$$p(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma} | \mathbb{Y}) \propto |\boldsymbol{\Sigma}|^{-\frac{n+k+\nu}{2}} \exp \left\{ -\frac{1}{2} \left(\text{Tr } \mathbb{F}\mathbb{F}^\top + \text{Tr } \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0)\mathbb{D}_{11}^{-1}(\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0)^\top + \right. \right. \\ \left. \left. + \text{Tr } \boldsymbol{\Sigma}^{-1} \left[(\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top - \mathbb{F}\boldsymbol{\Lambda}^\top)^\top (\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top - \mathbb{F}\boldsymbol{\Lambda}^\top) + \mathbb{Q} \right] \right) \right\}. \quad (3.37)$$

Plně podmíněná rozdělání parametrů \mathbb{F} a $\boldsymbol{\Lambda}$ jsou shodná s těmi uvedenými v tvrzení 6. Pro zbylé platí

$$\boldsymbol{\mu} | \mathbb{Y}, \mathbb{F}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma} \sim \mathbf{N}_m \left(\hat{\boldsymbol{\mu}}_\mu(\mathbb{F}, \boldsymbol{\Lambda}), \frac{1}{n} \boldsymbol{\Sigma} \right), \\ \boldsymbol{\Sigma} | \mathbb{Y}, \mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda} \sim \text{IW}_m(\boldsymbol{\Psi}_\mu(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda}), n + \nu + k),$$

$$\text{kde } \hat{\boldsymbol{\mu}}_\mu(\mathbb{F}, \boldsymbol{\Lambda}) = \bar{\mathbb{Y}} - \boldsymbol{\Lambda} \bar{\mathbb{F}} = \frac{1}{n} (\mathbb{Y} - \mathbb{F}\boldsymbol{\Lambda}^\top)^\top \mathbf{1}_n, \\ \boldsymbol{\Psi}_\mu(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = (\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top - \mathbb{F}\boldsymbol{\Lambda}^\top)^\top (\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top - \mathbb{F}\boldsymbol{\Lambda}^\top) + \\ + (\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0)\mathbb{D}_{11}^{-1}(\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0)^\top + \mathbb{Q}.$$

Důkaz. Jelikož jediné, co jsme změnili, je apriorní rozdělání $\boldsymbol{\mu}$, které závislo pouze na $\boldsymbol{\Sigma}$, tak se také změnila plně podmíněná rozdělání jen těchto dvou parametrů. Pro parametr $\boldsymbol{\Sigma}$ zde ubude jeden stupeň volnosti, který pramenil z hustoty (3.10). Z maticového parametru $\boldsymbol{\Psi}(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda})$ zmizí onen čtverec $(\boldsymbol{\mu} - \boldsymbol{\mu}_0)d_{0,0}^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)^\top$, tedy zbyde pouze $\boldsymbol{\Psi}_\mu(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda})$ definované ve znění tvrzení.

Pro plně podmíněné rozdělení $\boldsymbol{\mu}$ lze použít stejného postupu, jakého bylo použito v důkazu tvrzení 6, s tím rozdílem, že nyní $d_{0,0}^{-1} = 0$. Ovšem stále platí, že $\mathbf{1}_n^\top \mathbf{1}_n = n > 0$, a tak lze stejně použít tvrzení B7 (či B8). Výsledné $\hat{\boldsymbol{\mu}}_\mu(\mathbb{F}, \boldsymbol{\Lambda})$ tak jen vznikne zjednodušením $\hat{\boldsymbol{\mu}}(\mathbb{F}, \boldsymbol{\Lambda})$ za použití $d_{0,0}^{-1} = 0$. □

K analogickým závěrům lze podobnými úvahami dojít i u parametru faktorových zátěží $\boldsymbol{\Lambda}$, kde tentokrát pracujeme s $\mathbb{D}_{11}^{-1} = \mathbb{O}_{k \times k}$.

Tvrzení 10. *V modelu bayesovské faktorové analýzy, kde se apriorní specifikace od té zavedené v části 3.2.1 liší v apriorní neinformativnosti parametru $\boldsymbol{\Lambda}$, tj. $p(\boldsymbol{\Lambda}) \propto 1$, je aposteriorní hustota parametrů \mathbb{F} , $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}$ a $\boldsymbol{\Sigma}$ rovna*

$$p(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma} | \mathbb{Y}) \propto |\boldsymbol{\Sigma}|^{-\frac{n+1+\nu}{2}} \exp \left\{ -\frac{1}{2} \left(\text{Tr } \mathbb{F} \mathbb{F}^\top + \text{Tr } \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0) d_{0,0}^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0)^\top + \text{Tr } \boldsymbol{\Sigma}^{-1} \left[(\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top - \mathbb{F} \boldsymbol{\Lambda}^\top)^\top (\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top - \mathbb{F} \boldsymbol{\Lambda}^\top) + \mathbb{Q} \right] \right) \right\}. \quad (3.38)$$

Plně podmíněná rozdělení parametrů \mathbb{F} a $\boldsymbol{\mu}$ jsou shodná s těmi uvedenými v tvrzení 6. Jestliže má matice \mathbb{F} plnou sloupcovou hodnot, tedy $\mathbb{F}^\top \mathbb{F} > 0$, pak pro zbylé parametry platí

$$\boldsymbol{\Lambda} | \mathbb{Y}, \mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathbb{N}_{m \times k} \left(\hat{\boldsymbol{\Lambda}}_\Lambda(\mathbb{F}, \boldsymbol{\mu}), \boldsymbol{\Sigma} \otimes (\mathbb{F}^\top \mathbb{F})^{-1} \right),$$

$$\boldsymbol{\Sigma} | \mathbb{Y}, \mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda} \sim \text{IW}_m(\boldsymbol{\Psi}_\Lambda(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda}), n + \nu + 1),$$

$$\text{kde } \hat{\boldsymbol{\Lambda}}_\Lambda(\mathbb{F}, \boldsymbol{\mu}) = (\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top)^\top \mathbb{F} (\mathbb{F}^\top \mathbb{F})^{-1},$$

$$\boldsymbol{\Psi}_\Lambda(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = (\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top - \mathbb{F} \boldsymbol{\Lambda}^\top)^\top (\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top - \mathbb{F} \boldsymbol{\Lambda}^\top) + (\boldsymbol{\mu} - \boldsymbol{\mu}_0) d_{0,0}^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0)^\top + \mathbb{Q}.$$

Důkaz. Změnilo se pouze apriorní rozdělení $\boldsymbol{\Lambda}$, které již na $\boldsymbol{\mu}$ a \mathbb{F} nezáviselo, a tak jejich plně podmíněná rozdělení zůstávají stejná jako v tvrzení 6. Pro parametr $\boldsymbol{\Sigma}$ tentokrát ubude k stupňů volnosti, které byly obsaženy v původní apriorní hustotě $\boldsymbol{\Lambda}$ (3.10). Z maticového parametru $\boldsymbol{\Psi}(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda})$ zmizí $(\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0) \mathbb{D}_{11}^{-1} (\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0)^\top$, tedy zbyde pouze $\boldsymbol{\Psi}_\Lambda(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda})$ definované ve znění tvrzení.

Pro plně podmíněné rozdělení $\boldsymbol{\Lambda}$ lze použít stejného postupu, jakého bylo použito v důkazu tvrzení 6 s tím rozdílem, že nyní $\mathbb{D}_{11}^{-1} = \mathbb{O}_{k \times k}$. Ovšem stále za našich předpokladů platí, že $\mathbb{F}^\top \mathbb{F} > 0$, a tak lze stejně použít tvrzení B7 (či B8). Výsledné $\hat{\boldsymbol{\Lambda}}_\Lambda(\mathbb{F}, \boldsymbol{\mu})$ tak jen vznikne zjednodušením $\hat{\boldsymbol{\Lambda}}(\mathbb{F}, \boldsymbol{\mu})$ za použití $\mathbb{D}_{11}^{-1} = \mathbb{O}_{k \times k}$. □

V části 2.2.1 jsme si řekli, že za neinformativní rozdělení rozptylového parametru $\boldsymbol{\Sigma}$ považujeme $p(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{m+1}{2}}$. To v podstatě odpovídá hraniční nevlastní volbě $\nu = m + 1$ a $\mathbb{Q} = \mathbb{O}_{m \times m}$. Tato volba apriorního rozdělení vede ke zjednodušení aposteriorní hustoty a výsledného plně podmíněného rozdělení $\boldsymbol{\Sigma}$.

Tvrzení 11. *V modelu bayesovské faktorové analýzy, kde se apriorní specifikace od té zavedené v části 3.2.1 liší v apriorní neinformativnosti parametru $\boldsymbol{\Sigma}$, tj.*

$p(\Sigma) \propto |\Sigma|^{-\frac{m+1}{2}}$, je aposteriorní hustota parametrů \mathbb{F} , $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}$ a Σ rovna

$$p(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \Sigma | \mathbb{Y}) \propto |\Sigma|^{-\frac{n+k+m+2}{2}} \exp \left\{ -\frac{1}{2} \left(\text{Tr } \mathbb{F} \mathbb{F}^\top + \right. \right. \\ \left. \left. + \text{Tr } \Sigma^{-1} \left[\left(\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top - \mathbb{F} \boldsymbol{\Lambda}^\top \right)^\top \left(\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top - \mathbb{F} \boldsymbol{\Lambda}^\top \right) + \right. \right. \right. \\ \left. \left. \left. + (\boldsymbol{\mu} - \boldsymbol{\mu}_0) d_{0,0}^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0)^\top + (\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0) \mathbb{D}_{11}^{-1} (\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0)^\top \right] \right) \right\}. \quad (3.39)$$

Plně podmíněná rozdělení parametrů \mathbb{F} , $\boldsymbol{\mu}$ a $\boldsymbol{\Lambda}$ se oproti tvrzení 6 nijak nezmění. Pro parametr Σ však platí

$$\Sigma | \mathbb{Y}, \mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda} \sim \text{IW}_m(\Psi_\Sigma(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda}), n + k + m + 2),$$

$$\text{kde } \Psi_\Sigma(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = \left(\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top - \mathbb{F} \boldsymbol{\Lambda}^\top \right)^\top \left(\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top - \mathbb{F} \boldsymbol{\Lambda}^\top \right) + \\ + (\boldsymbol{\mu} - \boldsymbol{\mu}_0) d_{0,0}^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0)^\top + (\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0) \mathbb{D}_{11}^{-1} (\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0)^\top.$$

Důkaz. Dosazením $\nu = m + 1$ a $\mathbb{Q} = \mathbb{O}_{m \times m}$ do aposteriorní hustoty (3.14) dostaneme tvar (3.40). Nyní už analogicky jako v důkazu tvrzení 6 dostaneme plně podmíněné rozdělení parametru Σ uvedené ve znění tvrzení. \square

Takto bychom mohli zjednodušit apriorní rozdělení i pro více parametrů naráz, postup by však byl zcela analogický, a proto se tím již dále zabývat nebudeme. Tvrzení 9, 10 a 11 lze analogicky jako tvrzení 6 použít k sestavení Gibbsova algoritmu či ICM algoritmu pro získání odhadů jednotlivých parametrů.

3.5.2 Zobecněná apriorní rozdělení

První zobecnění apriorního rozdělení, které si zde ukážeme, bude odstranění předpokladu apriorní nezávislosti parametrů $\boldsymbol{\mu}$ a $\boldsymbol{\Lambda}$ za podmínky Σ . Původně jsme vyšli z toho, že $\mathbb{B} = (\boldsymbol{\mu}, \boldsymbol{\Lambda}) | \Sigma \sim \text{N}_{m \times (k+1)}((\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0), \Sigma \otimes \mathbb{D})$, kde \mathbb{D} byla blokově diagonální matice. Nyní budeme obecně předpokládat, že

$$\mathbb{D} = \begin{pmatrix} d_{0,0} & \mathbf{d}_{10}^\top \\ \mathbf{d}_{10} & \mathbb{D}_{11} \end{pmatrix},$$

což bude mít za následek, že $\boldsymbol{\mu}$ a $\boldsymbol{\Lambda}$ již nebudou apriorně při znalosti Σ nezávislé.

V tomto případě je daleko snazší a přehlednější pracovat se sdruženým parametrem \mathbb{B} , pro který se vše mnohem snáze odvozuje.

Tvrzení 12. V modelu bayesovské faktorové analýzy, kde se apriorní specifikace od té zavedené v části 3.2.1 liší v apriorní specifikaci sdruženého parametru parametru $\mathbb{B} = (\boldsymbol{\mu}, \boldsymbol{\Lambda}) | \Sigma \sim \text{N}_{m \times (k+1)}((\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0), \Sigma \otimes \mathbb{D})$, kde \mathbb{D} je obecná pozitivně definitní matice, je aposteriorní hustota parametrů \mathbb{F} , $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}$ a Σ rovna

$$p(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \Sigma | \mathbb{Y}) \propto |\Sigma|^{-\frac{n+\nu+k+1}{2}} \exp \left\{ -\frac{1}{2} \left(\text{Tr } \mathbb{F} \mathbb{F}^\top + \right. \right. \\ \left. \left. + \text{Tr } \Sigma^{-1} \left[\left(\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top - \mathbb{F} \boldsymbol{\Lambda}^\top \right)^\top \left(\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top - \mathbb{F} \boldsymbol{\Lambda}^\top \right) + \right. \right. \right. \\ \left. \left. \left. + (\mathbb{B} - \mathbb{B}_0) \mathbb{D}^{-1} (\mathbb{B} - \mathbb{B}_0)^\top + \mathbb{Q} \right] \right) \right\}. \quad (3.40)$$

Plně podmíněné rozdělení parametru \mathbb{F} se oproti tvrzení 6 nijak nezmění. Plně podmíněná rozdělení parametrů \mathbb{B} a Σ však jsou shodná s těmi v modelu mnohorozměrné bayesovské regrese s $\mathbb{X} = (\mathbf{1}_n, \mathbb{F})$, viz rovnice (2.12).

Důkaz. To, že nedojde ke změně plně podmíněného rozdělení faktorů \mathbb{F} , je opět zřejmé z toho, že nijak nefiguje ve změněné apriorní hustotě rozdělení $\mathbb{B}|\Sigma$. Určujeme-li plně podmíněné rozdělení parametrů \mathbb{B} a Σ , tak zde pracujeme s \mathbb{F} jako s maticí známých faktorů. To přesně odpovídá situaci mnohorozměrné bayesovské regrese se známou regresní maticí $\mathbb{X} = (\mathbf{1}_n, \mathbb{F})$. Proto lze použít výsledků obdržných v části 2.4. □

Nelze však využít marginálních aposteriorních rozdělení z části 2.4, neboť v našem případě jsou stále faktory \mathbb{F} neznámými parametry modelu, které je také třeba vyintegrovat. To je však stejně komplikované jako v části 3.2.3.

Pokud bychom však chtěli plně podmíněná rozdělení parametrů $\boldsymbol{\mu}$ a Λ , tak musíme použít vztahu (1.4) o podmínění jedné složky normálně rozděleného vektoru složkou zbývající pro $\mathbf{X}_1 = \boldsymbol{\mu}$ a $\mathbf{X}_2 = \text{vec } \Lambda$ za znalosti \mathbb{Y} , \mathbb{F} a Σ . Ono celkové sdružené rozdělení plyne z vektorizace (po sloupcích) rozdělení $\mathbb{B} = (\boldsymbol{\mu}, \Lambda)$ při \mathbb{Y} , \mathbb{F} a Σ tvaru (2.12), tedy

$$\text{vec } \mathbb{B} = \begin{pmatrix} \boldsymbol{\mu} \\ \text{vec } \Lambda \end{pmatrix} \Bigg|_{\mathbb{Y}, \mathbb{F}, \Sigma} \sim N_{m(k+1)} \left(\text{vec } \hat{\mathbb{B}}, [\mathbb{X}^\top \mathbb{X} + \mathbb{D}^{-1}]^{-1} \otimes \Sigma \right).$$

Označme si $\Phi = [\mathbb{X}^\top \mathbb{X} + \mathbb{D}^{-1}]^{-1}$ čtvercovou maticí řádu $k+1$, kterou lze blokově rozdělit na části o dimenzích 1 a k následovně

$$\Phi = \begin{pmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{pmatrix}.$$

Potom podle známého tvrzení o inverzi blokové matice (viz např. Anděl, 2007, Lemma 8.24) je matice $[\mathbb{X}^\top \mathbb{X} + \mathbb{D}^{-1}] = \Phi^{-1}$ zároveň rovna

$$\begin{aligned} & \begin{pmatrix} (\Phi_{11} - \Phi_{12}\Phi_{22}^{-1}\Phi_{21})^{-1} & -(\Phi_{11} - \Phi_{12}\Phi_{22}^{-1}\Phi_{21})^{-1}\Phi_{12}\Phi_{22}^{-1} \\ -(\Phi_{22} - \Phi_{21}\Phi_{11}^{-1}\Phi_{12})^{-1}\Phi_{21}\Phi_{11}^{-1} & (\Phi_{22} - \Phi_{21}\Phi_{11}^{-1}\Phi_{12})^{-1} \end{pmatrix} = \\ & = \begin{pmatrix} n + (d_{0,0} - \mathbf{d}_{10}^\top \mathbb{D}_{11}^{-1} \mathbf{d}_{10})^{-1} & n\bar{\mathbf{F}}^\top - (d_{0,0} - \mathbf{d}_{10}^\top \mathbb{D}_{11}^{-1} \mathbf{d}_{10})^{-1} \mathbf{d}_{10}^\top \mathbb{D}_{11}^{-1} \\ n\bar{\mathbf{F}} - (\mathbb{D}_{11} - \mathbf{d}_{10} \mathbf{d}_{0,0}^{-1} \mathbf{d}_{10}^\top)^{-1} \mathbf{d}_{10} \mathbf{d}_{0,0}^{-1} & \mathbf{F}^\top \mathbf{F} + (\mathbb{D}_{11} - \mathbf{d}_{10} \mathbf{d}_{0,0}^{-1} \mathbf{d}_{10}^\top)^{-1} \end{pmatrix}. \end{aligned}$$

Dle vlastnosti (1.4) a pravidel uvedených v tvrzeních A3 a A4, pak musí

$$\begin{aligned} \boldsymbol{\mu} | \mathbb{Y}, \mathbb{F}, \Lambda, \Sigma & \sim N_m \left(\check{\boldsymbol{\mu}} + (\Phi_{12}\Phi_{22}^{-1} \otimes \mathbb{I}_m) \text{vec} (\Lambda - \check{\Lambda}), (\Phi_{11} - \Phi_{12}\Phi_{22}^{-1}\Phi_{21}) \Sigma \right), \\ \text{vec } \Lambda | \mathbb{Y}, \mathbb{F}, \boldsymbol{\mu}, \Sigma & \sim N_{mk} \left(\text{vec } \check{\Lambda} + (\Phi_{21}\Phi_{11}^{-1} \otimes \mathbb{I}_m) (\boldsymbol{\mu} - \check{\boldsymbol{\mu}}), (\Phi_{22} - \Phi_{21}\Phi_{11}^{-1}\Phi_{12}) \otimes \Sigma \right), \end{aligned}$$

kde $\check{\boldsymbol{\mu}}$ a $\check{\boldsymbol{\Lambda}}$ jsou příslušné složky $\widehat{\mathbb{B}} = (\check{\boldsymbol{\mu}}, \check{\boldsymbol{\Lambda}})$ z tvrzení 4. Odtud už lze odvodit následující tvary plně podmíněných rozdělání

$$\boldsymbol{\mu}|\mathbb{Y}, \mathbb{F}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma} \sim \mathbf{N}_m \left(\check{\boldsymbol{\mu}} - (\boldsymbol{\Lambda} - \check{\boldsymbol{\Lambda}}) \mathbf{W}_1(\mathbb{F}), \left[n + (d_{0,0} - \mathbf{d}_{10}^\top \mathbb{D}_{11}^{-1} \mathbf{d}_{10})^{-1} \right]^{-1} \boldsymbol{\Sigma} \right),$$

$$\boldsymbol{\Lambda}|\mathbb{Y}, \mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathbf{N}_{m \times k} \left(\check{\boldsymbol{\Lambda}} - (\boldsymbol{\mu} - \check{\boldsymbol{\mu}}) \mathbf{W}_2(\mathbb{F}), \boldsymbol{\Sigma} \otimes \left[\mathbb{F}^\top \mathbb{F} + (\mathbb{D}_{11} - \mathbf{d}_{10} d_{0,0}^{-1} \mathbf{d}_{10}^\top)^{-1} \right]^{-1} \right),$$

kde

$$\mathbf{W}_1(\mathbb{F}) = \left(n \bar{\mathbf{F}} - \mathbb{D}_{11}^{-1} \mathbf{d}_{10} (d_{0,0} - \mathbf{d}_{10}^\top \mathbb{D}_{11}^{-1} \mathbf{d}_{10})^{-1} \right) \left[n + (d_{0,0} - \mathbf{d}_{10}^\top \mathbb{D}_{11}^{-1} \mathbf{d}_{10})^{-1} \right]^{-1},$$

$$\mathbf{W}_2(\mathbb{F}) = \left(n \bar{\mathbf{F}} - (\mathbb{D}_{11} - \mathbf{d}_{10} d_{0,0}^{-1} \mathbf{d}_{10}^\top)^{-1} \mathbf{d}_{10} d_{0,0}^{-1} \right)^\top \left[\mathbb{F}^\top \mathbb{F} + (\mathbb{D}_{11} - \mathbf{d}_{10} d_{0,0}^{-1} \mathbf{d}_{10}^\top)^{-1} \right]^{-1}.$$

Pro volbu $\mathbf{d}_{10} = \mathbf{0}_k$ odpovídající původnímu předpokladu apriorní nezávislosti $\boldsymbol{\mu}$ a $\boldsymbol{\Lambda}$ při $\boldsymbol{\Sigma}$ dostáváme alternativní vyjádření plně podmíněných rozdělání. Porovnej

$$\boldsymbol{\mu}|\mathbb{Y}, \mathbb{F}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma} \sim \mathbf{N}_m \left(\check{\boldsymbol{\mu}} - (\boldsymbol{\Lambda} - \check{\boldsymbol{\Lambda}}) \bar{\mathbf{F}} \frac{n}{n + d_{0,0}^{-1}}, \boldsymbol{\Sigma} \left[n + d_{0,0}^{-1} \right]^{-1} \right),$$

$$\boldsymbol{\Lambda}|\mathbb{Y}, \mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathbf{N}_{m \times k} \left(\check{\boldsymbol{\Lambda}} - (\boldsymbol{\mu} - \check{\boldsymbol{\mu}}) n \bar{\mathbf{F}}^\top \left[\mathbb{F}^\top \mathbb{F} + \mathbb{D}_{11}^{-1} \right]^{-1}, \boldsymbol{\Sigma} \otimes \left[\mathbb{F}^\top \mathbb{F} + \mathbb{D}_{11}^{-1} \right]^{-1} \right)$$

s plně podmíněnými rozděláními uvedenými v tvrzení 6. Vidíme tak, že $\check{\boldsymbol{\mu}}$ a $\widehat{\boldsymbol{\mu}}(\mathbb{F}, \boldsymbol{\Lambda})$ ani $\check{\boldsymbol{\Lambda}}$ a $\widehat{\boldsymbol{\Lambda}}(\mathbb{F}, \boldsymbol{\mu})$ nejsou ty samé výrazy.

Druhé zobecnění, které si ukážeme, bude spočívat v opuštění předpokladu, že $\boldsymbol{\mu}$ a $\boldsymbol{\Lambda}$ závisí apriorně na $\boldsymbol{\Sigma}$. Takovému přístupu se v literatuře říká použití zobecněného konjugovaného systému (Rowe, 2003). Budeme tak předpokládat, že parametry \mathbb{F} , $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}$ a $\boldsymbol{\Sigma}$ jsou apriorně nezávislé a že $\boldsymbol{\mu}$ a $\text{vec } \boldsymbol{\Lambda}$ jsou sice opět apriorně normálně rozdělány, ovšem se specifickými hyperparametry rozptylových matic. Tedy

$$\boldsymbol{\mu} \sim \mathbf{N}_m(\boldsymbol{\mu}_0, \boldsymbol{\Xi}) \quad \text{a} \quad \text{vec } \boldsymbol{\Lambda} \sim \mathbf{N}_{mk}(\text{vec } \boldsymbol{\Lambda}_0, \boldsymbol{\Delta}). \quad (3.41)$$

Aposteriorní hustota pak má tvar

$$p(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma}|\mathbb{Y}) \propto p(\mathbb{Y}|\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma}) \cdot p(\mathbb{F}) \cdot p(\boldsymbol{\mu}) \cdot p(\boldsymbol{\Lambda}) \cdot p(\boldsymbol{\Sigma})$$

$$\propto |\boldsymbol{\Sigma}|^{-\frac{n+\nu}{2}} \cdot \exp \left\{ -\frac{1}{2} \text{Tr } \boldsymbol{\Sigma}^{-1} (\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top - \mathbb{F} \boldsymbol{\Lambda}^\top)^\top (\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top - \mathbb{F} \boldsymbol{\Lambda}^\top) - \right.$$

$$\left. -\frac{1}{2} \left[\text{Tr } \mathbb{F}^\top \mathbb{F} + (\boldsymbol{\mu} - \boldsymbol{\mu}_0)^\top \boldsymbol{\Xi}^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0) + \text{vec } (\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0)^\top \boldsymbol{\Delta}^{-1} \text{vec } (\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0) \right] \right\}. \quad (3.42)$$

Tvrzení 13. *V modelu bayesovské faktorové analýzy s apriorním rozděláním, které se od toho zavedeného v části 3.2.1 liší v apriorní specifikaci parametrů $\boldsymbol{\mu}$ a $\boldsymbol{\Lambda}$, která je dána rovnicí (3.41), je posteriorní hustota dána rovnicí (3.42). Plně podmíněné rozdělání faktorů \mathbb{F} je shodné s tím uvedeným v tvrzení 6. Plně podmíněná rozdělání parametrů $\boldsymbol{\mu}$, $\text{vec } \boldsymbol{\Lambda}$ a $\boldsymbol{\Sigma}$ jsou potom*

$$\boldsymbol{\mu}|\mathbb{Y}, \mathbb{F}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma} \sim \mathbf{N}_m \left(\vec{\boldsymbol{\mu}}, \left[n \boldsymbol{\Sigma}^{-1} + \boldsymbol{\Xi}^{-1} \right]^{-1} \right),$$

$$\text{vec } \boldsymbol{\Lambda}|\mathbb{Y}, \mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathbf{N}_{mk} \left(\text{vec } \vec{\boldsymbol{\Lambda}}, \left[\mathbb{F}^\top \mathbb{F} \otimes \boldsymbol{\Sigma}^{-1} + \boldsymbol{\Delta}^{-1} \right]^{-1} \right),$$

$$\boldsymbol{\Sigma}|\mathbb{Y}, \mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda} \sim \text{IW}_m \left(\mathbb{Q} + (\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top - \mathbb{F} \boldsymbol{\Lambda}^\top)^\top (\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top - \mathbb{F} \boldsymbol{\Lambda}^\top), n + \nu \right),$$

kde

$$\begin{aligned}
\vec{\boldsymbol{\mu}} &= [n\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Xi}^{-1}]^{-1} \left((\mathbf{1}_n^\top \otimes \boldsymbol{\Sigma}^{-1}) \text{vec} (\mathbb{Y}^\top - \boldsymbol{\Lambda} \mathbb{F}^\top) + \boldsymbol{\Xi}^{-1} \boldsymbol{\mu}_0 \right) \\
&= [n\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Xi}^{-1}]^{-1} \left(n\boldsymbol{\Sigma}^{-1} (\bar{\mathbb{Y}} - \boldsymbol{\Lambda} \bar{\mathbb{F}}) + \boldsymbol{\Xi}^{-1} \boldsymbol{\mu}_0 \right), \\
\text{vec } \vec{\boldsymbol{\Lambda}} &= [\mathbb{F}^\top \mathbb{F} \otimes \boldsymbol{\Sigma}^{-1} + \boldsymbol{\Delta}^{-1}]^{-1} \left((\mathbb{F}^\top \otimes \boldsymbol{\Sigma}^{-1}) \text{vec} (\mathbb{Y}^\top - \boldsymbol{\mu} \mathbf{1}_n^\top) + \boldsymbol{\Delta}^{-1} \text{vec } \boldsymbol{\Lambda}_0 \right) \\
&= [\mathbb{F}^\top \mathbb{F} \otimes \boldsymbol{\Sigma}^{-1} + \boldsymbol{\Delta}^{-1}]^{-1} \left((\mathbb{F}^\top \mathbb{F} \otimes \boldsymbol{\Sigma}^{-1}) \text{vec} (\hat{\boldsymbol{\Lambda}}_\Lambda) + \boldsymbol{\Delta}^{-1} \text{vec } \boldsymbol{\Lambda}_0 \right), \\
\hat{\boldsymbol{\Lambda}}_\Lambda &= (\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top)^\top (\mathbb{F}^\top \mathbb{F})^{-1} \mathbb{F}^\top, \quad \text{pro } \mathbb{F}^\top \mathbb{F} > 0.
\end{aligned}$$

Důkaz. Plně podmíněné rozdělení $\boldsymbol{\Sigma}$ lze odvodit analogicky jako v důkazu tvrzení 6 jen s tím rozdílem, že nyní je nižší počet stupňů volnosti a v exponenciále se již $\boldsymbol{\Sigma}^{-1}$ vyskytuje jen u \mathbb{Q} a té části z věrohodnosti.

Pro plně podmíněná rozdělení $\boldsymbol{\mu}$ a $\boldsymbol{\Sigma}$ nejprve použijeme vlastnosti v) z tvrzení A4, abychom upravili stopu z věrohodnosti do tvarů, jejichž podstatnou částí budou

$$\begin{aligned}
& \left(\boldsymbol{\mu}^\top (\mathbf{1}_n^\top \otimes \mathbb{I}_m) - \text{vec} (\mathbb{Y}^\top - \boldsymbol{\Lambda} \mathbb{F}^\top) \right) [\mathbb{I}_n \otimes \boldsymbol{\Sigma}^{-1}] \\
& \qquad \qquad \qquad \left(\boldsymbol{\mu}^\top (\mathbf{1}_n^\top \otimes \mathbb{I}_m) - \text{vec} (\mathbb{Y}^\top - \boldsymbol{\Lambda} \mathbb{F}^\top) \right)^\top, \\
& \left((\text{vec } \boldsymbol{\Lambda})^\top (\mathbb{F}^\top \otimes \mathbb{I}_m) - \text{vec} (\mathbb{Y}^\top - \boldsymbol{\mu} \mathbf{1}_n^\top) \right) [\mathbb{I}_n \otimes \boldsymbol{\Sigma}^{-1}] \\
& \qquad \qquad \qquad \left((\text{vec } \boldsymbol{\Lambda})^\top (\mathbb{F}^\top \otimes \mathbb{I}_m) - \text{vec} (\mathbb{Y}^\top - \boldsymbol{\mu} \mathbf{1}_n^\top) \right)^\top.
\end{aligned}$$

Ostatní členy z tohoto rozkladu nás nezajímají, neboť již nezávisí na $\boldsymbol{\mu}$ či v druhém případě na $\boldsymbol{\Lambda}$. Nyní již k výrazům výše přičteme kvadratické formy plynoucí z apriorního rozdělení a dle tvrzení B7 doplníme na čtverec, kde hlavní proměnnou \mathbb{X} z tohoto tvrzení rozumíme nejprve $\boldsymbol{\mu}^\top$ a posléze $(\text{vec } \boldsymbol{\Lambda})^\top$. Odtud po několika úpravách s pomocí tvrzení A3 a A4 dostaneme příslušné $\vec{\boldsymbol{\mu}}$ a $\vec{\boldsymbol{\Lambda}}$ ve znění právě dokazovaného tvrzení. Doplněné kvadratické formy pak mají po několika úpravách tvar

$$\begin{aligned}
& \left(\boldsymbol{\mu} - \vec{\boldsymbol{\mu}} \right)^\top [n\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Xi}^{-1}] \left(\boldsymbol{\mu} - \vec{\boldsymbol{\mu}} \right), \\
& \left(\text{vec} (\boldsymbol{\Lambda} - \vec{\boldsymbol{\Lambda}}) \right)^\top [\mathbb{F}^\top \mathbb{F} \otimes \boldsymbol{\Sigma}^{-1} + \boldsymbol{\Delta}^{-1}] \text{vec} (\boldsymbol{\Lambda} - \vec{\boldsymbol{\Lambda}}).
\end{aligned}$$

Jelikož zbytek výrazů z doplnění na čtverec již neobsahuje $\boldsymbol{\mu}$ a v druhém případě $\boldsymbol{\Lambda}$, tak se jedná o stěžejní části hustot plně podmíněných rozdělení, která musí být vektorová normální, neboť tyto kvadratické formy vystupují v exponenciále s $-1/2$. Parametry těchto rozdělení tak vyčteme z těchto kvadratických forem. \square

Na základě tohoto tvrzení lze opět sestavit Gibbsův či ICM algoritmus, pomocí nichž lze odhadnout jednotlivé parametry za těchto obecnějších podmínek. Nezávislostí apriorního rozdělení $\boldsymbol{\mu}$ a $\boldsymbol{\Lambda}$ na $\boldsymbol{\Sigma}$ jsme si však poněkud ztížili úlohu nastavení výchozích hyperparametrů $\boldsymbol{\Xi}$ a $\boldsymbol{\Delta}$. Pro jednoduchost se někdy specifikují jako diagonální matice. Abychom zaručili, že se nebude jednat o příliš informativní rozdělení, tak často považujeme tyto parametry navíc za náhodné z inverzního Wishartova rozdělení, které už však zvolíme napevno.

3.6 Zobecnění bayesovské faktorové analýzy

Zobecnění uvedená v předchozí části pramenila ze změny uvažovaného apriorního rozdělení parametrů zavedených v části 3.2. Nyní se však podíváme, jakými způsoby lze rozšířit model bayesovské faktorové analýzy. Na následujícím příkladu si budeme ilustrovat, proč je v některých situacích nutné pracovat s obecnějším modelem.

Příklad 4 (Problém koktejlového večírku (Cocktail party problem)). Na koktejlovém večírku se nachází celkem k lidí, kteří spolu v různých skupinkách vedou konverzace. Rádi bychom zaznamenali, čím kdo do těchto konverzací přispěl, tedy hodnotu signálu l -tého člověka F_l . Proto za tímto účelem po místnosti rozmístíme m mikrofonů, které zaznamenávají hodnotu signálu $Y_j, j \in \{1, \dots, m\}$. Ovšem signál naměřený na mikrofonu je nějakou směsí (funkcí) signálů \mathbf{F} jednotlivých účastníků večírku a navíc nějakého šumu. Tuto obecnou funkci, která udává, jakým způsobem dochází ke složení signálů, nahradíme lineární aproximací z Taylorova rozvoje. Dostáváme tak, že vektor signálů z mikrofonů \mathbf{Y} se dá rozložit stejně jako v modelu faktorové analýzy (část 3.1) na $\mathbf{Y} = \boldsymbol{\mu} + \mathbf{\Lambda}\mathbf{F} + \boldsymbol{\varepsilon}$, kde význam parametrů $\boldsymbol{\mu}$ a $\mathbf{\Lambda}$ je zachován, ovšem $\boldsymbol{\varepsilon}$ v sobě nyní zahrnuje nejen chybu vzniklou šumem, ale i onu chybu z aproximace pomocí Taylorova polynomu. Celkem naměříme n různých pozorování vektoru \mathbf{Y} a naší úlohou je z nich co nejpřesněji stanovit původní hodnoty \mathbf{F} signálů jednotlivých účastníků.

◇

Na první pohled by se mohlo zdát, že takto zachycený problém přesně odpovídá modelu, kterému jsme se až doposud věnovali. Ovšem zdání zde klame, neboť je zde skryto několik detailů, které situaci komplikují. Jedním z nich může být například fakt, že v praxi zřejmě nastane situace, kdy mikrofonů nebude více než lidí. Doposud jsme totiž vždy předpokládali, že $m > k$. To nám znemožňuje dívat se na tento problém jako na model faktorové analýzy představený v části 3.1. Podmínky identifikace a řešitelnosti se totiž tímto hroutí. Lze však i nadále uvažovat bayesovskou specifikaci modelu.

3.6.1 Model s korelovanými faktory

Dejme tomu, že stále můžeme předpokládat, že hodnoty jednotlivých signálů (mikrofony i lidé) jsou normálně rozdělené. Podobně jako signály z mikrofonů nyní mohou mít i jednotlivé signály lidí jinou střední hodnotu a variabilitu. Když se navíc řádně zamyslíme nad tím, jak taková konverzace probíhá, tak přijdeme na to, že jednotlivé signály lidí mohou být (spíše negativně) korelovány. V rámci skupinky totiž v daný moment mluví spíše jediný člověk, zatímco ostatní mu naslouchají. Teprve až domluví tento řečník, tak začne mluvit někdo jiný. To nás nabádá k tomu, abychom opustili předpoklad

$$\mathbf{F} \sim \mathbf{N}_k(\mathbf{0}, \mathbb{I}_k) \quad \text{a nahradili ho předpokladem} \quad \mathbf{F} \sim \mathbf{N}_k(\mathbf{F}_0, \boldsymbol{\Omega}),$$

kde $\mathbf{F}_0 \in \mathbb{R}^k$ a $0 < \boldsymbol{\Omega} \in \mathbb{R}^{k \times k}$ jsou hyperparametry. Toto by se projevilo ve změně apriorního rozdělení zavedeném v části 3.2.1. Ovšem i když je následný přístup k odhadu tohoto modelu analogický, tak se ideově jedná o úplně jiný model.

Opět by se ukázalo, že jsme sice schopni odvodit tvar marginální aposteriorní hustoty \mathbb{F} , ale výsledné rozdělení by nepatřilo k žádným běžně používaným, se

kterými se jednoduše pracuje. Proto bychom se museli spokojit s plně podmíněným rozdělením a sestavením Gibbsova či ICM algoritmu pro odhad parametrů. Za předpokladu apriorní nezávislosti \mathbb{F} na ostatních parametrech modelu se problém určení plně podmíněného rozdělení \mathbb{F} zúží na doplnění výrazu

$$\left(\mathbb{F}\mathbf{\Lambda}^\top - \left(\mathbb{Y} - \mathbf{1}_n\boldsymbol{\mu}^\top\right)\right)\boldsymbol{\Sigma}^{-1}\left(\mathbb{F}\mathbf{\Lambda}^\top - \left(\mathbb{Y} - \mathbf{1}_n\boldsymbol{\mu}^\top\right)\right)^\top + (\mathbb{F} - \mathbb{F}_0)\boldsymbol{\Omega}^{-1}(\mathbb{F} - \mathbb{F}_0)^\top$$

na čtverec. K tomu lze opět použít tvrzení B7, kde hlavním parametrem je $\mathbb{X} = \mathbb{F}$. Dostaneme tak, že stěžejní částí závisující na \mathbb{F} je

$$\begin{aligned} & (\mathbb{F} - \vec{\mathbb{F}}) \left[\mathbf{\Lambda}^\top \boldsymbol{\Sigma}^{-1} \mathbf{\Lambda} + \boldsymbol{\Omega}^{-1} \right] (\mathbb{F} - \vec{\mathbb{F}})^\top, \\ \text{kde } \vec{\mathbb{F}} &= \left(\left(\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top \right) \boldsymbol{\Sigma}^{-1} \mathbf{\Lambda} + \mathbb{F}_0 \boldsymbol{\Omega}^{-1} \right) \left[\mathbf{\Lambda}^\top \boldsymbol{\Sigma}^{-1} \mathbf{\Lambda} + \boldsymbol{\Omega}^{-1} \right]^{-1}. \end{aligned}$$

Odtud plyne plně podmíněné rozdělení

$$\mathbb{F} | \mathbb{Y}, \boldsymbol{\mu}, \mathbf{\Lambda}, \boldsymbol{\Sigma} \sim \mathbf{N}_{n \times k} \left(\vec{\mathbb{F}}, \mathbb{I}_n \otimes \left[\mathbf{\Lambda}^\top \boldsymbol{\Sigma}^{-1} \mathbf{\Lambda} + \boldsymbol{\Omega}^{-1} \right]^{-1} \right).$$

Ovšem velmi problematické je zde určení hyperparametrů \mathbb{F}_0 a $\boldsymbol{\Omega}$, neboť typicky nemáme žádné historické údaje. Každý řádek matice \mathbb{F}_0 je tvořen vektorem \mathbf{F}_0 , to vyjadřuje, že je střední hodnota signálů od účastníků mezi všemi $i \in \{1, \dots, n\}$ stále stejná. Každému účastníkovi tedy musíme přiřadit apriorní představu o tom, jaká je jeho střední hodnota signálu. Specifikace rozptylu bývá ještě zásadnější z toho důvodu, abychom se vyhnuli přílišné informativnosti. Parametr $\boldsymbol{\Omega}$ proto často stanovujeme jako náhodný parametr z inverzního Wishartova rozdělení s již pevnými hyperparametry. Můžeme také zkusit matici obsahující mimo diagonálu záporná čísla, což je v souladu s negativní korelací signálů jednotlivých lidí.

3.6.2 Model se známými regresory

Příklad 4 nás může inspirovat k dalšímu zobecnění modelu, stačí položit si otázku: *Co by to bylo za večírek, kdyby na něm nehrála žádná hudba?* Tedy konverzace, kterou mikrofony zaznamenávají, nejsou jediné zdroje signálu v místnosti. Dejme tomu, že existuje dalších q zdrojů signálu $\mathbf{Z} = (Z_1, \dots, Z_q)$, které však známe předem¹. Tyto signály se tedy také podílí na výsledných hodnotách \mathbf{Y} . Funkce, která obecně stojí za tím, jakým způsobem je výsledný signál \mathbf{Y} stvořen, může být ještě daleko složitější. Opět se však omezme na lineární aproximaci, podle které by mělo $\mathbf{Y} = \boldsymbol{\mu} + \mathbb{C}\mathbf{Z} + \mathbf{\Lambda}\mathbf{F} + \boldsymbol{\varepsilon}$, kde $\mathbb{C}^\top = (\mathbf{c}_1, \dots, \mathbf{c}_q) \in \mathbb{R}^{m \times q}$ je matice parametrů, jejíž řádky \mathbf{c}_j udávají, jakým způsobem \mathbf{Z} ovlivňuje jednotlivé výsledné signály Y_j , $j \in \{1, \dots, m\}$.

Pro n realizací vektoru \mathbf{Y} pak v maticovém zápisu platí

$$\mathbb{Y} = \mathbf{1}_n \boldsymbol{\mu}^\top + \mathbb{Z} \mathbb{C}^\top + \mathbb{F} \mathbf{\Lambda}^\top + \mathbb{E},$$

kde $\mathbb{Z}^\top = (\mathbf{Z}_1, \dots, \mathbf{Z}_n) \in \mathbb{R}^{n \times q}$ a ostatní matice jsou definovány jako v úvodu části 3.1.2. To lze opět zkrátit zavedením matic $\mathbb{B} = (\boldsymbol{\mu}, \mathbb{C}, \mathbf{\Lambda})$ a $\mathbb{X} = (\mathbf{1}_n, \mathbb{Z}, \mathbb{F})$ na $\mathbb{Y} = \mathbb{X} \mathbb{B}^\top + \mathbb{E}$. Za předpokladu normality chyb modelu $\mathbb{E} \sim \mathbf{N}_{n \times m}(\mathbb{O}_{n \times m}, \mathbb{I}_n \otimes \boldsymbol{\Sigma})$

¹Pokud víme, jakou píseň reproduktory přehrávaly v danou chvíli, tak víme i přesnou hodnotu signálu, který byl těmito reproduktory emitován.

potom sestavíme věrohodnost modelu analogicky jako v rovnici (2.8). Jestliže si potom zavedeme apriorní rozdělení jednotlivých parametrů analogicky jako v částech 3.2.1 či 3.5, tak můžeme opět využívat tvrzení B7 a odvodit tak plně podmíněná rozdělení. Ukažme si výsledek za předpokladu, že $\mathbb{B}|\Sigma \sim \mathbf{N}_{m \times (1+q+k)}(\mathbb{B}_0, \mathbb{D})$, kde \mathbb{D} je blokově diagonální matice:

$$\mathbb{D} = \begin{pmatrix} d_{0,0} & \mathbf{0}_q^\top & \mathbf{0}_k^\top \\ \mathbf{0}_q & \mathbb{D}_{11} & \mathbb{O}_{q \times k} \\ \mathbf{0}_k & \mathbb{O}_{k \times q} & \mathbb{D}_{22} \end{pmatrix}.$$

Tvrzení 14. *Uvažme model vycházející z bayesovské faktorové analýzy nastíněný výše, kde máme zahrnuty také známé regresory \mathbf{Z} . Nechť apriorní rozdělení je dáno následovně:*

$$\begin{aligned} \mathbb{F} &\sim \mathbf{N}_{n \times k}(\mathbb{F}_0, \mathbb{I}_n \otimes \boldsymbol{\Omega}), \\ \boldsymbol{\mu}|\Sigma &\sim \mathbf{N}_m(\boldsymbol{\mu}_0, \Sigma d_{0,0}), \\ \mathbb{C}|\Sigma &\sim \mathbf{N}_{m \times q}(\mathbb{C}_0, \Sigma \otimes \mathbb{D}_{11}), \\ \boldsymbol{\Lambda}|\Sigma &\sim \mathbf{N}_{m \times k}(\boldsymbol{\Lambda}_0, \Sigma \otimes \mathbb{D}_{22}), \\ \Sigma &\sim \text{IW}_m(\mathbb{Q}, \nu), \end{aligned}$$

kde faktory \mathbb{F} jsou apriorně nezávislé na ostatních parametrech. Pak pro plně podmíněná rozdělení parametrů \mathbb{F} , $\boldsymbol{\mu}$, \mathbb{C} , $\boldsymbol{\Lambda}$ a Σ platí

$$\begin{aligned} \mathbb{F}|\mathbb{Y}, \boldsymbol{\mu}, \mathbb{C}, \boldsymbol{\Lambda}, \Sigma &\sim \mathbf{N}_{n \times k}(\widehat{\mathbb{F}}(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \mathbb{C}, \Sigma), \mathbb{I}_n \otimes [\boldsymbol{\Lambda}^\top \Sigma^{-1} \boldsymbol{\Lambda} + \boldsymbol{\Omega}^{-1}]^{-1}), \\ \boldsymbol{\mu}|\mathbb{Y}, \mathbb{F}, \mathbb{C}, \boldsymbol{\Lambda}, \Sigma &\sim \mathbf{N}_m(\widehat{\boldsymbol{\mu}}(\mathbb{F}, \mathbb{C}, \boldsymbol{\Lambda}), \Sigma [n + d_{0,0}^{-1}]^{-1}), \\ \mathbb{C}|\mathbb{Y}, \mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \Sigma &\sim \mathbf{N}_{m \times q}(\widehat{\mathbb{C}}(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda}), \Sigma \otimes [\mathbf{Z}^\top \mathbf{Z} + \mathbb{D}_{11}^{-1}]^{-1}), \\ \boldsymbol{\Lambda}|\mathbb{Y}, \mathbb{F}, \boldsymbol{\mu}, \mathbb{C}, \Sigma &\sim \mathbf{N}_{m \times k}(\widehat{\boldsymbol{\Lambda}}(\mathbb{F}, \boldsymbol{\mu}, \mathbb{C}), \Sigma \otimes [\mathbb{F}^\top \mathbb{F} + \mathbb{D}_{22}^{-1}]^{-1}), \\ \Sigma|\mathbb{Y}, \mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \mathbb{C} &\sim \text{IW}_m(\boldsymbol{\Psi}(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \mathbb{C}), n + \nu + k + q + 1), \end{aligned}$$

kde

$$\begin{aligned} \widehat{\mathbb{F}}(\boldsymbol{\mu}, \mathbb{C}, \boldsymbol{\Lambda}, \Sigma) &= ((\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top - \mathbf{Z} \mathbb{C}^\top) \Sigma^{-1} \boldsymbol{\Lambda} + \mathbb{F}_0 \boldsymbol{\Omega}^{-1}) [\boldsymbol{\Lambda}^\top \Sigma^{-1} \boldsymbol{\Lambda} + \boldsymbol{\Omega}^{-1}]^{-1}, \\ \widehat{\boldsymbol{\mu}}(\mathbb{F}, \mathbb{C}, \boldsymbol{\Lambda}) &= \frac{n}{n + d_{0,0}^{-1}} (\overline{\mathbb{Y}} - \mathbb{C} \overline{\mathbf{Z}} - \boldsymbol{\Lambda} \overline{\mathbf{F}}) + \frac{d_{0,0}^{-1}}{n + d_{0,0}^{-1}} \boldsymbol{\mu}_0, \\ \text{kde } \overline{\mathbb{Y}} &= \frac{1}{n} \mathbb{Y}^\top \mathbf{1}_n, \quad \overline{\mathbf{Z}} = \frac{1}{n} \mathbf{Z}^\top \mathbf{1}_n, \quad \overline{\mathbf{F}} = \frac{1}{n} \mathbb{F}^\top \mathbf{1}_n, \\ \widehat{\mathbb{C}}(\mathbb{F}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) &= [(\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top - \mathbb{F} \boldsymbol{\Lambda}^\top)^\top \mathbf{Z} + \mathbb{C}_0 \mathbb{D}_{11}^{-1}] [\mathbf{Z}^\top \mathbf{Z} + \mathbb{D}_{11}^{-1}]^{-1} \\ &= [\widehat{\mathbb{C}}_C \mathbf{Z}^\top \mathbf{Z} + \mathbb{C}_0 \mathbb{D}_{11}^{-1}] [\mathbf{Z}^\top \mathbf{Z} + \mathbb{D}_{11}^{-1}]^{-1}, \\ \widehat{\mathbb{C}}_C &= (\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top - \mathbb{F} \boldsymbol{\Lambda}^\top)^\top \mathbf{Z} (\mathbf{Z}^\top \mathbf{Z})^{-1}, \quad \text{pokud } \mathbf{Z}^\top \mathbf{Z} > 0, \\ \widehat{\boldsymbol{\Lambda}}(\mathbb{F}, \boldsymbol{\mu}, \mathbb{C}) &= [(\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top - \mathbf{Z} \mathbb{C}^\top)^\top \mathbb{F} + \boldsymbol{\Lambda}_0 \mathbb{D}_{22}^{-1}] [\mathbb{F}^\top \mathbb{F} + \mathbb{D}_{22}^{-1}]^{-1} \\ &= [\widehat{\boldsymbol{\Lambda}}_\Lambda \mathbb{F}^\top \mathbb{F} + \boldsymbol{\Lambda}_0 \mathbb{D}_{22}^{-1}] [\mathbb{F}^\top \mathbb{F} + \mathbb{D}_{22}^{-1}]^{-1}, \\ \widehat{\boldsymbol{\Lambda}}_\Lambda &= (\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top - \mathbf{Z} \mathbb{C}^\top)^\top \mathbb{F} (\mathbb{F}^\top \mathbb{F})^{-1}, \quad \text{pokud } \mathbb{F}^\top \mathbb{F} > 0, \end{aligned}$$

$$\begin{aligned}\Psi(\mathbb{F}, \boldsymbol{\mu}, \mathbb{C}, \boldsymbol{\Lambda}) &= \left(\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top - \mathbb{Z} \mathbb{C}^\top - \mathbb{F} \boldsymbol{\Lambda}^\top \right)^\top \left(\mathbb{Y} - \mathbf{1}_n \boldsymbol{\mu}^\top - \mathbb{Z} \mathbb{C}^\top - \mathbb{F} \boldsymbol{\Lambda}^\top \right) + \\ &+ (\boldsymbol{\mu} - \boldsymbol{\mu}_0) d_{0,0}^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0)^\top + (\mathbb{C} - \mathbb{C}_0) \mathbb{D}_{11}^{-1} (\mathbb{C} - \mathbb{C}_0)^\top + \\ &+ (\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0) \mathbb{D}_{22}^{-1} (\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0)^\top + \mathbb{Q}.\end{aligned}$$

Důkaz. Lze dokázat analogickým způsobem jako tvrzení 6. □

Na základě tohoto tvrzení pak lze sestavit Gibbsův či ICM algoritmus, pomocí kterých lze získat odhady jednotlivých parametrů. Algoritmy budou velmi podobné těm sestaveným v částech 3.3 a 3.4. Stačí přidat parametr \mathbb{C} , regresory \mathbb{Z} a modifikovat příslušné parametry po vzoru tvrzení 14.

Ovšem u možných zobecnění jsme zdaleka neskončili. Jedno z nich nabízí samotná Taylorova aproximace funkce, která mixuje jednotlivé signály. Místo lineární aproximace, bychom mohli použít například kvadratickou aproximaci. To by ale vyžadovalo pracovat s třírozměrnými formami, tj. útvary z $\mathbb{R}^{n \times k \times l}$. S těmi už se pracuje daleko hůře než s maticovými parametry, a proto se tomuto zobecnění nebudeme nadále věnovat.

Měření $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ většinou pořizujeme v časech $t_1 < \dots < t_n$. Tedy by se na ně šlo dívat jako na mnohorozměrnou časovou řadu. V příkladě 4 lze očekávat, že střední hodnota signálu se nějakým způsobem vyvíjí v čase. Na počátku večírku by signály byly nízké, pak by se s přibývajícím počtem lidí velikost signálů zvětšovala, až by dosáhla bodu, kdy se drží střední hodnota signálu přibližně na stejné úrovni. S blížícím se koncem samotného večírku by měřené signály slábly. Stačilo by tedy ke známým regresorům přidat vhodnou parametrizaci času pro podchycení vlivu času na střední hodnotu modelu.

Ovšem ošetření toho, že po sobě měřené signály jsou více korelované než ty, které jsou měřeny dále od sebe, již není jednoduché. Museli bychom použít metod mnohorozměrných časových řad. Do nich se zde ale již pouštět nehodláme.

V celé této kapitole jsme se seznámili se základním modelem bayesovské faktorové analýzy. Za různých předpokladů modelu jsme si odvodili plně podmíněná rozdělení neznámých parametrů. Na jejich základě jsme sestavili algoritmy, pomocí nichž lze tyto neznámé parametry odhadnout.

Tyto myšlenky však v kontextu této problematiky nejsou nijak pokročilé. Jedná se pouze o odrazový můstek, na základě kterého lze budovat složitější a v praxi daleko lépe použitelnější algoritmy. Jeden z modelů, který se v praxi používá pro své elegantní řešení odhadu počtu latentních faktorů, si představíme v následující kapitole.

4. Bayesovský odhad počtu skrytých faktorů

V minulé kapitole jsme si odvodili, jak odhadovat parametry modelu faktorové analýzy pomocí bayesovského přístupu, je-li počet skrytých faktorů předem znám (viz příklad 3). Pokud není počet faktorů předem znám, tak se při použití klasického přístupu zkusí několik různých voleb počtu faktorů k a vybere se ten s přijatelně interpretovatelnými výsledky. V literatuře byla navržena celá řada různých kritérií, která se ale v celé řadě případů neshodují. Bayesovský pohled na tento problém však nabízí hned několik výhod. Především ušetří statistika zásadních rozhodnutí ovlivňujících strukturu modelu a nabízí alternativní způsob odhadu počtu skrytých faktorů. Za tímto účelem ale budeme muset zásadně pozměnit definici modelu bayesovské faktorové analýzy.

Existuje vícero způsobů, jak si zadefinovat model s proměnlivým počtem faktorů. Zde se budeme držet postupu navrženém v článku Contiové, Frühwirth-Schnatterové, Heckmana a Piateka (Conti a kol., 2014). Pracuje se zde s modelem určujících faktorů (*dedicated factor model*), který si nyní představíme.

4.1 Model určujících faktorů

Zásadním principem zde bude předpoklad, že za naším modelem může stát maximálně K^{\max} latentních faktorů \mathbf{F} . Předpokládáme, že pro každé měření Y_j , $j \in \{1, \dots, m\}$ existuje *nejvýše jeden* faktor F_k , $k \in \{1, \dots, K^{\max}\}$, který ovlivnil měření Y_j . Může se tak stát, že se zde vyskytnou faktory ovlivňující jeden a více měření (tzv. aktivní faktory), ale mohou zde být i faktory, které neovlivní jedině z těchto měření, tzv. neaktivní faktory. Přidělení faktoru k nějakému pozorování podléhá pravděpodobnostnímu mechanismu, který si ještě podrobněji představíme. Počet aktivních faktorů K je tak náhodnou veličinou s hodnotami v $\{1, \dots, K^{\max}\}$. Bayesovským odhadem počtu faktorů by tak mohla být podmíněná střední hodnota K při dostupných datech.

Uvažme zobecněný model faktorové analýzy se známými regresory představený v části 3.6.2, kde modelujeme m -rozměrný vektor \mathbf{Y} jako součet střední hodnoty $\boldsymbol{\mu}$, lineární kombinace $\mathbb{C}\mathbf{Z}$ známých regresorů \mathbf{Z} dimenze q , součinu faktorových zátěží $\boldsymbol{\Lambda}$ se skrytými faktory \mathbf{F} dimenze K^{\max} a vektoru chyb modelu $\boldsymbol{\varepsilon}$, tj. $\mathbf{Y} = \boldsymbol{\mu} + \mathbb{C}\mathbf{Z} + \boldsymbol{\Lambda}\mathbf{F} + \boldsymbol{\varepsilon}$. Slučme zde pro zjednodušení zápisu $\boldsymbol{\mu}$ a $\mathbb{C}\mathbf{Z}$ do jediného výrazu, tedy položíme $\mathbb{B} = (\boldsymbol{\mu}, \mathbb{C})$ (matice neznámých koeficientů) a $\mathbf{X} = (1, \mathbf{Z}^\top)^\top$ (známé regresory). Rozklad \mathbf{Y} tedy lze zapsat ve tvaru

$$\mathbf{Y} = \mathbb{B}\mathbf{X} + \boldsymbol{\Lambda}\mathbf{F} + \boldsymbol{\varepsilon}. \quad (4.1)$$

Zásadní změna však nastává ve struktuře faktorových zátěží $\boldsymbol{\Lambda}$ (nyní o rozměrech m a K^{\max}). Jelikož zde předpokládáme, že každou složku vektoru \mathbf{Y} ovlivňuje *nejvýše jeden* z faktorů, tak $\boldsymbol{\Lambda}$ bude velmi řídkou maticí (bude obsahovat nejméně $m(K^{\max} - 1)$ nul). Zavedme si pomocnou $(m \times K^{\max})$ -matici $\boldsymbol{\Delta}$ tvořenou prvky $\Delta_{j,k} \in \{0, 1\}$, která určuje, které zátěže $\lambda_{j,k}$ jsou nulové. Tedy

$$\Delta_{j,k} = 0 \implies \lambda_{j,k} = 0, \quad \Delta_{j,k} = 1 \implies \lambda_{j,k} \in \mathbb{R}.$$

Počet aktivních faktorů K odpovídá počtu nenulových sloupců matice Δ .

Nechť Δ_j značí jednotlivé řádky matice Δ , tj. $\Delta^\top = (\Delta_1, \dots, \Delta_m)$. Pak $\Delta_j = \mathbf{0}_{K^{\max}} =: e_0$ znamená, že ani jeden z faktorů \mathbf{F} neovlivňuje měření Y_j . Je-li $\Delta_j = (0, \dots, 0, 1, 0, \dots, 0)^\top =: e_k$, kde 1 se vyskytuje na k -té pozici, tak měření Y_j bylo ovlivněno faktorem F_k . Velikost tohoto vlivu je pak dána hodnotou $\lambda_{j,k}$. Dále budeme také používat značení δ_j , které je definováno jako $\delta_j = k$ právě tehdy, když $\Delta_j = e_k$. Jednotlivé rovnosti z rovnice (4.1) se tak pro každé $j \in \{1, \dots, m\}$ zjednodušují na tvar

$$\begin{aligned} Y_j &= \mathbf{b}_j^\top \mathbf{X} + \varepsilon_j, & \text{pokud } \delta_j = 0, \\ Y_j &= \mathbf{b}_j^\top \mathbf{X} + \lambda_{j,\delta_j} F_{\delta_j} + \varepsilon_j, & \text{pokud } \delta_j \in \{1, \dots, K^{\max}\}, \end{aligned}$$

kde \mathbf{b}_j je j -tý řádek matice \mathbb{B} , tj. $\mathbb{B}^\top = (\mathbf{b}_1, \dots, \mathbf{b}_m)$.

Pro identifikaci tohoto modelu bude zapotřebí učinit několik předpokladů. Opět potřebujeme, aby jak faktory \mathbf{F} , tak chyby modelu ε měly nulovou střední hodnotu, tj. $\mathbf{E} \mathbf{F} = \mathbf{0}_{K^{\max}}$ a $\mathbf{E} \varepsilon = \mathbf{0}_m$. O rozptylové matici Σ chyb modelu budeme předpokládat, že je diagonální s kladnými prvky na diagonále, tj. $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_m^2)$, $\sigma_j^2 > 0, j \in \{1, \dots, m\}$. Nadále předpokládáme, že chyby modelu ε a faktory \mathbf{F} jsou nekorelované. Ovšem mezi jednotlivými faktory povolíme korelovanost jako v modelu v části 3.6.1. Za předpokladů výše lze pro varianční matici vektoru \mathbf{Y} snadno odvodit, že platí

$$\text{var } \mathbf{Y} = \Lambda \Omega \Lambda^\top + \Sigma, \quad \text{kde } \text{var } \mathbf{F} = \Omega > 0 \quad (4.2)$$

je varianční matice faktorů \mathbf{F} . Tuto matici jsme ovšem schopni rozložit na

$$\Omega = \mathbb{V}^{\frac{1}{2}} \mathbb{K} \mathbb{V}^{\frac{1}{2}}, \quad \text{kde } \mathbb{V} = \text{diag } \Omega, \quad \mathbb{K} = \text{corr } \mathbf{F}. \quad (4.3)$$

Potom bychom mohli zapsat $\Lambda \mathbf{F} = (\Lambda \mathbb{V}^{-\frac{1}{2}}) (\mathbb{V}^{\frac{1}{2}} \mathbf{F}) = \Lambda^* \mathbf{F}^*$ a dostat tak jiné zátěže a faktory, které splňují požadavky výše. Ovšem takovéto \mathbf{F}^* má pak varianční matici rovnou přímo korelační matici \mathbb{K} . Proto jedna z podmínek zaručující jednoznačnost bude, že faktory \mathbf{F} mají jednotkový rozptyl, tj. $\text{diag } \Omega = \mathbf{1}_{K^{\max}}$. Ovšem toto stále neurčuje faktory jednoznačně, ještě je totiž například možné považovat za zátěže $-\Lambda^*$ a za faktory $-\mathbf{F}^*$. Nežádoucí je také situace, kdy $\text{rank } \Omega < K^{\max}$, neboť pak by existoval takový faktor, který by byl lineární kombinací ostatních.

Anderson a Rubin (1956) ve svém článku uvádějí různá kritéria pro identifikaci v modelu faktorové analýzy. Zde budeme potřebovat větu 5.1 z tohoto článku.

Věta 15. (*Theorem 5.1 Anderson a Rubin, 1956*) *Nechť $\Lambda \in \mathbb{R}^{m \times k}$ a $\Sigma \in \mathbb{R}^{m \times m} > 0$ je diagonální matice. Jestliže po odstranění libovolného řádku matice Λ existují dvě podmatice složené z různých řádků, jejichž hodnota je stejná jako hodnota Λ , pak Σ je identifikováno jednoznačně a Λ je identifikováno jednoznačně až na přenásobení libovolnou ortonormální maticí zprava. To znamená, že pokud nastane $\Lambda \Lambda^\top + \Sigma = \tilde{\Lambda} \tilde{\Lambda}^\top + \tilde{\Sigma}$, pak $\Sigma = \tilde{\Sigma}$ a existuje ortonormální matice \mathbb{G} taková, že $\Lambda = \tilde{\Lambda} \mathbb{G}$.*

Tuto větu použijeme na rozklad (4.2) tak, abychom získali identifikovatelnost Σ v modelu určujících faktorů. Pro splnění podmínek této věty však budeme muset přidat dodatečný předpoklad o tom, že každý faktor F_k je buď neaktivní (k -tý sloupec Δ je nulový), nebo ovlivňuje alespoň 3 pozorování. Za tímto účelem si definujme počet nenulových políček v k -tém sloupci matice Δ jako $n_k(\Delta) = \sum_{j=1}^m \Delta_{j,k}$.

Věta 16. *V modelu určujících faktorů, kde $\text{diag } \Omega = \mathbf{1}_{K^{\max}}$, $\text{rank } \Omega = K^{\max}$ a pro každé $k \in \{1, \dots, K^{\max}\}$ je buď $n_k(\Delta) = 0$, nebo $n_k(\Delta) \geq 3$, jsou parametry identifikovatelné až na triviální rotace. Tedy matice Δ je identifikovatelná až na permutaci sloupců. Faktorové zátěže Λ a podmatice Ω odpovídající nenulovým sloupcům matice Δ jsou identifikovatelné až na stejnou permutaci sloupců a změnu znaménka.*

Důkaz. Podrobný důkaz je k nalezení v článku Contiové a kol. (Conti a kol., 2014). Zde si ukážeme hlavní myšlenky tohoto důkazu.

Nejprve si uvědomme, že díky našemu předpokladu je splněna podmínka z věty 15. Ať už je odstraněn nulový či nenulový řádek matice Λ , tak ve zbylé matici vždy najdeme každý nenulový řádek alespoň dvakrát. Proto lze vždy nalézt dvě disjunktní podmatice, které by měly stejnou hodnotu (rovnou počtu aktivních faktorů K). Odtud dostáváme, že matice Σ je identifikovatelná jednoznačně. Tedy když $\Lambda\Omega\Lambda^\top + \Sigma = \tilde{\Lambda}\tilde{\Omega}\tilde{\Lambda}^\top + \tilde{\Sigma}$, tak $\Sigma = \tilde{\Sigma}$. Tedy také dostáváme, že $\Lambda\Omega\Lambda^\top = \tilde{\Lambda}\tilde{\Omega}\tilde{\Lambda}^\top$.

Na diagonále matice $\Lambda\Omega\Lambda^\top$ nalezneme $\lambda_j^\top \lambda_j, j \in \{1, \dots, m\}$, kde λ_j jsou řádky matice Λ , tj. $\Lambda^\top = (\lambda_1, \dots, \lambda_m)$, neboť dle našeho předpokladu jsou diagonální prvky $\omega_{k,k} = 1, k \in \{1, \dots, K^{\max}\}$. Tedy musí být splněno, že $\lambda_j^\top \lambda_j = \tilde{\lambda}_j^\top \tilde{\lambda}_j$, a přitom víme, že nejvýše jeden prvek tohoto vektoru je nenulový. Odtud dostáváme, že λ_j je nulový právě tehdy, když $\tilde{\lambda}_j$ je nulový. Zároveň odtud vidíme, že případný nenulový prvek λ_j musí být až na znaménko shodný s nenulovým prvkem vektoru $\tilde{\lambda}_j$. Nicméně toto neznamená, že tento nenulový prvek musí být určen tomu samému faktoru k .

Podívejme se nyní na prvky mimo diagonálu matice $\Lambda\Omega\Lambda^\top$. Uvažme takový pár j a j' , který přísluší veličinám Y_j a $Y_{j'}$, které byly ovlivněny nějakým faktorem, tj. $\delta_j \neq 0 \neq \delta_{j'}$ (nenulové budou i $\tilde{\delta}_j$ a $\tilde{\delta}_{j'}$ z alternativního rozkladu). Musí platit, že $\lambda_j^\top \omega_{\delta_j, \delta_{j'}} \lambda_{j'} = \tilde{\lambda}_j^\top \tilde{\omega}_{\tilde{\delta}_j, \tilde{\delta}_{j'}} \tilde{\lambda}_{j'}$. Jelikož nenulové prvky λ_j a $\tilde{\lambda}_j$ musí být až na znaménko stejné, tak musí i $\omega_{\delta_j, \delta_{j'}}$ a $\tilde{\omega}_{\tilde{\delta}_j, \tilde{\delta}_{j'}}$ až na znaménko stejné. Pokud jsou měření j a j' určeny tím samým faktorem $k = \delta_j = \delta_{j'}$, pak jsme zpět u diagonálního prvku $\omega_{k,k} = 1$. Jsou-li však ve druhé reprezentaci určeny jinými faktory, tj. $\tilde{\delta}_j \neq \tilde{\delta}_{j'}$, tak dostáváme, že nediagonální prvek $\tilde{\omega}_{\tilde{\delta}_j, \tilde{\delta}_{j'}} = \pm \omega_{k,k} = \pm 1$. Ve druhé reprezentaci se tak vyskytují dva perfektně korelované faktory, což je ve sporu s předpokladem $\text{rank } \tilde{\Omega} = K^{\max}$. Tedy musí platit, že $k = \delta_j = \delta_{j'}$ právě tehdy, když $\tilde{k} = \tilde{\delta}_j = \tilde{\delta}_{j'}$.

Odtud už je zřejmé, že reprezentace Δ a $\tilde{\Delta}$ mají stejný počet nenulových sloupců. Nicméně stále se nemusí jednat o stejné faktory k a \tilde{k} . Ovšem díky shodným nulovým řádkům již můžeme tvrdit, že Δ je identifikovatelná, až na permutaci sloupců. Potom Λ je identifikovatelná až na tu samou permutaci a možnou změnu znaménka. Identifikovatelnost Ω je nutno zúžit jen na sloupce odpovídající aktivním faktorům, neboť jen pro nenulové $\delta_j, \delta_{j'}, \tilde{\delta}_j, \tilde{\delta}_{j'}$ jsme ukázali, že $\omega_{\delta_j, \delta_{j'}}^2 = \omega_{\tilde{\delta}_j, \tilde{\delta}_{j'}}^2$.

□

Neidentifikovatelnost kovarianční matice neaktivních faktorů ani kovariancí mezi aktivními a neaktivními faktory nás však trápit nemusí, neboť skutečně používané jsou pouze aktivní faktory.

Mějme daný počet měření m . Potom nejvyšší možný počet aktivních faktorů K může být $\lfloor \frac{m}{3} \rfloor$, neboť každý aktivní faktor musí mít nejméně 3 měření, která ovlivňuje (dle předpokladu identifikovatelnosti ve znění věty 16). Jak můžeme vidět v tabulce 3.1, pro $m = 3$ je omezení $\lfloor \frac{m}{3} \rfloor$ na maximální počet faktorů stejné jako Ledermannova mez, pro $m > 3$ však již $m/3 < \phi(m)$. Požadavek $K^{\max} < m/3$ tak v sobě rovnou zaručí i splnění Ledermannovy meze.

4.2 Věrohodnost a apriorní rozdělení

Stejně jako doposud budeme předpokládat, že chyby modelu $\boldsymbol{\varepsilon}$ a faktory \mathbf{F} pocházejí z normálního rozdělení, tedy

$$\boldsymbol{\varepsilon} \sim \mathbf{N}_m(\mathbf{0}_m, \boldsymbol{\Sigma}) \quad \text{a} \quad \mathbf{F} \sim \mathbf{N}_{K^{\max}}(\mathbf{0}_{K^{\max}}, \mathbb{K}).$$

Pro vektor \mathbf{Y} tak platí

$$\mathbf{Y}|\mathbf{X} \sim \mathbf{N}_m(\mathbb{B}\mathbf{X}, \boldsymbol{\Upsilon} := \boldsymbol{\Lambda}\mathbb{K}\boldsymbol{\Lambda}^\top + \boldsymbol{\Sigma}) \quad \text{či} \quad \mathbf{Y}|\mathbf{X}, \mathbf{F} \sim \mathbf{N}_m(\mathbb{B}\mathbf{X} + \boldsymbol{\Lambda}\mathbf{F}, \boldsymbol{\Sigma}).$$

K dispozici budeme mít opět n realizací $\mathbf{Y}_i, i \in \{1, \dots, n\}$ náhodného vektoru \mathbf{Y} spolu s příslušnými regresory $\mathbf{X}_i = (1, \mathbf{Z}_i^\top)^\top$. Zavedme si analogicky jako v předchozích modelech matice $\mathbb{Y}^\top = (\mathbf{Y}_1, \dots, \mathbf{Y}_m)$, $\mathbb{X}^\top = (\mathbf{X}_1, \dots, \mathbf{X}_n)$, $\mathbb{F}^\top = (\mathbf{F}_1, \dots, \mathbf{F}_n)$ a $\mathbb{E}^\top = (\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n)$. Potom v maticovém zápisu platí

$$\mathbb{Y} = \mathbb{X}\mathbb{B}^\top + \mathbb{F}\boldsymbol{\Lambda}^\top + \mathbb{E}.$$

Věrohodnost tohoto modelu sestavíme analogicky jako v předchozích modelech:

$$p(\mathbb{Y}|\mathbb{X}, \mathbb{F}, \mathbb{B}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \cdot \exp \left\{ -\frac{1}{2} \text{Tr} \boldsymbol{\Sigma}^{-1} (\mathbb{Y} - \mathbb{X}\mathbb{B}^\top - \mathbb{F}\boldsymbol{\Lambda}^\top)^\top (\mathbb{Y} - \mathbb{X}\mathbb{B}^\top - \mathbb{F}\boldsymbol{\Lambda}^\top) \right\} \quad (4.4)$$

$$\propto \prod_{j=1}^m (\sigma_j^2)^{-\frac{n}{2}} \cdot \exp \left\{ -\frac{1}{2\sigma_j^2} (\mathbb{Y}_{\bullet j} - \mathbb{X}\mathbf{b}_j - \mathbb{F}\boldsymbol{\lambda}_j)^\top (\mathbb{Y}_{\bullet j} - \mathbb{X}\mathbf{b}_j - \mathbb{F}\boldsymbol{\lambda}_j) \right\} \quad (4.5)$$

$$\propto \prod_{j=1}^m (\sigma_j^2)^{-\frac{n}{2}} \cdot \exp \left\{ -\frac{1}{2\sigma_j^2} \sum_{i=1}^n (Y_{i,j} - \mathbf{X}_i^\top \mathbf{b}_j - \mathbb{1}(\delta_j \neq 0) F_{i,\delta_j} \lambda_{j,\delta_j})^2 \right\}, \quad (4.6)$$

kde jsme zavedli označení $\mathbb{Y}_{\bullet j}$ pro j -tý sloupec matice \mathbb{Y} .

O koeficientech \mathbb{B} spojených se známými regresory \mathbb{X} budeme (podobně jako v předchozích modelech v kapitole 3) předpokládat, že jsou apriorně normálně rozdělené. Budeme zde uvažovat zobecněnou verzi konjugovaného rozdělení, tedy rozptylovou matici nezávislou na $\boldsymbol{\Sigma}$. Konkrétně budeme předpokládat, že

$$\mathbb{B} \sim \mathbf{N}_{m \times (1+q)}(\mathbb{B}_0, \mathbb{I}_m \otimes \mathbb{D}), \quad \text{tedy} \quad \mathbf{b}_j \sim \mathbf{N}_{1+q}(\mathbf{b}_{0,j}, \mathbb{D}), \quad \forall j \in \{1, \dots, m\}, \quad (4.7)$$

kde $\mathbb{B}_0 \in \mathbb{R}^{m \times (1+q)}$ a $0 < \mathbb{D} \in \mathbb{R}^{(1+q) \times (1+q)}$ jsou pevně stanovené hyperparametry. Řádky \mathbf{b}_j jsou tak apriorně mezi sebou nezávislé s různými středními hodnotami, ale stejnou rozptylovou maticí \mathbb{D} . Apriorní hustotu pak lze po vzoru hustot (1.5) a (1.2) zapsat ve tvaru

$$p(\mathbb{B}|\mathbb{B}_0, \mathbb{D}) \propto \exp \left\{ -\frac{1}{2} \text{Tr} (\mathbb{B} - \mathbb{B}_0) \mathbb{D}^{-1} (\mathbb{B} - \mathbb{B}_0)^\top \right\} \propto \prod_{j=1}^m \exp \left\{ -\frac{1}{2} (\mathbf{b}_j - \mathbf{b}_{0,j})^\top \mathbb{D}^{-1} (\mathbf{b}_j - \mathbf{b}_{0,j}) \right\}. \quad (4.8)$$

Dalším parametrem, jehož apriorní rozdělení bude velmi podobné tomu používanému v předchozí kapitole, je rozptylová matice Σ . Nyní však trváme na diagonalitě této matice, a proto nebudeme považovat Σ za reprezentanta z maticového inverzního Wishartova rozdělení, ale jednotlivé diagonální prvky budou pocházet z jednorozměrného Wishartova rozdělení (inverzní gama rozdělení v jiné parametrizaci), a to nezávisle na sobě. Tedy

$$\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_m^2), \quad \sigma_j^2 \sim \text{IW}_1(q_j, \nu_\sigma), \quad \forall j \in \{1, \dots, m\}, \quad (4.9)$$

kde $q_j > 0$ a $\nu_\sigma > 2$ jsou pevné hyperparametry. Frühwirth-Schnatter a Lopes (2012) doporučují použít $q_j = (\nu_\sigma - 4)/2 \left(\mathbb{S}_Y^{-1}\right)_{j,j}$, kde \mathbb{S}_Y je výběrová kovarianční matice vektoru \mathbf{Y} z dat \mathbb{Y} a kde $\nu_\sigma > 4$ není příliš nízké, např. $\nu_\sigma = 7$. Apriorní hustotu parametru Σ pak lze po vzoru hustoty (1.8) zapsat ve tvaru

$$p(\Sigma|q_1, \dots, q_m, \nu_\sigma) \propto \prod_{j=1}^m (\sigma_j^2)^{-\frac{\nu_\sigma}{2}} \exp\left\{-\frac{q_j}{2\sigma_j^2}\right\}. \quad (4.10)$$

O faktorových zátěžích Λ budeme opět předpokládat, že pochází z normálního rozdělení. Ovšem je třeba vynulovat případy, kdy $\Delta_{j,k} = 0$. Přesněji – podmíněné rozdělení $\lambda_{j,k}|\Delta_{j,k} = 0$ je degenerované rozdělení v nule, naopak $\lambda_{j,k}|\Delta_{j,k} = 1, \sigma_j^2$ bude normální rozdělení s rozptylem závislejícím na σ_j^2 pro konjugovanost. Konkrétně pro každé $j \in \{1, \dots, m\}$ a $k \in \{1, \dots, K^{\max}\}$ platí

$$\begin{aligned} \lambda_{j,k}|\Delta_{j,k} = 1, \sigma_j^2 &\sim \text{N}(\lambda_{0,j}, \sigma_j^2 a_j), \\ \lambda_{j,k}|\Delta_{j,k}, \sigma_j^2 &\sim \mathbb{1}(\Delta_{j,k} = 0)\text{D}_{\{0\}} + \mathbb{1}(\Delta_{j,k} = 1)\text{N}(\lambda_{0,j}, \sigma_j^2 a_j), \end{aligned} \quad (4.11)$$

kde $\lambda_{0,j} \in \mathbb{R}$ a $a_j > 0$ jsou hyperparametry. Všimněme si, že ať už je $\delta_j \neq 0$ jakékoliv, tak předpokládáme stejnou apriorní střední hodnotu $\lambda_{0,j}$ a stejnou konstantu a_j upravující rozptyl. Často se pro zjednodušení volí $a_j = a > 0$ shodné pro všechna j . Potom a má význam parametru ovlivňujícího naši obecnou míru informativnosti. Nemělo by být příliš nízké, ale ani volba $a^{-1} = 0$ vedoucí na neinformativní rozdělení není žádoucí, neboť značně zpomaluje konvergenci algoritmu. Hustotu podmíněného rozdělení $\Lambda|\Delta, \Sigma$ pak lze zapsat ve tvaru

$$\begin{aligned} p(\Lambda|\Delta, \Sigma) &\propto \prod_{j=1}^m \prod_{k=1}^{K^{\max}} \left[\mathbb{1}_{\{0\}}(\lambda_{j,k})\right]^{\mathbb{1}(\Delta_{j,k}=0)} \left[\left(\sigma_j^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{(\lambda_{j,k} - \lambda_{0,j})^2}{2a_j\sigma_j^2}\right\}\right]^{\mathbb{1}(\Delta_{j,k}=1)} \\ &\propto \prod_{j=1}^m \left[\left(\sigma_j^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{(\lambda_{j,\delta_j} - \lambda_{0,j})^2}{2a_j\sigma_j^2}\right\}\right]^{\mathbb{1}(\delta_j \neq 0)}. \end{aligned} \quad (4.12)$$

Rozdělení $\Lambda|\Sigma$ nepodmíněné Δ odpovídá směsi Diracova rozdělení v 0 a normálního rozdělení. Váhy těchto rozdělení pro jediný prvek $\lambda_{j,k}$ pak odpovídají pravděpodobnostem jevů $\Delta_{j,k} = 0$ a $\Delta_{j,k} = 1$, kterým se budeme věnovat nyní.

4.2.1 Apriorní rozdělení indikátorů

Na indikátorovou matici Δ budeme také nahlížet jako na náhodný element, což dělá náš model flexibilním. Rozdělení této matice budeme specifikovat po jednotlivých řádcích $\Delta_j, j \in \{1, \dots, m\}$, které mají ucelený význam. Tyto řádky

mohou nezávisle na sobě nabývat pouze nulového vektoru $e_0 = \mathbf{0}_{K^{\max}}$ nebo jednotkových vektorů e_k s 1 na k -té pozici, $k \in \{1, \dots, K^{\max}\}$. Označme si $\boldsymbol{\tau} = (\tau_0, \tau_1, \dots, \tau_{K^{\max}})^\top$ pravděpodobnosti těchto jevů, tj.

$$P(\Delta_j = e_k | \boldsymbol{\tau}) = \tau_k > 0, \quad j \in \{1, \dots, m\}, k \in \{1, \dots, K^{\max}\}, \sum_{k=0}^{K^{\max}} \tau_k = 1. \quad (4.13)$$

Hustotu celkového maticového parametru Δ (vůči čítací míře) získáme součinem těchto pravděpodobností, ovšem potřebujeme zajistit, aby výsledná matice Δ splňovala identifikační podmínky ve větě 16. Zavedme si množinu \mathcal{D} všech takových matic Δ , které tyto podmínky splňují, tj.

$$\mathcal{D} = \left\{ \begin{array}{l} \Delta \in \{0,1\}^{m \times K^{\max}} : \sum_{k=1}^{K^{\max}} \Delta_{j,k} \leq 1, \forall j \in \{1, \dots, m\}; \\ n_k(\Delta) = \sum_{j=1}^m \Delta_{j,k} \in \{0, 3, 4, \dots\}, \forall k \in \{1, \dots, K^{\max}\} \end{array} \right\}. \quad (4.14)$$

Musíme tedy zúžit apriorní rozdělení pouze na množinu \mathcal{D} . Jak toho docílit v praxi si ukážeme v samotné implementaci MCMC algoritmu v části 4.4.3. Celkově pak lze zapsat hustotu Δ ve tvaru

$$p(\Delta | \boldsymbol{\tau}) \propto \mathbb{1}_{\mathcal{D}}(\Delta) \cdot \prod_{k=0}^{K^{\max}} \tau_k^{n_k(\Delta)}, \quad (4.15)$$

kde $n_0(\Delta) = m - \sum_{k=1}^{K^{\max}} n_k(\Delta)$ je počet měření Y_j takových, které nejsou určeny žádným z faktorů F_k . Přesnou hodnotu multiplikační konstanty v proporcionalitě (4.15) pro naše účely nebude zapotřebí dopočítávat.

Není však prospěšné mít pravděpodobnosti $\boldsymbol{\tau}$ stanoveny jako konstanty, budeme je tedy považovat za náhodné parametry se svým apriorním rozdělením. To však zavedeme skrz parametr $\boldsymbol{\tau}^* = (\tau_0, \tau_1^*, \dots, \tau_{K^{\max}}^*)^\top$. Parametr τ_0 určuje pravděpodobnost, zda bude řádek Δ_j nulový. Jestliže nebude Δ_j nulové, tak potom s pravděpodobnostmi $\tau_k^* > 0$, $\sum_{k=1}^{K^{\max}} \tau_k^* = 1$, které si souhrnně označme $\boldsymbol{\tau}_{>0}^* = (\tau_1^*, \dots, \tau_{K^{\max}}^*)$, určíme, který z faktorů ovlivňuje Y_j . Budeme předpokládat, že nezávisle na sobě pochází τ_0 a $\boldsymbol{\tau}_{>0}^*$ z beta a Dirichletova rozdělení (viz část 1.5), konkrétně

$$\tau_0 \sim \text{Beta}(\alpha, \beta) \quad \text{a} \quad \boldsymbol{\tau}_{>0}^* \sim \text{Dir}_{K^{\max}}(\boldsymbol{\alpha}^*). \quad (4.16)$$

Hyperparametry $\alpha, \beta > 0$ a $\boldsymbol{\alpha}^*$ již volíme pevně. Nechceme-li zde vnášet nějakou informaci, tak použijeme $\alpha = \beta = 1$ a $\boldsymbol{\alpha}^* = \mathbf{1}_{K^{\max}}$ vedoucí na neinformativní rozdělení. Hodnoty nižší než 1 by odpovídaly situaci, kdy bychom chtěli preferovat pravděpodobnosti $\boldsymbol{\tau}^*$ blízké 0 či 1 (naopak pro hodnoty vyšší než 1). Apriorní hustota parametru $\boldsymbol{\tau}^*$ má tvar

$$p(\boldsymbol{\tau}^* | \alpha, \beta, \boldsymbol{\alpha}^*) \propto \tau_0^{\alpha-1} (1 - \tau_0)^{\beta-1} \cdot \prod_{k=1}^{K^{\max}} (\tau_k^*)^{\alpha_k^*-1}. \quad (4.17)$$

Použijme převodních vztahů mezi $\boldsymbol{\tau}$ a $\boldsymbol{\tau}^*$:

$$\boldsymbol{\tau} = \begin{pmatrix} \tau_0 \\ \boldsymbol{\tau}_{>0} \end{pmatrix} = \begin{pmatrix} \tau_0 \\ (1 - \tau_0) \boldsymbol{\tau}_{>0}^* \end{pmatrix} \quad \text{a} \quad \boldsymbol{\tau}^* = \begin{pmatrix} \tau_0 \\ \boldsymbol{\tau}_{>0}^* \end{pmatrix} = \begin{pmatrix} \tau_0 \\ \frac{\boldsymbol{\tau}_{>0}}{1 - \tau_0} \end{pmatrix},$$

abychom upravili hustotu (4.15) do tvaru

$$p(\mathbf{\Delta}|\boldsymbol{\tau}^*) \propto \mathbb{1}_{\mathcal{D}}(\mathbf{\Delta}) \cdot \tau_0^{n_0(\mathbf{\Delta})} (1 - \tau_0)^{m - n_0(\mathbf{\Delta})} \prod_{k=1}^{K^{\max}} (\tau_k^*)^{n_k(\mathbf{\Delta})}. \quad (4.18)$$

Spojením (4.17) a (4.18) tak dostáváme sdruženou hustotu dvojice $\mathbf{\Delta}$ a $\boldsymbol{\tau}^*$:

$$p(\mathbf{\Delta}, \boldsymbol{\tau}^* | \alpha, \beta, \boldsymbol{\alpha}^*) \propto \mathbb{1}_{\mathcal{D}}(\mathbf{\Delta}) \cdot \tau_0^{\alpha + n_0(\mathbf{\Delta}) - 1} (1 - \tau_0)^{\beta + m - n_0(\mathbf{\Delta}) - 1} \prod_{k=1}^{K^{\max}} (\tau_k^*)^{\alpha_k^* + n_k(\mathbf{\Delta}) - 1}. \quad (4.19)$$

Odtud je například vidět, že podmíněné rozdělení $\boldsymbol{\tau}^* | \mathbf{\Delta}$ lze opět vyjádřit jako kombinaci beta a Dirichletova rozdělení, konkrétně

$$\tau_0 | \mathbf{\Delta} \sim \text{Beta}(\alpha + n_0(\mathbf{\Delta}), \beta + m - n_0(\mathbf{\Delta})) \quad \text{a} \quad \boldsymbol{\tau}_{>0}^* | \mathbf{\Delta} \sim \text{Dir}_{K^{\max}}(\boldsymbol{\alpha}^* + \mathbf{n}(\mathbf{\Delta})), \quad (4.20)$$

kde jsme zavedli $\mathbf{n}(\mathbf{\Delta}) = (n_1(\mathbf{\Delta}), \dots, n_{K^{\max}}(\mathbf{\Delta}))^\top$.

Správné modelování indikátorové matice $\mathbf{\Delta}$ a pravděpodobností $\boldsymbol{\tau}^*$ je klíčové pro ideální funkčnost algoritmu. Conti a kol. (2014) provedli simulační studii, ve které se ukázalo, že pokud dokonce pro každé j -té měření uvažujeme jinou sadu pravděpodobností $\boldsymbol{\tau}_j^*$, tak algoritmus lépe rozeznává aktivní a neaktivní faktory. Dokonce to odbourává vliv informativnosti apriorního rozdělení $\boldsymbol{\tau}^*$.

4.2.2 Apriorní rozdělení faktorů

Jak jsme již avizovali v úvodu této části, budeme opět předpokládat, že skryté faktory \mathbf{F} jsou normálně rozdělené s nulovou střední hodnotou a s varianční maticí \mathbb{K} . Matice \mathbb{K} je zároveň korelační maticí pro účely identifikace, viz věta 16. Tu však budeme také považovat za náhodný parametr modelu (pro lepší adaptivitu algoritmu). Pro korelační matici neexistuje žádné pěkné rozdělení, které by šlo jednoduše propojit s apriorní hustotou faktorů

$$p(\mathbb{F} | \mathbf{\Omega}) \propto |\mathbf{\Omega}|^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \text{Tr} \mathbf{\Omega}^{-1} \mathbb{F}^\top \mathbb{F} \right\}, \quad (4.21)$$

kde v tomto případě $\mathbf{\Omega} = \mathbb{K}$.

Použijeme metody marginálního rozšíření dat (*marginal data augmentation*), která zde bude spočívat v dočasném převodu na model s obecnou rozptylovou maticí $\mathbf{\Omega}$ skrytých faktorů. Jakmile budeme mít nagenеровané $\mathbf{\Omega}$, již budeme moci přejít ke korelační matici \mathbb{K} . Toto korektně provedeme až při implementaci samotného algoritmu v části 4.4.1. Připravme si tedy pravděpodobnostní schéma, kterému bude parametr $\mathbf{\Omega}$ podléhat.

Pro $\mathbf{\Omega} = \mathbb{V}^{\frac{1}{2}} \mathbb{K} \mathbb{V}^{\frac{1}{2}}$ (viz rozklad (4.3)) obecnou pozitivně definitní maticí budeme předpokládat apriorní maticové inverzní Wishartovo rozdělení s parametry $\mathbb{S} = \text{diag}(S_1, \dots, S_{K^{\max}})$ a ν_Ω , tj. $\mathbf{\Omega} \sim \text{IW}_{K^{\max}}(\mathbb{S}, \nu_\Omega)$. To znamená, že apriorní hustota je

$$p(\mathbf{\Omega} | \mathbb{S}; \nu_\Omega) \propto |\mathbb{S}|^{\frac{\nu_\Omega - K^{\max} - 1}{2}} |\mathbf{\Omega}|^{-\frac{\nu_\Omega}{2}} \exp \left\{ -\frac{1}{2} \text{Tr} \mathbf{\Omega}^{-1} \mathbb{S} \right\}. \quad (4.22)$$

Pomocí věty o transformaci C9 s jakobiánem $|\mathbb{V}|^{\frac{K^{\max}-1}{2}}$ se snadno ukáže, že sdružená hustota \mathbb{V} a \mathbb{K} má tvar

$$\begin{aligned} p(\mathbb{V}, \mathbb{K} | \mathbb{S}; \nu_\Omega) &\propto |\mathbb{S}|^{\frac{\nu_\Omega - K^{\max} - 1}{2}} |\mathbb{V}|^{-\frac{\nu_\Omega - K^{\max} + 1}{2}} |\mathbb{K}|^{-\frac{\nu_\Omega}{2}} \exp \left\{ -\frac{1}{2} \text{Tr} \mathbb{V}^{-1} \mathbb{K}^{-1} \mathbb{S} \right\} \\ &\propto |\mathbb{K}|^{-\frac{\nu_\Omega}{2}} \prod_{k=1}^{K^{\max}} S_k^{\frac{\nu_\Omega - K^{\max} - 1}{2}} V_k^{-\frac{\nu_\Omega - K^{\max} + 1}{2}} \exp \left\{ -\frac{S_k (\mathbb{K}^{-1})_{k,k}}{2V_k} \right\}. \end{aligned} \quad (4.23)$$

Odtud vidíme, že podmíněná hustota k -tého diagonálního prvku $V_k = \omega_{k,k}$ matice \mathbb{V} při znalosti S_k i \mathbb{K} nabývá tvaru hustoty skalárního inverzního Wishartova rozdělení, konkrétně

$$V_k | \mathbb{K}, S_k \sim \text{IW}_1 \left(S_k (\mathbb{K}^{-1})_{k,k}, \nu_\Omega - K^{\max} + 1 \right), \quad \forall k \in \{1, \dots, K^{\max}\}. \quad (4.24)$$

Použitím tvrzení 3 jsme schopni vyintegrovat všechna V_k z hustoty (4.23) a určit tak hustotu \mathbb{K} nepodmíněnou znalostí \mathbb{V} . Ta má tvar

$$p(\mathbb{K}) \propto |\mathbb{K}|^{-\frac{\nu_\Omega}{2}} \prod_{k=1}^{K^{\max}} \left(\mathbb{K}^{-1} \right)_{k,k}^{-\frac{\nu_\Omega - K^{\max} - 1}{2}}, \quad (4.25)$$

kde konstanta v této proporcionalitě již *nezávisí* na \mathbb{S} . Ovšem již zde nepozorujeme žádné běžně používané rozdělení.

Parametr $\mathbb{S} = \text{diag}(S_1, \dots, S_{K^{\max}})$ můžeme volit jako fixní hyperparametr. Huang a Wand (2013) však navrhují tento parametr znáhodnit. O jednotlivých diagonálních prvcích S_k se v tomto případě předpokládá, že pochází z Wishartova rozdělení (gamma rozdělení v jiné parametrizaci), konkrétně

$$S_k \sim \text{W}_1 \left(\nu^* \xi_k^2, 1 \right) = \Gamma \left(\frac{1}{2}, \frac{1}{2\nu^* \xi_k^2} \right), \quad \forall k \in \{1, \dots, K^{\max}\}, \quad (4.26)$$

kde $\nu^* = (\nu_\Omega - K^{\max} - 1) - K^{\max} + 1 = \nu_\Omega - 2K^{\max}$ a $\xi_k^2 > 0$ jsou již pevné hyperparametry. Hustotu parametru \mathbb{S} pak lze po vzoru (1.6) psát v součinném tvaru

$$p(\mathbb{S}) = \prod_{k=1}^{K^{\max}} S_k^{-\frac{1}{2}} \exp \left\{ -\frac{S_k}{2\nu^* \xi_k^2} \right\}. \quad (4.27)$$

4.2.3 Shrnutí apriorního rozdělení

Ve výše popsaném modelu se vyskytuje velké množství parametrů. Proto zde uvádíme přehledné shrnutí.

Primárními parametry modelu jsou latentní faktory \mathbf{A} , regresní koeficienty \mathbb{B} , faktorové zátěže $\mathbf{\Lambda}$ a diagonální rozptylová matice chyb modelu $\mathbf{\Sigma}$. Tyto parametry přímo vystupují ve věrohodnosti (4.4). Sekundárními znáhodněnými parametry jsou indikátorová matice $\mathbf{\Delta}$ a rozptylová matice latentních faktorů $\mathbf{\Omega}$ (případně \mathbb{K}). Tyto parametry ovlivňují apriorní rozdělení primárních parametrů. Pro větší flexibilitu algoritmu zavádíme i terciární náhodné parametry $\boldsymbol{\tau}^*$ a \mathbb{S} , které ovlivňují rozdělení sekundárních parametrů. Pevnými (předem stanovenými) hyperparametry modelu jsou $b_{0,j}$, \mathbb{D} , q_j , ν_σ , $\lambda_{0,j}$, a_j , α , β , $\boldsymbol{\alpha}^*$, ν_Ω a ξ_k^2 , kde $j \in \{1, \dots, m\}$ a $k \in \{1, \dots, K^{\max}\}$. Nadále tyto hyperparametry nebudeme vypisovat do vzorců hustot pro zkrácení zápisu.

Celkové apriorní rozdělení všech znáhodněných parametrů modelu lze rozdělit na tři na sobě nezávislé části. První částí je apriorní rozdělení čtveřice Λ , Σ , Δ a τ^* , jejíž hustotu lze pomocí postupného podmiňování rozložit na následující součin hustot

$$p(\Lambda, \Sigma, \Delta, \tau^*) = \underbrace{p(\Lambda|\Sigma, \Delta)}_{(4.12)} \cdot \underbrace{p(\Sigma)}_{(4.10)} \cdot \underbrace{p(\Delta|\tau^*)}_{(4.15)} \cdot \underbrace{p(\tau^*)}_{(4.17)}.$$

Nacházíme-li se v (neidentifikovaném) rozšířeném modelu s obecnou rozptylovou maticí faktorů Ω , tak zde lze apriorní hustotu trojice \mathbb{F} , Ω a \mathbb{S} zapsat ve tvaru

$$p(\mathbb{F}, \Omega, \mathbb{S}) = \underbrace{p(\mathbb{F}|\Omega)}_{(4.21)} \cdot \underbrace{p(\Omega|\mathbb{S})}_{(4.22)} \cdot \underbrace{p(\mathbb{S})}_{(4.27)}.$$

Vyjádříme-li parametr Ω ekvivalentně pomocí dvojice \mathbb{K} (korelační matice) a \mathbb{V} (diagonální matice s jednotlivými rozptyly na diagonále), tak dostaneme apriorní hustotu

$$p(\mathbb{F}, \mathbb{V}, \mathbb{K}, \mathbb{S}) = \underbrace{p(\mathbb{F}|\Omega = \mathbb{V}^{\frac{1}{2}}\mathbb{K}\mathbb{V}^{\frac{1}{2}})}_{(4.21)} \cdot \underbrace{p(\mathbb{V}|\mathbb{K}, \mathbb{S})}_{(4.24)} \cdot \underbrace{p(\mathbb{K})}_{(4.25)} \cdot \underbrace{p(\mathbb{S})}_{(4.27)}.$$

Zbývá apriorní hustota (4.8) parametru \mathbb{B} , který není nijak spjatý s ostatními znáhodněnými parametry.

Dohromady má tedy apriorní hustota znáhodněných paramerů tvar

$$\begin{aligned} p(\mathbb{F}, \mathbb{B}, \Lambda, \Sigma, \Delta, \tau^*, \Omega, \mathbb{S}) &= \\ &= p(\mathbb{B}) \cdot p(\Lambda|\Delta, \Sigma) \cdot p(\Sigma) \cdot p(\Delta|\tau^*) \cdot p(\tau^*) \cdot p(\mathbb{F}|\Omega) \cdot p(\Omega|\mathbb{S}) \cdot p(\mathbb{S}). \end{aligned} \quad (4.28)$$

Názorněji je struktura našeho modelu vidět v diagramu 4.1. V kruzích jsou vyznačeny znáhodněné parametry a v obdélnících pevně stanovené hyperparametry. Dostupná data \mathbb{X} a \mathbb{Y} jsou zakreslena uvnitř kruhu i čtverce. Jsou zde zachyceny jednotlivé úrovně modelu (ovlivňování funguje směrem odshora dolů). Jsou zde vyznačeny i jednotlivá rozdělení, která byla předpokládána.

4.3 Aposteriorní rozdělení

Nyní dejme dohromady věrohodnost modelu (4.6) a hustotu (4.28) apriorního rozdělení znáhodněných parametrů definovaných v předchozí kapitole. Podle Bayesovy věty 2 je aposteriorní hustota znáhodněných parametrů modelu určujících faktorů proporcionální

$$\begin{aligned} p(\mathbb{F}, \mathbb{B}, \Lambda, \Sigma, \Delta, \tau^*, \Omega, \mathbb{S}|\mathbb{Y}, \mathbb{X}) &\propto p(\mathbb{Y}|\mathbb{X}, \mathbb{F}, \mathbb{B}, \Lambda, \Sigma) \cdot \\ &\cdot p(\mathbb{B}) \cdot p(\Lambda|\Delta, \Sigma) \cdot p(\Sigma) \cdot p(\Delta|\tau^*) \cdot p(\tau^*) \cdot p(\mathbb{F}|\Omega) \cdot p(\Omega|\mathbb{S}) \cdot p(\mathbb{S}). \end{aligned} \quad (4.29)$$

Je zřejmé, že ani v tomto složitějším modelu nebudeme schopni najít marginální aposteriorní rozdělení všech těchto parametrů. Ovšem díky vhodným volbám apriorních rozdělení jsme schopni určit plně podmíněná rozdělení a na jejich základě sestavit algoritmus pro MCMC odhad parametrů.

Algoritmus, který použijeme, však bude daleko sofistikovanější než Gibbsův algoritmus, který jsme používali doposud. Budeme pro něj potřebovat nejen plně podmíněná rozdělení, ale také částečně marginalizovaná aposteriorní rozdělení. Tímto zajistíme lepší vlastnosti tohoto algoritmu (menší autokorelovanost, rychlejší konvergence k aposteriornímu rozdělení).

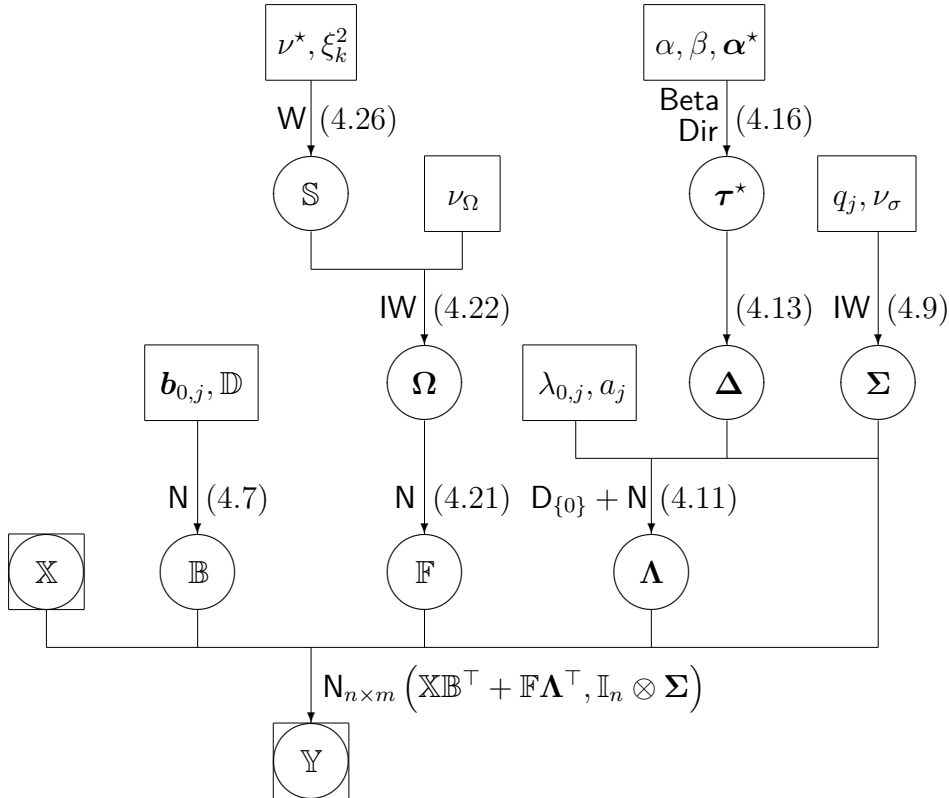
Následující tvrzení ukazuje, jak si poradit s parametry Λ , Σ a Δ zároveň.

Tvrzení 17. *V modelu určujících faktorů definovaném v části 4.1 s apriorním rozdělením zavedeným v části 4.2 lze hustotu plně podmíněného rozdělení trojice $(\Lambda, \Sigma, \Delta)$ rozložit na součin tří hustot:*

$$p(\Lambda, \Sigma, \Delta | \mathbb{Y}, \mathbb{X}, \mathbb{F}, \mathbb{B}, \boldsymbol{\tau}^*) \propto p(\Lambda | \mathbb{Y}, \mathbb{X}, \mathbb{F}, \mathbb{B}, \Sigma, \Delta) \cdot p(\Sigma | \mathbb{Y}, \mathbb{X}, \mathbb{F}, \mathbb{B}, \Delta) \cdot p(\Delta | \mathbb{Y}, \mathbb{X}, \mathbb{F}, \mathbb{B}, \boldsymbol{\tau}^*). \quad (4.30)$$

Podmíněná rozdělení náležící těmto hustotám jsou

$$\begin{aligned} \lambda_{j,k} | \mathbb{Y}_{\bullet,j}, \mathbb{X}, \mathbb{F}_{\bullet,k}, \mathbf{b}_j, \sigma_j^2, \Delta_{j,k} &\sim \mathbf{1}(\Delta_{j,k} = 0) \mathbf{D}_{\{0\}} + \mathbf{1}(\Delta_{j,k} = 1) \mathbf{N}(\hat{\lambda}_{j,k}, r_{j,k} \sigma_j^2), \\ \sigma_j^2 | \mathbb{Y}_{\bullet,j}, \mathbb{X}, \mathbb{F}, \mathbf{b}_j, \Delta_j &\sim \text{IW}_1(\Psi_{j,\delta_j}, n + \nu_\sigma), \\ P(\Delta_j = e_k | \mathbb{Y}_{\bullet,j}, \mathbb{X}, \mathbb{F}, \mathbf{b}_j, \boldsymbol{\tau}^*) &= \left[\sum_{l=0}^{K^{max}} \exp(O_j^{(k \rightarrow l)}) \right]^{-1}, \end{aligned}$$



Obrázek 4.1: Diagram popisující strukturu modelu určujících faktorů.

kde pro $j \in \{1, \dots, m\}$ a $k \in \{1, \dots, K^{\max}\}$ jsou

$$\begin{aligned}
r_{j,k} &= \begin{cases} 1, & k = 0, \\ (\mathbb{F}_{\bullet,k}^\top \mathbb{F}_{\bullet,k} + a_j^{-1})^{-1}, & k \in \{1, \dots, K^{\max}\}, \end{cases} \\
\widehat{\lambda}_{j,k} &= r_{j,k} \left[(\mathbb{Y}_{\bullet,j} - \mathbb{X}\mathbf{b}_j)^\top \mathbb{F}_{\bullet,k} + \frac{\lambda_{0,j}}{a_j} \right], \\
\Psi_{j,k} &= q_j + (\mathbb{Y}_{\bullet,j} - \mathbb{X}\mathbf{b}_j)^\top (\mathbb{Y}_{\bullet,j} - \mathbb{X}\mathbf{b}_j) + \mathbf{1}(k \neq 0) \left(\lambda_{0,j}^2 a_j^{-1} - (\widehat{\lambda}_{j,k})^2 r_{j,k}^{-1} \right), \\
O_j^{(k \rightarrow l)} &= \begin{cases} 0, & k = l, \\ \frac{1}{2} \log \left(\frac{r_{j,l}}{a_j} \right) - \frac{n + \nu_\sigma - 2}{2} \log \left(\frac{\Psi_{j,l}}{\Psi_{j,0}} \right) + \log \left(\frac{1 - \tau_0}{\tau_0} \tau_l^* \right), & k = 0, l \neq 0, \\ \frac{1}{2} \log \left(\frac{r_{j,l}}{r_{j,k}} \right) - \frac{n + \nu_\sigma - 2}{2} \log \left(\frac{\Psi_{j,l}}{\Psi_{j,k}} \right) + \log \left(\frac{\tau_l^*}{\tau_k^*} \right), & k \neq 0, l \neq 0, \\ -O_j^{(l \rightarrow k)}. \end{cases}
\end{aligned}$$

Důkaz. Rozložení hustoty (4.30) na součin tří jiných hustot je zřejmé ze známé vlastnosti vyjádření hustoty více parametrů pomocí postupného podmiňování. Dále si odvodíme, o jaká se jedná podmíněná rozdělení, na něž tuto hustotu rozkládáme. Hledaná rozdělení nezávisí na $\boldsymbol{\Omega}$ a \mathbb{S} (také náhodné parametry modelu), neboť ty z hierarchie našeho modelu ovlivňují pouze \mathbb{F} , kterým ve všech případech podmiňujeme.

Podmíněné rozdělení $\boldsymbol{\Lambda}$ uvažované v tomto tvrzení je vlastně klasické plně podmíněné rozdělení, jak bylo zadefinováno v části 2.3.2, neboť podmiňujeme znalostí všech ostatních parametrů modelu. Podle úvah v této části lze plně podmíněnou hustotu získat vyjádřením hustoty

$$\begin{aligned}
p(\boldsymbol{\Lambda} | \mathbb{Y}, \mathbb{X}, \mathbb{F}, \mathbb{B}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}) &\propto p(\mathbb{Y} | \mathbb{X}, \mathbb{F}, \mathbb{B}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma}) \cdot p(\boldsymbol{\Lambda} | \boldsymbol{\Sigma}, \boldsymbol{\Delta}) \\
&\propto \prod_{j=1}^m \prod_{k=1}^{K^{\max}} [\mathbf{1}_{\{0\}}(\lambda_{j,k})]^{\mathbf{1}(\delta_j=0)} \cdot \prod_{j=1}^m (\sigma_j^2)^{-\frac{n}{2}} (a_j \sigma_j^2)^{-\frac{1}{2} \mathbf{1}(\delta_j \neq 0)} \\
&\cdot \exp \left\{ - \frac{(\mathbb{Y}_{\bullet,j} - \mathbb{X}\mathbf{b}_j - \mathbb{F}\boldsymbol{\lambda}_j)^\top (\mathbb{Y}_{\bullet,j} - \mathbb{X}\mathbf{b}_j - \mathbb{F}\boldsymbol{\lambda}_j) + \mathbf{1}(\delta_j \neq 0) \frac{(\lambda_{j,\delta_j} - \lambda_{0,j})^2}{a_j}}{2\sigma_j^2} \right\}
\end{aligned}$$

jako funkce v proměnné $\boldsymbol{\Lambda}$. Rozložme to na jednotlivé prvky $\lambda_{j,k}$ při znalosti σ_j^2 , $\Delta_{j,k}$ a ostatních. Za znalosti $\Delta_{j,k} = 0$ víme, že musí být $\lambda_{j,k} = 0$. Pokud ovšem $\Delta_{j,k} = 1$, tj. $0 \neq k = \delta_j$, tak zbývá výraz

$$\exp \left\{ - \frac{1}{2\sigma_j^2} \left[(\mathbb{Y}_{\bullet,j} - \mathbb{X}\mathbf{b}_j - \mathbb{F}_{\bullet,\delta_j} \lambda_{j,\delta_j})^\top (\mathbb{Y}_{\bullet,j} - \mathbb{X}\mathbf{b}_j - \mathbb{F}_{\bullet,\delta_j} \lambda_{j,\delta_j}) + \frac{(\lambda_{j,\delta_j} - \lambda_{0,j})^2}{a_j} \right] \right\}$$

vyjádřit jako exponencializovanou kvadratickou formu v λ_{j,δ_j} . Za tímto účelem použijeme tvrzení B7 s hlavní proměnnou $\mathbb{X}^* = \lambda_{j,\delta_j}$. Dostáváme tak rozklad

$$\exp \left\{ - \frac{1}{2\sigma_j^2} \left[(\lambda_{j,\delta_j} - \widehat{\lambda}_{j,\delta_j})^2 (\mathbb{F}_{\bullet,\delta_j}^\top \mathbb{F}_{\bullet,\delta_j} + a_j^{-1}) + (\Psi_{j,\delta_j} - q_j) \right] \right\}, \quad (4.31)$$

kde Ψ_{j,δ_j} a $\widehat{\lambda}_{j,\delta_j}$ jsou definovány jako ve znění tvrzení. Z tohoto rozpisu je již zřejmé, že plně podmíněné rozdělení $\lambda_{j,k}$ při $\Delta_{j,k} = 1$ musí být normální rozdělení se střední hodnotou $\widehat{\lambda}_{j,k}$ a rozptylem $r_{j,k}\sigma_j^2$.

Dále potřebujeme získat hustotu rozdělení $\Sigma|\mathbb{Y}, \mathbb{X}, \mathbb{F}, \mathbb{B}, \Delta$. Všimněme si, že tato hustota nezávisí na Λ , tedy se nejedná o plně podmíněné rozdělení (to by na Λ jistě záviselo). Musíme tedy uvažovat hustotu sdruženého rozdělení $\Lambda, \Sigma|\mathbb{Y}, \mathbb{X}, \mathbb{F}, \mathbb{B}, \Delta$, ze které je třeba vyintegrovat parametr Λ . Tato úloha se zjednoduší pomocí Bayesovy věty 2 následovně:

$$\begin{aligned} p(\Sigma|\mathbb{Y}, \mathbb{X}, \mathbb{F}, \mathbb{B}, \Delta) &= \int p(\Lambda, \Sigma|\mathbb{Y}, \mathbb{X}, \mathbb{F}, \mathbb{B}, \Delta) d\Lambda \\ &\propto p(\Sigma) \cdot \int p(\mathbb{Y}|\mathbb{X}, \mathbb{F}, \mathbb{B}, \Lambda, \Sigma) \cdot p(\Lambda|\Sigma, \Delta) d\Lambda. \end{aligned}$$

Již jsme si odvodili, že $p(\mathbb{Y}|\mathbb{X}, \mathbb{F}, \mathbb{B}, \Lambda, \Sigma) \cdot p(\Lambda|\Sigma, \Delta)$ je v Λ až na konstantu hustotou směsi Diracova a normálního rozdělení. Lze tedy jednoduše vyintegrovat Λ (stále podmiňujeme Δ , takže víme, zda je $\lambda_{j,k}$ nulové nebo zda pochází z normálního rozdělení).

Za tímto účelem použijeme tvrzení 3, kde pro j takové, že $\delta_j \neq 0$, je

$$\int_{\mathbb{R}} \left(2\pi r_{j,\delta_j} \sigma_j^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{\left(\lambda_{j,\delta_j} - \widehat{\lambda}_{j,\delta_j}\right)^2}{2r_{j,\delta_j} \sigma_j^2}\right\} d\lambda_{j,\delta_j} = 1.$$

Dejme proto dohromady apriorní hustotu Σ (rovnice (4.10)), nevyintegrovaný zbytek z rozkladu (4.31) a konstantu r_{j,δ_j} , která byla zapotřebí při integraci. Dostaneme, že

$$p(\sigma_j^2|\mathbb{Y}_{\bullet j}, \mathbb{X}, \mathbb{F}, \mathbf{b}_j, \Delta_j) \propto \left(\frac{r_{j,\delta_j}}{a_j}\right)^{\frac{1}{2}\mathbb{1}(\delta_j \neq 0)} \left(\sigma_j^2\right)^{-\frac{n+\nu_\sigma}{2}} \exp\left\{-\frac{1}{2\sigma_j^2} [q_j + \Psi_{j,\delta_j} - q_j]\right\}, \quad (4.32)$$

$$\propto \left(\sigma_j^2\right)^{-\frac{n+\nu_\sigma}{2}} \exp\left\{-\frac{\Psi_{j,\delta_j}}{2\sigma_j^2}\right\}. \quad (4.33)$$

Odtud již vidíme, že pro každé j (i ta, pro která $\delta_j = 0$) je podmíněné rozdělení σ_j^2 při znalosti ostatních parametrů až na Λ opět inverzní Wishartovo s parametry Ψ_{j,δ_j} a $n + \nu_\sigma$.

Zbývá určit rozdělení $\Delta|\mathbb{Y}, \mathbb{X}, \mathbb{F}, \mathbb{B}, \tau^*$, tedy rozdělení nepodmíněné Λ a Σ . Podobným způsobem jako výše u Σ lze odvodit, že

$$\begin{aligned} p(\Delta|\mathbb{Y}, \mathbb{X}, \mathbb{F}, \mathbb{B}, \tau^*) &= \int p(\Delta, \Lambda, \Sigma|\mathbb{Y}, \mathbb{X}, \mathbb{F}, \mathbb{B}, \tau^*) d(\Lambda, \Sigma) \\ &\propto p(\Delta|\tau^*) \underbrace{\int \underbrace{p(\mathbb{Y}|\mathbb{X}, \mathbb{F}, \mathbb{B}, \Lambda, \Sigma) \cdot p(\Lambda|\Sigma, \Delta) \cdot p(\Sigma)}_{p(\mathbb{Y}, \Lambda, \Sigma|\mathbb{X}, \mathbb{F}, \mathbb{B}, \Delta)} d(\Lambda, \Sigma)}_{p(\mathbb{Y}|\mathbb{X}, \mathbb{F}, \mathbb{B}, \Delta)} \\ &\propto p(\Delta|\tau^*) \int p(\Sigma) \int p(\mathbb{Y}|\mathbb{X}, \mathbb{F}, \mathbb{B}, \Lambda, \Sigma) \cdot p(\Lambda|\Sigma, \Delta) d\Lambda d\Sigma. \end{aligned}$$

Integrovaní podle Λ jsme již provedli při odvozování podmíněného rozdělení Σ . Tentokrát použijeme tvrzení 3 za účelem integrace všech σ_j^2 , kde nyní musíme

doplnit konstantu $\Gamma(\frac{n+\nu_\sigma}{2} - 1)(\Psi_{j,\delta_j}/2)^{-\frac{n+\nu_\sigma-2}{2}}$, aby zde vznikla hustota, která se vyintegruje na 1. Tuto konstantu přidejme k nevyintegrovanému zbytku, který můžeme vidět ve výrazu (4.32). Mějme však na paměti, že důležité jsou pouze ty členy, které závisí na Δ .

Dostáváme tak, že

$$p(\mathbb{Y}|\mathbb{X}, \mathbb{F}, \mathbb{B}, \Delta) \propto \prod_{j=1}^m \left(\frac{r_{j,\delta_j}}{a_j} \right)^{\frac{1}{2}\mathbb{1}(\delta_j \neq 0)} (\Psi_{j,\delta_j})^{-\frac{n+\nu_\sigma-2}{2}}, \quad (4.34)$$

$$p(\mathbb{Y}_{\bullet j}|\mathbb{X}, \mathbb{F}, \mathbf{b}_j, \Delta_j) \propto \left(\frac{r_{j,\delta_j}}{a_j} \right)^{\frac{1}{2}\mathbb{1}(\delta_j \neq 0)} (\Psi_{j,\delta_j})^{-\frac{n+\nu_\sigma-2}{2}}, \quad \forall j \in \{1, \dots, m\}.$$

Odtud již dostáváme, že částečně marginalizovaná aposteriorní pravděpodobnost toho, že Δ_j nabyde hodnoty e_k pro nějaké $k \in \{0, \dots, K^{\max}\}$ (ekvivalentně, že $\delta_j = k$), je proporcionální výrazu

$$\begin{aligned} P(\Delta_j = e_k | \mathbb{Y}_{\bullet j}, \mathbb{X}, \mathbb{F}, \mathbf{b}_j, \boldsymbol{\tau}^*) &\propto \\ &\propto \left(\frac{r_{j,k}}{a_j} \right)^{\frac{1}{2}\mathbb{1}(k \neq 0)} (\Psi_{j,k})^{-\frac{n+\nu_\sigma-2}{2}} (\tau_0)^{\mathbb{1}(k=0)} ((1-\tau_0)\tau_k^*)^{\mathbb{1}(k \neq 0)}. \end{aligned} \quad (4.35)$$

Jelikož součet pravděpodobností musí dát součtem opět 1, tak dostáváme, že

$$\begin{aligned} P(\Delta_j = e_k | \mathbb{Y}_{\bullet j}, \mathbb{X}, \mathbb{F}, \mathbf{b}_j, \boldsymbol{\tau}^*) &= \\ &= \frac{\left(\frac{r_{j,k}}{a_j} \right)^{\frac{1}{2}\mathbb{1}(k \neq 0)} (\Psi_{j,k})^{-\frac{n+\nu_\sigma-2}{2}} (\tau_0)^{\mathbb{1}(k=0)} ((1-\tau_0)\tau_k^*)^{\mathbb{1}(k \neq 0)}}{\sum_{l=0}^{K^{\max}} \left(\frac{r_{j,l}}{a_j} \right)^{\frac{1}{2}\mathbb{1}(l \neq 0)} (\Psi_{j,l})^{-\frac{n+\nu_\sigma-2}{2}} (\tau_0)^{\mathbb{1}(l=0)} ((1-\tau_0)\tau_l^*)^{\mathbb{1}(l \neq 0)}} \\ &= \left[\sum_{l=0}^{K^{\max}} \sqrt{\frac{r_{j,l}}{r_{j,k}}} \left(\frac{\Psi_{j,l}}{\Psi_{j,k}} \right)^{-\frac{n+\nu_\sigma-2}{2}} \left(\frac{\tau_0 \sqrt{a_j}}{(1-\tau_0)} \right)^{\mathbb{1}(l=0) - \mathbb{1}(k=0)} \frac{(\tau_l^*)^{\mathbb{1}(l \neq 0)}}{(\tau_k^*)^{\mathbb{1}(k \neq 0)}} \right]^{-1}. \end{aligned} \quad (4.36)$$

Nyní již jen stačí vzít logaritmus sčítance této sumy a dosadit pro různé volby k a l , abychom získali logaritmované poměry $O_j^{(k \rightarrow l)}$ ve znění tvrzení. \square

Převod na logaritmus se zde provádí z výpočetních důvodů. Eliminuje se tím tak numerická chyba, která by mohla vzniknout dělením číslem velmi blízkým nule. Zatím jsme však neřešili, jak zajistit, aby Δ splňovalo podmínku věty 15, tj. $\Delta \in \mathcal{D}$. S tímto problémem se vypořádáme později v samotném MCMC algoritmu (část 4.4.3).

S touto trojicí parametrů $(\Lambda, \Sigma, \Delta)$ je ještě spojen terciární parametr $\boldsymbol{\tau}^*$, který ovlivňuje pouze Δ . Kromě svého apriorního rozdělení a rozdělení $\Delta | \boldsymbol{\tau}^*$ již nikde jinde nevystupuje. Proto plně podmíněné rozdělení $\boldsymbol{\tau}^*$ vlastně odpovídá podmíněnému rozdělení $\boldsymbol{\tau}^* | \Delta$, které jsme si již jednoduše odvodili v části 4.2.1, viz rovnice (4.20).

Následující tvrzení udává plně podmíněná rozdělení druhé skupiny parametrů, která se pojí s latentními faktory \mathbb{F} .

Tvrzení 18. V modelu určujících faktorů definovaném v části 4.1 s apriorním rozdělením zavedeným v části 4.2 jsou plně podmíněná rozdělení parametrů \mathbb{F} , $\mathbf{\Omega}$ a \mathbb{S} následující:

$$\begin{aligned}\mathbb{F}|\mathbb{Y}, \mathbb{X}, \mathbb{B}, \mathbf{\Lambda}, \mathbf{\Sigma}, \mathbf{\Omega} &\sim \mathbf{N}_{n \times K^{\max}} \left(\hat{\mathbb{F}}, \mathbb{I}_n \otimes [\mathbf{\Lambda}^\top \mathbf{\Sigma}^{-1} \mathbf{\Lambda} + \mathbf{\Omega}^{-1}]^{-1} \right), \\ \mathbf{\Omega}|\mathbb{F}, \mathbb{S} &\sim \text{IW}_{K^{\max}} \left(\mathbb{F}^\top \mathbb{F} + \mathbb{S}, n + \nu_\Omega \right), \\ S_k|\mathbf{\Omega} &\sim \text{W}_1 \left(\left(\frac{1}{\nu^* \xi_k^2} + (\mathbf{\Omega}^{-1})_{k,k} \right)^{-1}, \nu_\Omega - K^{\max} \right), \quad k \in \{1, \dots, K^{\max}\}, \\ \text{kde } \hat{\mathbb{F}} &= (\mathbb{Y} - \mathbb{X}\mathbb{B}^\top) \mathbf{\Sigma}^{-1} \mathbf{\Lambda} [\mathbf{\Lambda}^\top \mathbf{\Sigma}^{-1} \mathbf{\Lambda} + \mathbf{\Omega}^{-1}]^{-1}.\end{aligned}$$

Důkaz. Plně podmíněné rozdělení \mathbb{F} lze odvodit analogicky jako v části 3.6.1 s $\mathbb{F}_0 = \mathbb{O}_{n \times K^{\max}}$. Výsledek je vlastně obdobou plně podmíněného rozdělení v tvrzení 14. Důležité je uvědomit si, že podmiňujeme znalostí parametru $\mathbf{\Lambda}$, v němž je již promítnuté $\mathbf{\Delta}$, které jinak přímo ve vztahu s \mathbb{F} není. Proto již $\mathbf{\Delta}$, natož $\boldsymbol{\tau}^*$ v plně podmíněném rozdělení nefigurují.

Parametr $\mathbf{\Omega}$ ve věrohodnosti (4.4) nijak nevystupuje, proto při určování plně podmíněného rozdělení parametru $\mathbf{\Omega}$ nebudou data \mathbb{Y} , \mathbb{X} a parametry \mathbb{B} , $\mathbf{\Lambda}$, $\mathbf{\Sigma}$, $\mathbf{\Delta}$ a $\boldsymbol{\tau}^*$ hrát žádnou roli. Plně podmíněné rozdělení $\mathbf{\Omega}$ závisející pouze na \mathbb{F} a \mathbb{S} se tak získá pouze spojením apriorní hustoty (4.22) a hustoty rozdělení $\mathbb{F}|\mathbf{\Omega}$ (4.21). Podle Bayesovy věty 2 tak

$$p(\mathbf{\Omega}|\mathbb{F}, \mathbb{S}) \propto p(\mathbb{F}|\mathbf{\Omega})p(\mathbf{\Omega}|\mathbb{S}) \propto |\mathbf{\Omega}|^{-\frac{n+\nu_\Omega}{2}} \cdot \exp \left\{ -\frac{1}{2} \text{Tr} \mathbf{\Omega}^{-1} (\mathbb{F}^\top \mathbb{F} + \mathbb{S}) \right\}, \quad (4.37)$$

a tedy podmíněné rozdělení $\mathbf{\Omega}|\mathbb{F}, \mathbb{S}$ je maticové inverzní Wishartovo s parametry $\mathbb{F}^\top \mathbb{F} + \mathbb{S}$ a $n + \nu_\Omega$.

Podobně jako u předchozího $\mathbf{\Omega}$ můžeme i u \mathbb{S} vyloučit data \mathbb{Y} , \mathbb{X} a parametry \mathbb{B} , $\mathbf{\Lambda}$, $\mathbf{\Sigma}$, $\mathbf{\Delta}$, $\boldsymbol{\tau}^*$, a dokonce i faktory \mathbb{F} z plně podmíněného rozdělení. Dejme proto dohromady apriorní hustotu (4.27) parametru \mathbb{S} s informací o \mathbb{S} obsažené v hustotě rozdělení $\mathbf{\Omega}|\mathbb{S}$ v rovnici (4.22). To jsou totiž jediné hustoty ze specifikovaného modelu, ve kterých se \mathbb{S} vyskytuje. Tím z Bayesovy věty 2 získáme hustotu podmíněného rozdělení $\mathbb{S}|\mathbf{\Omega}$

$$p(\mathbb{S}|\mathbf{\Omega}) \propto p(\mathbf{\Omega}|\mathbb{S})p(\mathbb{S}) \propto \prod_{k=1}^{K^{\max}} S_k^{\frac{\nu_\Omega - K^{\max} - 1}{2}} \exp \left\{ -\frac{S_k}{2} \left(\frac{1}{\nu^* \xi_k^2} + (\mathbf{\Omega}^{-1})_{k,k} \right) \right\}, \quad (4.38)$$

které je zároveň plně podmíněným rozdělením parametru \mathbb{S} . Pro jednotlivé diagonální prvky S_k pozorujeme hustotu Wishartova rozdělení (1.6) s parametry uvedenými ve znění tvrzení. □

Zbývá již jen určit plně podmíněné rozdělení parametru \mathbb{B} , který není apriorně nijak propojen s žádným ze znárodných parametrů. Toto rozdělení určíme pro jednotlivé řádky \mathbf{b}_j této matice stejně, jako jsme specifikovali apriorní rozdělení (4.7).

Tvrzení 19. V modelu určujících faktorů definovaném v části 4.1 s apriorním rozdělením zavedeným v části 4.2 je plně podmíněné rozdělení parametru \mathbb{B} následující:

$$\mathbf{b}_j | \mathbb{Y}_{\bullet j}, \mathbb{X}, \mathbb{F}, \boldsymbol{\lambda}_j, \sigma_j^2 \sim \mathbf{N}_{1+q} \left(\hat{\mathbf{b}}_j, \left[\sigma_j^{-2} \mathbb{X}^\top \mathbb{X} + \mathbb{D}^{-1} \right]^{-1} \right), \quad \forall j \in \{1, \dots, m\},$$

$$\text{kde } \hat{\mathbf{b}}_j = \left[\sigma_j^{-2} \mathbb{X}^\top \mathbb{X} + \mathbb{D}^{-1} \right]^{-1} \left(\sigma_j^{-2} \mathbb{X}^\top (\mathbb{Y}_{\bullet j} - \mathbb{F} \boldsymbol{\lambda}_j) + \mathbb{D}^{-1} \mathbf{b}_{0,j} \right).$$

Důkaz. Pro plně podmíněnou hustotu parametru \mathbb{B} budeme potřebovat dát dohromady věrohodnost (4.5) a apriorní hustotu (4.8). Zřejmě

$$p(\mathbb{B} | \mathbb{Y}, \mathbb{X}, \mathbb{F}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma}) \propto \prod_{j=1}^m \exp \left\{ -\frac{1}{2} \left[(\mathbf{b}_j - \mathbf{b}_{0,j})^\top \mathbb{D}^{-1} (\mathbf{b}_j - \mathbf{b}_{0,j}) + \sigma_j^{-2} (\mathbb{Y}_{\bullet j} - \mathbb{X} \mathbf{b}_j - \mathbb{F} \boldsymbol{\lambda}_j)^\top (\mathbb{Y}_{\bullet j} - \mathbb{X} \mathbf{b}_j - \mathbb{F} \boldsymbol{\lambda}_j) \right] \right\}.$$

Zaměříme se dále na jednotlivé řádky \mathbf{b}_j matice \mathbb{B} , které budou zjevně nezávislé i v aposteriorním rozdělení. Pomocí tvrzení B7 s hlavní proměnnou $\mathbb{X}^* = \mathbf{b}_j^\top$ doplníme následující výraz na čtverec:

$$\begin{aligned} & \sigma_j^{-2} \left(\mathbf{b}_j^\top \mathbb{X}^\top - (\mathbb{Y}_{\bullet j}^\top - \boldsymbol{\lambda}_j^\top \mathbb{F}^\top) \right) \left(\mathbf{b}_j^\top \mathbb{X}^\top - (\mathbb{Y}_{\bullet j}^\top - \boldsymbol{\lambda}_j^\top \mathbb{F}^\top) \right)^\top + \\ & + \left(\mathbf{b}_j^\top - \mathbf{b}_{0,j}^\top \right) \mathbb{D}^{-1} \left(\mathbf{b}_j^\top - \mathbf{b}_{0,j}^\top \right)^\top = \left(\mathbf{b}_j - \hat{\mathbf{b}}_j \right)^\top \left[\sigma_j^{-2} \mathbb{X}^\top \mathbb{X} + \mathbb{D}^{-1} \right] \left(\mathbf{b}_j - \hat{\mathbf{b}}_j \right) - \\ & - \hat{\mathbf{b}}_j^\top \left[\sigma_j^{-2} \mathbb{X}^\top \mathbb{X} + \mathbb{D}^{-1} \right] \hat{\mathbf{b}}_j + \sigma_j^{-2} (\mathbb{Y}_{\bullet j} - \mathbb{F} \boldsymbol{\lambda}_j)^\top (\mathbb{Y}_{\bullet j} - \mathbb{F} \boldsymbol{\lambda}_j) + \mathbf{b}_{0,j}^\top \mathbb{D}^{-1} \mathbf{b}_{0,j}, \end{aligned}$$

kde $\hat{\mathbf{b}}_j$ má tvar, jenž je uveden ve znění tvrzení. Na \mathbf{b}_j již pak závisí pouze první člen tohoto rozkladu, a tedy

$$p(\mathbf{b}_j | \mathbb{Y}_{\bullet j}, \mathbb{X}, \mathbb{F}, \boldsymbol{\lambda}_j, \sigma_j^2) \propto \exp \left\{ -\frac{1}{2} \left(\mathbf{b}_j - \hat{\mathbf{b}}_j \right)^\top \left[\sigma_j^{-2} \mathbb{X}^\top \mathbb{X} + \mathbb{D}^{-1} \right] \left(\mathbf{b}_j - \hat{\mathbf{b}}_j \right) \right\}, \quad (4.39)$$

což odpovídá hustotě (1.2) vektorového normálního rozdělení. □

4.4 MCMC algoritmus

V předchozí části jsme si připravili vše potřebné pro to, abychom mohli definovat algoritmus, pomocí kterého budeme vhodným způsobem generovat z odvozených rozdělení. Tento algoritmus byl prezentován v článku Conti a kol. (2014) v obecnější verzi. Zde si tento algoritmus na několika místech zjednodušíme, ovšem zápis jednotlivých kroků provedeme velmi konkrétně a přehledně.

Základem bude opět Gibbsův algoritmus, který však bude zapotřebí rozšířit o metodu marginálního doplňování dat. Ani tato procedura však nezajišťuje, že nagenované hodnoty v i -tém kroku budou splňovat předpoklady modelu určujících faktorů. Bude proto nutné rozšířit algoritmus o takzvaný Metropolisův-Hastingsův krok, který zajistí splnění podmínek věty 15. Celý algoritmus tak bude rozdělen na tři části.

4.4.1 Metoda marginálního doplnění dat

Nejprve se vypořádáme s generováním skrytých faktorů \mathbb{F} a rozptylové (zároveň korelační) matice \mathbb{K} z jejich společného plně podmíněného rozdělení $\mathbb{F}, \mathbb{K} | \mathbb{Y}, \mathbb{X}, \mathbb{B}, \mathbf{\Lambda}, \mathbf{\Sigma}$. Připomínáme, že pro identifikaci našeho modelu vyžadujeme, aby rozptylová matice faktorů měla na diagonále samé jedničky (věta 15), tedy, aby to byla právě korelační matice \mathbb{K} a nikoliv obecná pozitivně definitní matice $\mathbf{\Omega}$. Generování přímo korelační matice je však velmi náročné. Převědeme tak dočasně model s korelační maticí \mathbb{K} na model s obecnou varianční maticí $\mathbf{\Omega}$ pomocí rozptylů $\mathbb{V} = \text{diag}(V_1, \dots, V_{K^{\max}})$. Po nagerování nového $\mathbf{\Omega}$ model převedeme zpět na model s varianční maticí totožnou s korelační maticí \mathbb{K} .

Předpokládejme, že máme k dispozici data \mathbb{Y} a \mathbb{X} a k nim poslední nagerované hodnoty $\mathbb{B}^i, \mathbf{\Lambda}^i, \mathbf{\Sigma}^i$ (tyto budou pocházet ze současného i -tého kroku, viz algoritmus 6) a také korelační maticí \mathbb{K}^{i-1} z minulého kroku. Následující algoritmus nageruje nové parametry \mathbb{F}^i a \mathbb{K}^i z rozdělení $\mathbb{F}, \mathbb{K} | \mathbb{Y}, \mathbb{X}, \mathbb{B}, \mathbf{\Lambda}, \mathbf{\Sigma}$. V rámci procedury také dochází k přeškálování matice faktorových zátěží $\mathbf{\Lambda}^i$.

Algoritmus 5. *Generování \mathbb{F} a \mathbb{K} pomocí marginálního rozšíření dat.*

Vstupy: data \mathbb{Y} a \mathbb{X} , poslední nagerované hodnoty $\mathbb{B}^i, \mathbf{\Lambda}^i, \mathbf{\Sigma}^i$ a \mathbb{K}^{i-1} .

Výstup: nové \mathbb{F}^i a \mathbb{K}^i z rozdělení $\mathbb{F}, \mathbb{K} | \mathbb{Y}, \mathbb{X}, \mathbb{B}, \mathbf{\Lambda}, \mathbf{\Sigma}$, dále upravené $\mathbf{\Lambda}^i$.

Postup: V následujícím pořadí proveď tyto kroky:

- a) *Generuj \mathbb{F}^i z plně podmíněného rozdělení $\mathbb{F} | \mathbb{Y}, \mathbb{X}, \mathbb{B}^i, \mathbf{\Lambda}^i, \mathbf{\Sigma}^i, \mathbf{\Omega} = \mathbb{K}^{i-1}$, tj.*

$$\mathbb{N}_{n \times K^{\max}} \left(\widehat{\mathbb{F}}, \mathbb{I}_n \otimes \left[(\mathbf{\Lambda}^i)^\top (\mathbf{\Sigma}^i)^{-1} \mathbf{\Lambda}^i + (\mathbb{K}^{i-1})^{-1} \right]^{-1} \right), \quad \text{viz tvrzení 18.}$$

- b) *Generuj $\mathbb{S}_{\text{prior}}^i$ z apriorního rozdělení parametru \mathbb{S} , tj. k -tý diagonální prvek z rozdělení $\mathbb{W}_1(\nu^* \xi_k^2, 1)$, viz (4.26).*

- c) *Generuj rozptyly $\mathbb{V}_{\text{prior}}^i$ z rozdělení $\mathbb{V} | \mathbb{K} = \mathbb{K}^{i-1}, \mathbb{S} = \mathbb{S}_{\text{prior}}^i$, tj. k -tý diagonální prvek z rozdělení*

$$\mathbb{W}_1 \left((\mathbb{S}_{\text{prior}}^i)_{k,k} \left((\mathbb{K}^{i-1})^{-1} \right)_{k,k}, \nu_\Omega - K^{\max} + 1 \right), \quad \text{viz (4.24).}$$

- d) *Převěď model s $(\mathbf{\Lambda}^i, \mathbb{F}^i, \mathbb{K}^{i-1})$ na model s $(\tilde{\mathbf{\Lambda}}^i, \tilde{\mathbb{F}}^i, \tilde{\mathbf{\Omega}}^i)$, kde*

$$\tilde{\mathbf{\Lambda}}^i := \mathbf{\Lambda}^i (\mathbb{V}_{\text{prior}}^i)^{-\frac{1}{2}}, \quad \tilde{\mathbb{F}}^i := \mathbb{F}^i (\mathbb{V}_{\text{prior}}^i)^{\frac{1}{2}}, \quad \tilde{\mathbf{\Omega}}^i := (\mathbb{V}_{\text{prior}}^i)^{\frac{1}{2}} \mathbb{K}^{i-1} (\mathbb{V}_{\text{prior}}^i)^{\frac{1}{2}}.$$

- e) *Generuj \mathbb{S}^i z plně podmíněného rozdělení $\mathbb{S} | \mathbf{\Omega} = \tilde{\mathbf{\Omega}}^i$, tj. k -tý diagonální prvek z rozdělení*

$$\mathbb{W}_1 \left(\left(\frac{1}{\nu^* \xi_k^2} + \left((\tilde{\mathbf{\Omega}}^i)^{-1} \right)_{k,k} \right)^{-1}, \nu_\Omega - K^{\max} \right), \quad \text{viz tvrzení 18.}$$

- f) *Generuj $\mathbf{\Omega}^i$ z plně podmíněného rozdělení $\mathbf{\Omega} | \mathbb{F} = \tilde{\mathbb{F}}^i, \mathbb{S} = \mathbb{S}^i$, tj. z rozdělení*

$$\mathbb{W}_{K^{\max}} \left((\tilde{\mathbb{F}}^i)^\top \tilde{\mathbb{F}}^i + \mathbb{S}^i, n + \nu_\Omega \right), \quad \text{viz tvrzení 18.}$$

g) Z nové rozptylové matice $\mathbf{\Omega}^i$ extrahuj diagonální prvky do matice $\mathbb{V}_{post}^i = \text{diag } \mathbf{\Omega}^i$. Nová korelační matice tak má tvar

$$\mathbb{K}^i := \left(\mathbb{V}_{post}^i \right)^{-\frac{1}{2}} \mathbf{\Omega}^i \left(\mathbb{V}_{post}^i \right)^{-\frac{1}{2}}.$$

h) Převed model s $(\tilde{\mathbf{\Lambda}}^i, \tilde{\mathbb{F}}^i, \mathbf{\Omega}^i)$ na model s $(\mathbf{\Lambda}^i, \mathbb{F}^i, \mathbb{K}^i)$, kde předdefinuj

$$\mathbf{\Lambda}^i := \tilde{\mathbf{\Lambda}}^i \left(\mathbb{V}_{post}^i \right)^{\frac{1}{2}}, \quad \mathbb{F}^i := \tilde{\mathbb{F}}^i \left(\mathbb{V}_{post}^i \right)^{-\frac{1}{2}}.$$

Konec algoritmu.

Všimněme si, že součin $\mathbb{F}^i (\mathbf{\Lambda}^i)^\top$ zůstává zachován, neboť

$$\mathbb{F}^i (\mathbf{\Lambda}^i)^\top = \underbrace{\mathbb{F}^i \left(\mathbb{V}_{prior}^i \right)^{\frac{1}{2}} \left(\mathbb{V}_{post}^i \right)^{-\frac{1}{2}}}_{\text{nové } \mathbb{F}^i} \cdot \underbrace{\left(\mathbb{V}_{post}^i \right)^{\frac{1}{2}} \left(\mathbb{V}_{prior}^i \right)^{-\frac{1}{2}} (\mathbf{\Lambda}^i)^\top}_{\text{nové } (\mathbf{\Lambda}^i)^\top}.$$

To je také důvod, proč je nutno parametr $\mathbf{\Lambda}$ předdefinovat.

Algoritmus 5 zde byl zjednodušen v porovnání s tím, který uvádějí Conti a kol. (2014). Autoři v onom článku navíc navrhují rozdělit faktory \mathbb{F} na aktivní a neaktivní. Neaktivní faktory pak lze jednoduše vyintegrovat, neboť ve věrohodnosti vůbec nevystupují. Generují se nejprve aktivní faktory, posléze se blokově generuje $\mathbf{\Omega}$ a až na závěr neaktivní faktory. Tato procedura však vyžaduje znalost toho, jak pracovat s jednotlivými bloky matice pocházející z inverzního Wishartova rozdělení. Vzhledem ke komplikovanosti této procedury ji zde již více nebudeme rozvádět.

Nicméně tato sofistikovaná modifikace velmi zlepšuje vlastnosti algoritmu. Snižuje například vysokou autokorelovanost generovaných korelačních matic, čímž zároveň zlepšuje schopnost algoritmu objevovat potenciální nové faktory mezi těmi neaktivními. Urychluje také konvergenci generovaného řetězce ke svému limitnímu (stacionárnímu) rozdělení.

4.4.2 Základní Gibbsův algoritmus

Nyní si ukažme algoritmus, pomocí kterého budeme generovat celý nový soubor (znáhodněných) parametrů $\boldsymbol{\theta}^i = \text{vec} (\mathbb{F}^i, \mathbb{B}^i, \mathbf{\Lambda}^i, \mathbf{\Sigma}^i, \mathbf{\Delta}^i, \boldsymbol{\tau}^{*,i}, \mathbb{K}^i)$, jestliže již máme k dispozici předchozí hodnotu vektoru $\boldsymbol{\theta}^{i-1}$. Potřebujeme opět vhodně zvolit pořadí. Zde začneme generováním trojice $(\mathbf{\Lambda}^i, \mathbf{\Sigma}^i, \mathbf{\Delta}^i)$ podle tvrzení 17. Díky tomu v úplně první iteraci nebude zapotřebí pro tyto parametry stanovovat počáteční hodnoty. Pokračovat budeme regresními koeficienty \mathbb{B}^i podle tvrzení 19. Nyní pomocí algoritmu 5 z předchozí části nagegenerujeme faktory \mathbb{F}^i a jejich korelační matici \mathbb{K}^i . Na závěr už jen upravíme pravděpodobnosti $\boldsymbol{\tau}^*$ pro příští iteraci.

Algoritmus 6. *Generování nového souboru parametrů $\boldsymbol{\theta}$ modelu určujících faktorů.*

Vstupy: data \mathbb{Y} a \mathbb{X} , poslední nagegenerované hodnoty $\boldsymbol{\theta}^{i-1}$.

Výstup: nové $\boldsymbol{\theta}^i$.

Postup: V následujícím pořadí proved tyto kroky:

1) Po vzoru tvrzení 17 postupně generuj

- a) Δ^i z podmíněného rozdělení $\Delta|\mathbb{Y}, \mathbb{X}, \mathbb{F}^{i-1}, \mathbb{B}^{i-1}, \boldsymbol{\tau}^{*,i-1}$, tj. j -tý řádek Δ_j^i nabyde hodnoty vektoru $e_k, k \in \{0, 1, \dots, K^{\max}\}$ s pravděpodobností

$$P\left(\Delta_j^i = e_k | \mathbb{Y}_{\bullet,j}, \mathbb{X}, \mathbb{F}^{i-1}, \mathbb{B}_j^{i-1}, \boldsymbol{\tau}^{*,i-1}\right) = \left[\sum_{l=0}^{K^{\max}} \exp(O_j(k \rightarrow l)) \right]^{-1},$$

- b) Σ^i z podmíněného rozdělení $\Sigma|\mathbb{Y}, \mathbb{X}, \mathbb{F}^{i-1}, \mathbb{B}^{i-1}, \Delta^i$, tj. j -tý diagonální prvek $(\sigma_j^2)^i$ z rozdělení $\text{IW}_1\left(\Psi_{j,\delta_j^i}, n + \nu_\sigma\right)$,
- c) Λ^i z podmíněného rozdělení $\Lambda|\mathbb{Y}, \mathbb{X}, \mathbb{F}^{i-1}, \mathbb{B}^{i-1}, \Sigma^i, \Delta^i$, tj. pro každé $j \in \{1, \dots, m\}$ a $k \in \{1, \dots, K^{\max}\}$ takové, že $\Delta_{j,k}^i = 0$, nastav $\lambda_{j,k} = 0$ a pro takové, že $\Delta_{j,k}^i = 1$, generuj $\lambda_{j,k}$ z rozdělení $\text{N}\left(\hat{\lambda}_{j,k}, r_{j,k} (\sigma_j^2)^i\right)$.

- 2) Generuj regresní koeficienty \mathbb{B}^i z plně podmíněného rozdělení, tj. j -tý řádek \mathbf{b}_j^i z rozdělení

$$\text{N}_{1+q}\left(\hat{\mathbf{b}}_j, \left[\frac{1}{(\sigma_j^2)^i} \mathbb{X}^\top \mathbb{X} + \mathbb{D}^{-1} \right]^{-1}\right), \quad \text{viz tvrzení 19.}$$

- 3) Generuj \mathbb{F}^i a \mathbb{K}^i pomocí algoritmu 5.

- 4) Generuj $\boldsymbol{\tau}^{*,i}$ z plně podmíněného rozdělení $\boldsymbol{\tau}^*|\Delta^i$, tj. první složku τ_0^i z rozdělení $\text{Beta}(\alpha + n_0(\Delta^i), \beta + m - n_0(\Delta^i))$ a zbývající složky $\boldsymbol{\tau}_{>0}^{*,i}$ z rozdělení $\text{Dir}_{K^{\max}}(\boldsymbol{\alpha}^* + \mathbf{n}(\Delta^i))$, viz rovnice 4.20.

Konec algoritmu.

Tento algoritmus lze ještě vylepšit tím, že v kroku 4) generujeme pro každé $j \in \{1, \dots, m\}$ jiné $\boldsymbol{\tau}_j^{*,i}$. Je však nutné namísto $n_0(\Delta^i)$ a $\mathbf{n}(\Delta^i)$ uvažovat pouze j -tý sčítanec, který se pro $k \in \{1, \dots, K^{\max}\}$ zjednodušuje na $\mathbb{1}(\delta_j^i = k)$. Takto jsou pravděpodobnosti aktivity faktoru $F_{j,k}$ daleko flexibilnější a dojde k zrychlení konvergence tvořeného markovského řetězce k posteriornímu rozdělení.

Ovšem algoritmus 6 bohužel sám o sobě nestačí pro generování onoho markovského řetězce. Tento postup totiž nezaručuje podmínku identifikovatelnosti z věty 15, tj. aby $n_k(\Delta)$ bylo buď nulové (k -tý faktor je neaktivní), nebo má aktivní faktor k alespoň 3 měření ($n_k(\Delta) \geq 3$), která ovlivňuje. Za účelem zajištění identifikovatelnosti ještě budeme potřebovat rozšířit algoritmus o takzvaný Metropolisův-Hastingsův algoritmus.

4.4.3 Metropolisův-Hastingsův algoritmus

Potřebujeme docílit toho, že generovaný soubor parametrů $\boldsymbol{\theta}^i$ splňuje podmínky identifikace z věty 15. V rámci generování tohoto i -tého souboru provedeme algoritmus 6 několikrát za sebou, posléze obrátíme pořadí generování a provedeme stejný počet opakování algoritmu 6. Poslední generovaný soubor budeme považovat za kandidáta na další stav $\boldsymbol{\theta}^i$. V takzvaném Metropolisově-Hastingsově kroku pak tohoto kandidáta přijmeme, pokud splňuje podmínky identifikovatelnosti ve větě 15. Jestliže tyto podmínky nespĺňuje, tak markovský řetězec setrvá v předchozím stavu $\boldsymbol{\theta}^{i-1}$.

Zde je matematický zápis tohoto algoritmu.

Algoritmus 7. *Metropolisův-Hastingsův algoritmus generující nový stav markovského řetězce splňující podmínky identifikovatelnosti.*

Vstupy: data \mathbb{Y} a \mathbb{X} , poslední nagenovaný soubor parametrů $\boldsymbol{\theta}^{i-1}$ splňující podmínky identifikovatelnosti, číslo $S \in \mathbb{N}$.

Výstup: soubor nových parametrů $\boldsymbol{\theta}^i$ splňující podmínky identifikovatelnosti.

Postup: V následujícím pořadí proved tyto kroky:

i) *Polož výchozí bod $\widehat{\boldsymbol{\theta}}_0 := \boldsymbol{\theta}^{i-1}$.*

ii) *Pro $s \in \{1, \dots, S\}$ generuj pomocné stavy $\widehat{\boldsymbol{\theta}}_s$ pomocí algoritmu 6 s výchozí hodnotou $\widehat{\boldsymbol{\theta}}_{s-1}$ (v pořadí od 1) po 4)).*

iii) *Polož $\check{\boldsymbol{\theta}}_S := \widehat{\boldsymbol{\theta}}_S$.*

iv) *Pro $s \in \{0, \dots, S-1\}$ generuj pomocné stavy $\check{\boldsymbol{\theta}}_s$ pomocí algoritmu 6 s výchozí hodnotou $\check{\boldsymbol{\theta}}_{s+1}$, ovšem tentokrát v opačném pořadí (od 4) po 1)).*

v) *Jestliže poslední generovaný pomocný stav $\check{\boldsymbol{\theta}}_0$ splňuje podmínky identifikovatelnosti z věty 15, tak polož $\boldsymbol{\theta}^i := \check{\boldsymbol{\theta}}_0$. V opačném případě polož $\boldsymbol{\theta}^i := \boldsymbol{\theta}^{i-1}$.*

Konec algoritmu.

I když v kroku iv) postupujeme při použití algoritmu 6 v opačném pořadí od 4) po 1), tak v rámci kroku 1), příp. 3), generujeme stále v pořadí od a) po c), příp. h). Částečná marginalizace a struktura algoritmu 5 totiž vyžadují generování přesně v tomto pořadí.

Během tohoto Metropolisova-Hastingsova algoritmu celkem generujeme $2S$ stavů, které nemusí nutně splňovat podmínky identifikace, ovšem výsledné $\boldsymbol{\theta}^i$ již tyto podmínky splňovat musí. Těchto $2S$ mezikroků je po skončení tohoto algoritmu zapomenuto a není již použito pro finální inferenci. Číslo S může být stanoveno pevně, ovšem autoři článku doporučují při každé iteraci i tento počet náhodně měnit (např. generované z Poissonova rozdělení). Tato změna může pomoci v situacích, kdy se například algoritmus na nějakou dobu zastaví v jednom stavu, ze kterého je obtížné se dostat.

Pro úplně první spuštění algoritmu 6 nepotřebujeme znát hodnoty $\boldsymbol{\Lambda}$, $\boldsymbol{\Sigma}$ ani $\boldsymbol{\Delta}$. Ovšem v případě, že hned na počátku se rozhodneme po dokončení pomocných mezikroků zůstat ve výchozím stavu, potřebujeme tyto hodnoty znát předem (navíc musí splňovat podmínky identifikovatelnosti). Výchozí hodnota $\boldsymbol{\Delta}$ pak často má velmi rozhodující roli v rychlosti konvergence markovského řetězce, a tedy i v délce potřebné zážehové fáze. Proto Conti a kol. (2014) doporučují použít několika předběžných iterací algoritmu 6, který ovšem nakonec nemusí nutně dát $\boldsymbol{\Delta}$ s požadovanou vlastností. V tomto případě se musí faktory s pouze jedním či dvěma spjatými měřeními ($n_k(\boldsymbol{\Delta}) \in \{1,2\}$) předefinovat na neaktivní faktory (ani jedno spjaté měření) a s tím provést i potřebné úpravy ostatních parametrů.

Conti a kol. (2014) ukázali, že takto sestrojený algoritmus splňuje tzv. *detailní podmínku rovnováhy*, ze které plyne reversibilita výsledného markovského řetězce. Markovský řetězec vzniklý neustálým opakováním algoritmu 7 konverguje ke svému limitnímu (stacionárnímu) rozdělení. Můžeme ho proto použít pro odhady parametrů modelu a především také k odhadu vhodného počtu faktorů. Podrobné důkazy těchto vlastností však jdou nad rámec této diplomové práce.

4.5 Bayesovský odhad parametrů a počtu aktivních faktorů K

V této části si ukážeme, jak využít algoritmů představených v předchozí části k odhadu parametrů modelu a především k odhadu počtu aktivních faktorů K .

Již jsme zmínili, že začneme iterováním algoritmu 6 pro nalezení identifikovatelného a mnohem vhodnějšího výchozího stavu θ^0 . Dle našeho uvážení zvolíme dostatečně velké $B \in \mathbb{N}$ určující délku zážehové fáze. Jak jsme si uvedli v části 2.3.1, dopředu nevíme, jak dlouho by mělo řetězci trvat, než dosáhne svého limitního rozdělení. Jediné podle čeho můžeme usoudit potřebnou délku B jsou tedy rozměry našich dat n a m a naše zkušenosti. Pomocí B iterací algoritmu 7 zažehneme markovský řetězec se stavy $\theta^1, \dots, \theta^B$. Tyto stavy však pro výsledné odhady nepoužijeme. Dále si zvolíme $M \in \mathbb{N}$ – počet opakování algoritmu 7. Těchto M stavů $\theta^{B+1}, \dots, \theta^{B+M}$ již použijeme pro odhad parametrů modelu, viz část 2.3.1.

Počet aktivních faktorů v i -tém stavu určíme z indikátorové matice Δ^i jako počet faktorů, které ovlivňují alespoň 3 měření z \mathbf{Y} . Ostatní faktory jsou neaktivní – neovlivňují žádné z měření z \mathbf{Y} . Připomínáme značení n_k počtu měření, které ovlivňuje faktor $k \in \{1, \dots, K^{\max}\}$, které lze spočítat jako

$$n_k(\Delta^i) = \sum_{j=1}^m \Delta_{j,k}^i = \sum_{j=1}^m \mathbb{1}(\delta_j^i = k).$$

Jelikož všechny stavy $\theta^i, i \in \{B+1, \dots, B+M\}$ splňují podmínky identifikovatelnosti, tak buď $n_k(\Delta^i) = 0$, nebo $n_k(\Delta^i) \geq 3$. Počet aktivních faktorů ve stavu i spočteme jako

$$K^i = \sum_{k=1}^{K^{\max}} \mathbb{1}(n_k(\Delta^i) \neq 0). \quad (4.40)$$


Za odhad počtu aktivních faktorů K pak považujeme výběrový průměr $\bar{K} = \frac{1}{M} \sum_{i=B+1}^{B+M} K^i$ či výběrový modus.

Dále používáme $n_0(\Delta^i) = m - \sum_{k=1}^{K^{\max}} n_k(\Delta^i) = \sum_{j=1}^m \mathbb{1}(\delta_j^i = 0)$. Pomocí výběrového průměru (modu) těchto $n_0(\Delta^i)$ tak lze odhadnout očekávaný počet měření, které nejsou ovlivněny žádným z faktorů. Stejně tak lze pomocí průměru hodnot $n_k(\Delta^i)$ odhadnout očekávaný počet měření, jež jsou ovlivněny právě faktorem k .

Jediným nevyřešeným problémem zůstává identifikovatelnost našeho modelu, až na změnu pořadí faktorů a změnu jejich znaménka. Neexistuje žádné přirozené řazení sloupců matice Δ , lze si proto zvolit vlastní způsob řazení a ten dodržovat v každém stavu. Conti a kol. (2014) doporučují ještě jinou alternativu, kde vždy srovnají sloupce vzestupně podle čísla příslušného řádku, kde se v daném sloupci poprvé (od shora) vyskytuje nenulový prvek. Takto je třeba přeuspořádat nejen Δ , ale také zátěže Λ , faktory \mathbb{F} a korelační matici \mathbb{K} .

Pro účely identifikace znaménka sloupce k je zapotřebí si určit takovou zátěž $\lambda_{j,k}$, která s největší aposteriorní pravděpodobností bude různá od nuly. Podle významu faktoru k a měření j se potom statistik rozhodne, zda má být tato zátěž kladná či záporná. Jde zde čistě jen o interpretační záležitost pramenící především z podstaty zkoumaného problému. Vždy, když tato kontrolní zátěž pro sloupec k má opačné znaménko, tak se celý k -tý sloupec zátěží a příslušné k -té faktory přenásobí (-1) .

4.6 Zobecnění bayesovského modelu určujících faktorů

Tímto jsme vyložili teoretické pozadí modelu určujících faktorů a představili si jednotlivé kroky algoritmu, pomocí kterého lze sestavit odhady parametrů tohoto modelu. Piatek (2017) implementoval celý tento algoritmus v balíčku `BayesFM` statistického programu . Pomocí funkce `befa` budeme v následující kapitole provádět simulační studii.

Autoři dokonce uvažují model, kde jednotlivá měření Y_j mohou být i binární veličiny (hodnoty pouze 0 nebo 1). Ty jsou modelovány tak, že na pozadí figuruje latentní normálně rozdělená veličina Y_j^* , jejíž hodnoty jsou hodnotou 0 rozděleny na kladnou a nekladnou část:

$$Y_j^* \in \begin{cases} (-\infty, 0] & \implies Y_j = 0, \\ (0, \infty) & \implies Y_j = 1, \end{cases}.$$

V modelu určujících faktorů pak vystupuje Y_j^* , ovšem u apriorní specifikace parametrů je rozptyl σ_j^2 fixován na konstantě 1. To má své důsledky na nepatrné změny parametrů rozdělení v algoritmu 6. Zde pak přibývá jeden krok, ve kterém se generují latentní veličiny $Y_{\bullet j}^*$ z jejich plně podmíněného rozdělení. Podmiňuje se zde naměřenými binárními veličinami, regresory a ostatními parametry modelu. Toto vede na generování z takzvaného *useknutého* normálního rozdělení (na kladnou nebo nekladnou část).

Model určujících faktorů pak lze rozšířit i pro ordinální kategorické veličiny (o více jak dvou kategoriích), u kterých lze předpokládat existenci skryté reálné latentní proměnné. Mluvíme tedy například o případech, kdy bychom měli k dispozici pouze výškové či váhové kategorie, ale jistě víme, že každý subjekt zde měl svou konkrétní výšku a váhu. Dalším příkladem pak mohou být zaokrouhlované veličiny. Nicméně doposud pro tyto víceúrovňové ordinální kategorické veličiny algoritmus implementován nebyl.

Autoři dále pracují na zobecnění modelu určujících faktorů. Hlavním předmětem zájmu je opuštění předpokladu, že každé měření může být ovlivněno nejvýše jediným z faktorů. V reálném světě totiž často o výsledku zkoumané veličiny rozhoduje vícero různých faktorů.

5. Simulační studie

V této závěrečné části pomocí simulační studie prozkoumáme schopnost algoritmu představeného v předchozí části odhadovat správně skutečný počet faktorů. Tato schopnost nás bude zajímat nejen v případě platnosti modelu určujících faktorů (podrobněji část 4.1), ale také v obecném modelu faktorové analýzy. Tyto odhady skutečného počtu faktorů porovnáme s několika běžnými metodami pro určení počtu faktorů.

Pro bayesovský odhad modelu určujících faktorů využijeme již zmiňovaného balíčku `BayesFM` (Piatek, 2017) pro statistický program \mathbb{R} . Obsahuje funkci `simul.dedic.facmod`, kterou lze použít pro generování modelu určujících faktorů (bez regresorů) s předem stanovenou indikátorovou maticí Δ . Hlavní funkcí je však `befa`, která provádí algoritmus pro generování markovského řetězce. Na výstup této funkce lze aplikovat funkce `post.column.switch` a `post.sign.switch`, které sjednotí pořadí a znaménka jednotlivých faktorů pro ucelenou interpretaci. My se však budeme především zaměřovat na složku `nfac` výstupu funkce `befa`, která obsahuje počty aktivních faktorů v jednotlivých iteracích markovského řetězce. Nutno dodat, že funkce `befa` má celou řadu vstupních parametrů, jejichž označení odpovídá značení v článku Conti a kol. (2014), které se v řadě případů liší od toho našeho. Rozdíl ve značení lze jednoduše získat srovnáním analogických diagramů na obrázcích 4.1 a D1.

Klasické metody odhadu počtu faktorů, se kterými budeme srovnávat bayesovský přístup, jsou implementovány v balíčku `nFactors` (Raiche, 2010). Funkce `nScree` dává čtvery různé odhady počtu faktorů založené na vlastních číslech výběrové korelační matice. Kaiserovo pravidlo (*Kaiser rule*) stanovuje počet faktorů jako počet vlastních čísel vyšších než 1. Metoda paralelní analýzy (*parallel analysis*) je velmi podobná, jen srovnává s jinou hodnotou, která je určena pomocí daného kvantilu. Metoda faktoru akcelerace (*acceleration factor*) si seřadí jednotlivá vlastní čísla a určí, od jakého okamžiku dochází k výraznému nárůstu vlastních čísel, výsledkem je potom počet vlastních čísel vyšších než bod tohoto zlomu a splňujících podmínku paralelní analýzy. Poslední metodou je metoda optimálních souřadnic (*optimal coordinates*), která je založena na extrapolaci vlastních čísel. Poslední dvě metody nemohou navrhnout vyšší číslo, než které navrhuje metoda paralelní analýzy. Podrobnější definice těchto metod viz dokumentace balíčku `nFactors`.

5.1 Popis scénáře a volba parametrů

Budeme zkoumat tři rozdílné situace. První bude situace, kdy je model určujících faktorů splněn, tedy faktorové zátěže jsou až na maximálně jednu hodnotu v řádku nulové. Ve druhé situaci ponecháme strukturu faktorových zátěží modelu určujících faktorů, ovšem hodnoty, které by měly být nulové, nahradíme číslem blízkým nule (v intervalu $(-0.2, 0.2)$). Očekáváme zde, že by bayesovský přístup mohl dobře fungovat i v tomto případě, kde máme obecný model faktorové analýzy, ovšem s nejvýše jedním dominantním faktorem pro každé pozorování. Třetí situací pak bude případ, kdy faktorové zátěže nemají žádnou strukturu řídké matice, a tedy výsledná měření jsou vždy lineární kombinací všech faktorů (do-

minimálních faktorů může být více).

Pro každou z těchto tří situací budeme uvažovat skutečné počty faktorů $K \in \{2, 3, 5\}$. Pro každé takové K prozkoumáme vždy trojici různých počtů měření m (hodnoty viz tabulka 5.1). V článku Conti a kol. (2014) zkoumali za platnosti modelu určujících faktorů pouze scénář, kdy každý aktivní faktor ovlivňuje vždy předem daný *stejný* počet měření. Zde budeme uvažovat pro každou dvojici (K, m) jak tento *vyvážený* scénář, tak scénář mírně *nevyvážený*, a dokonce i *extrémní* scénář, kde máme jeden faktor ovlivňující většinu měření, zatímco ostatní faktory ovlivňují minimální počet měření (tedy 3). V posledním zmíněném scénáři pak také některá měření nejsou spojena s žádným faktorem. Tuto (ne)vyváženost však lze provést pouze v prvních dvou situacích – třetí situace žádnou strukturu faktorových zátěží nepředpokládá.

Faktorové zátěže dominantních faktorů v prvních dvou situacích stanovujeme jako ekvidistantní dělení intervalu $[0.4, 0.8]$ (počet dělicích bodů je roven počtu měření, které jsou spojeny s tímto faktorem dle určeného Δ). Těmto dělicím bodům navíc ještě náhodně přiřadíme znaménko $+$, nebo $-$ (s pravděpodobností 0.5). Ve třetí situaci nemusíme dodržovat žádnou strukturu Λ , generují se proto z rovnoměrného rozdělení na intervalu $[-0.8, 0.8]$.

Diagonální prvky σ_j^2 rozptylové matice Σ stanovíme jako $\sigma_j^2 = 0.5$ pro měření j , které je spojeno s nějakým faktorem ($\delta_j \neq 0$), a $\sigma_j^2 = 1$, pokud $\delta_j = 0$. Korelační matici faktorů \mathbb{K} zvolíme jako jednotkovou matici \mathbb{I}_K , což odpovídá nezávislým faktorům.

Pro obdržení jednoho pozorovaného vektoru měření \mathbf{Y} je nejprve zapotřebí generovat vektor faktorů \mathbf{F} z $\mathbf{N}_K(\mathbf{0}_K, \mathbb{I}_K)$ a vektor chyb $\boldsymbol{\varepsilon}$ z $\mathbf{N}_m(\mathbf{0}_m, \Sigma)$. Výsledná měření dopočteme jako součet $\mathbf{Y} = \Lambda\mathbf{F} + \boldsymbol{\varepsilon}$. Jelikož se zde zaměřujeme pouze na odhad počtu faktorů, tak si dovoluujeme neuvažovat žádný vliv regresorů či posun ve střední hodnotě. Sníží to počet odhadovaných parametrů a tím i rychlost algoritmu. Navíc platí základní model faktorové analýzy, pro nějž lze použít klasické metody odhadu ideálního počtu faktorů. Ve všech situacích a scénářích takto generujeme celkem $n = 100$ pozorování.

Takto nagenеровanou matici pozorování \mathbb{Y} pak dále předáme funkci `befa`, která provede bayesovský odhad, a funkci `nScree` pro klasické odhady počtu faktorů. V této fázi již postupujeme tak, jako by nám K , Λ , Δ , Σ a \mathbb{K} nebyly známé. V samotné funkci `befa` lze upřesnit nastavení hyperparametrů modelu či počátečních hodnot parametrů. V této simulaci se budeme držet přednastavených hodnot, viz níže. Změníme pouze několik parametrů. Nastavíme si délku zážehové fáze $B = 20\,000$ (`burnin`) a délku markovského řetězce, který se použije pro odhad parametrů modelu $M = 20\,000$ (`niter`). Maximální počet faktorů $K^{\max} = K_{\max}$ určujeme jako dolní celou část z podílu $m/3$ (což je nejvyšší možný počet faktorů takový, aby byl platný model určujících faktorů).

Hyperparametry jsou přednastaveny funkcí `befa` následovně:

- $b_{0,j} = 0$ (*nulová apriorní střední hodnota regresních koeficientů*),
- $\mathbb{D} = \text{diag } \mathbf{B0} = 10\mathbb{I}_2$ (*apriorní varianční matice regresních koeficientů*),
- $q_j = 2\mathbf{C0} = 2$ (parametr inverzního Wishartova rozdělení pro σ_j^2),
- $\nu_\sigma = 2(\mathbf{c0} + 1) = 2(2 + 1) = 6$ (stupně volnosti inverzního Wishartova rozdělení pro σ_j^2),

- $\lambda_{0,j} = 0$ (nulová apriorní střední hodnota nenulových faktorových zátěží),
- $a_j = A0 = 10$ (apriorní rozptyl nenulových faktorových zátěží),
- $\alpha = \text{kappa0} = 2$ (první parametr beta rozdělení pro τ_0),
- $\beta = \text{xi0} = 1$ (druhý parametr beta rozdělení pro τ_0),
- $\alpha^* = \text{kappa} = \mathbf{1}_{K^{\max}}/K^{\max}$ (parametr Dirichletova rozdělení pro $\tau_{>0}^*$),
- ν_Ω a ξ_k^2 jsou stanoveny dle doporučení v práci Huang a Wand (2013).

Každé měření j má dle doporučení zmíněných v části 4.2.1 svůj vlastní parametr $\tau_{0,j}$. Počet mezikroků S , které jsou provedeny v Metropolisově-Hastingsově algoritmu, je volen náhodně jako $S = 1 + P$, kde náhodná veličina P je generována z Poissonova rozdělení se střední hodnotou 4 (`n.step - 1`), tj. $P \sim \text{Pois}(4)$.

Výchozí hodnoty parametrů Λ , Σ a \mathbb{B} budeme generovat z jejich apriorního rozdělení. Výchozí \mathbb{K} je přednastaveno na $\mathbb{I}_{K^{\max}}$. Jelikož výchozí hodnota Δ má značný vliv na rychlost konvergence řetězce k posteriornímu rozdělení, tak Δ nebudeme přidělovat zcela náhodně. V každém scénáři s m měřeními iterujeme $m \cdot 10\,000$ algoritmus 6 (funkce `befa` s nastavením `Nid = 1`, `burnin = 0` a `niter = m \cdot 10\,000`). Poslední nagenеровané Δ použijeme jako výchozí hodnotu. Je ale zapotřebí ošetřit, že toto Δ splňuje identifikační podmínky. Je-li j -té měření přisuzováno faktoru, kterému není celkem přisouzeno 3 a více faktorů, tak toto přisouzení zrušíme, tedy nastavíme $\delta_j = 0$.

Celý tento proces generování \mathbb{Y} a následného odhadu zopakujeme 100-krát se stejnými Λ , Σ , Δ a \mathbb{K} .

5.2 Výsledky studie

V této poslední části si ukážeme výsledky provedené simulační studie. Ty se nám podařilo zachytit ve třech tabulkách 5.1, 5.2 a 5.3, každá se věnuje jedné z uvažovaných situací.

Nejprve se podívejme do tabulky 5.1, která zachycuje výsledky v situaci, kdy je model určujících faktorů splněn. Vidíme zde, že bayesovské odhady jsou velmi blízké skutečnému počtu aktivních faktorů K , zejména v případě vyrovnaného počtu měření přisouzeného jednotlivým faktorům. Potvrzujeme tak výsledky, které uvádějí Conti a kol. (2014) ve svém článku. Ovšem zaměříme-li se na scénáře s nevyrovnaným rozložením měření na faktory, tak zvláště u těch extrémních pozorujeme mírné vychýlení odhadu. Zjevně je v takto extrémním scénáři těžké objevit skrytý faktor, když ovlivňuje pouze troje měření (nejnižší povolený počet). Mezi klasickými metodami zde dávají srovnatelné výsledky metody optimálních souřadnic a paralelní analýzy. Metoda faktoru akcelerace selhává v nevyvážených scénářích (navrhuje daleko menší počet faktorů) a Kaiserovo pravidlo výrazně nadhodnocuje skutečný počet skrytých faktorů.

Druhou situací je případ, kdy je téměř dodržen model určujících faktorů. V modelu se vyskytují dominantní faktory, které jsou určeny stejným způsobem jako v předchozí situaci, avšak navíc ostatní faktory také přispívají malým dílem do výsledného měření. Není tak dodržen model určujících faktorů, ale je ho téměř dosaženo. Očekávali jsme, že i v takovéto situaci by MCMC metody navržené ve čtvrté kapitole mohly rozumně dobře fungovat. Z tabulky 5.2 však můžeme vyčíst,

že bayesovské odhady jsou mírně nadhodnocené. Zvláště pak v případech s vyšším počtem měření m pro dané K . Nevyvážení rozložení počtu měření příslušných faktorům má opět tendenci snižovat počet odhadovaných faktorů. V této situaci to tedy odhady přibližuje skutečnosti, ovšem vychýlenost je zřejmá. Klasické metody odhadu (zejména optimální souřadnice a paralelní analýza) však stále navrhuji přibližně správný počet faktorů. Metoda faktoru akcelerace a Kaiserovo pravidlo mají stejné nedostatky jako v předchozí situaci.

Třetí situací byl obecný model faktorové analýzy bez žádné dané struktury faktorových zátěží. Tabulka 5.3 jasně ukazuje, že bayesovské odhady, které předpokládají platnost modelu určujících faktorů, jsou výrazně vychýlené. Navržením daleko vyššího počtu faktorů se snaží podchytit společné efekty jednotlivých faktorů, které model určujících faktorů nepředpokládá. Čím větší je počet měření, tím vyšší počet faktorů bayesovský přístup navrhuje. Na druhou stranu klasické odhady v tomto případě navrhuji téměř vždy správný počet faktorů (až na drobná vychýlení faktoru akcelerace).

Příkládáme zde také obrázky 5.1, 5.2 a 5.3, které se zabývají rozložením počtu faktorů ve všech generovaných markovských řetězcích (až od dosažení zážehové fáze $B = 20000$). Vidíme zde vždy procentuální zastoupení jednotlivých navrhovaných počtů faktorů dělené podle uvažované situace a také dle vyváženosti rozložení faktorů. Jedná se tak o odhad aposteriorního rozdělení počtu aktivních faktorů K . Zelené sloupce odpovídající první situaci, kdy platí model určujících faktorů, dosahují nejvyšších výšek právě ve vymezeném fialovém obdélníku označujícím skutečný počet faktorů. Pozorujeme zde také mírné posunutí modrých sloupců odpovídajících situaci s dominantními faktory. Oranžové sloupce odpovídající situaci obecného počtu faktorů jsou již znatelně posunuté vpravo, tedy navrhuji daleko vyšší počet faktorů, než je ten skutečný.

Objevené problémy s bayesovským odhadem mohou spočívat v nízkém počtu pozorování $n = 100$ (málo informace obsažené v datech) či nízkém pevném $B = 20000$ určujícím délkou zážehové fáze. Pro výpočetní náročnost bylo nutné stanovit n , B a M pevně, a to nepříliš malé ani velké. Zřejmě však bude nutná jiná délka zážehové fáze v případě $K = 2$ a $m = 6$ než v případě $K = 5$ a $m = 50$. Ovšem při generování tolika markovských řetězců nebylo možné zabývat se každým z nich zvlášť a ověřovat zda již dosahuje stacionárního rozdělení. Takto opatrní a poctiví můžeme být při analýze reálných dat, kde aplikujeme bayesovské metody faktorové analýzy pouze jednou. V takovémto případě doporučujeme porovnat s výsledky klasických metod. Pokud bayesovský model určujících faktorů navrhuje daleko více skrytých faktorů, než klasické metody, tak zřejmě se na výsledku jednoho měření může podílet i více faktorů, což popírá model určujících faktorů. Vidíme zde tedy, jak je nutné zobecnit tento model i pro tuto možnost, že měření mohou být ovlivňována i více faktory.

Tabulka 5.1: První situace – model určujících faktorů. Odhady skutečného počtu faktorů K při různém počtu celkových měření m a různém rozvržení daném indikátorovou maticí Δ . Počet pozorování $n = 100$, počet opakování simulace 100, délka zážehové fáze $B = 20000$, délka markovského řetězce použitého pro inferenci $M = 20000$. Průměr přes všechna opakování {směrodatná odchylka} (nejčastější navrhovaná hodnota) [počet případů navržení nejčastější hodnoty ze 100 možných].

K	m	K^{\max}	Rozvržení ($\mathbf{n}_k(\Delta)$) ¹					Bayesovské odhady ²		Klasické odhady					
			0	1	2	3	4	5	Průměr	Modus	OC ³	AF ⁴	P ⁵	Kaiser ⁶	
2	6	2	3	3				1.76 {0.34}	1.80 (2) [80]	0.88 (0) [56]	2.01 (2) [95]	2.02 (2) [98]	2.02 (2) [98]		
			6	6					2.06 {0.08}	2.02 (2) [98]	1.64 (2) [82]	2.00 (2) [100]	2.00 (2) [100]	2.28 (2) [73]	
	12	4	7	5				2.07 {0.15}	2.03 (2) [97]	1.98 (2) [99]	2.00 (2) [100]	2.00 (2) [100]	2.34 (2) [66]		
			3	6	3				1.93 {0.26}	1.91 (2) [89]	2.64 (2) [46]	1.35 (1) [65]	2.71 (3) [44]	4.04 (4) [86]	
	24	8		12	12				2.28 {0.24}	2.09 (2) [91]	1.90 (2) [95]	2.00 (2) [100]	2.00 (2) [100]	4.62 (5) [46]	
				14	10				2.32 {0.26}	2.14 (2) [86]	2.00 (2) [100]	2.00 (2) [100]	2.00 (2) [100]	4.52 (4) [43]	
6			15	3				2.03 {0.30}	1.85 (2) [81]	2.12 (2) [84]	1.00 (1) [100]	2.12 (2) [84]	6.98 (7) [58]		
15	5		5	5	5			2.99 {0.08}	2.99 (3) [99]	2.38 (3) [73]	2.99 (3) [98]	3.01 (3) [99]	3.74 (4) [62]		
			7	5	3			2.71 {0.45}	2.69 (3) [65]	2.95 (3) [97]	2.60 (3) [71]	3.01 (3) [99]	3.55 (4) [49]		
		3	6	3	3			2.72 {0.41}	2.72 (3) [73]	3.18 (3) [43]	1.24 (1) [85]	3.60 (3) [48]	5.15 (5) [77]		
3	21	7	7	7	7			3.09 {0.15}	3.03 (3) [97]	2.54 (3) [82]	3.00 (3) [100]	3.00 (3) [100]	4.49 (5) [49]		
			10	7	4				3.03 {0.15}	2.98 (3) [98]	3.00 (3) [100]	2.70 (3) [82]	3.00 (3) [100]	4.61 (4) [46]	
			6	9	3	3			2.61 {0.46}	2.55 (3) [56]	3.29 (3) [52]	1.00 (1) [100]	3.58 (3) [54]	7.08 (7) [66]	
30	10		10	10	10			3.18 {0.21}	3.09 (3) [91]	2.83 (3) [94]	3.00 (3) [100]	3.00 (3) [100]	6.43 (6) [49]		
			15	10	5			3.13 {0.18}	3.03 (3) [95]	3.00 (3) [100]	2.52 (3) [70]	3.00 (3) [100]	6.44 (7) [41]		
		9	15	3	3			2.52 {0.56}	2.40 (3) [47]	3.17 (3) [55]	1.00 (1) [100]	3.31 (3) [59]	9.74 (10) [57]		
25	8		5	5	5	5	5	4.98 {0.13}	4.97 (5) [97]	3.93 (5) [71]	4.73 (5) [92]	4.99 (5) [99]	6.88 (7) [60]		
			7	6	5	4	3		4.56 {0.46}	4.55 (5) [55]	4.45 (5) [81]	3.20 (5) [44]	4.97 (5) [95]	6.79 (7) [56]	
		3	9	3	3	3	3	3.64 {0.75}	3.64 (4) [40]	4.51 (5) [53]	1.00 (1) [100]	5.28 (5) [68]	8.36 (8) [56]		
5	35	11	7	7	7	7	7	5.04 {0.06}	5.00 (5) [100]	4.45 (5) [85]	4.87 (5) [96]	5.00 (5) [100]	8.92 (9) [59]		
				11	9	7	5	3		4.66 {0.46}	4.63 (5) [64]	4.84 (5) [94]	2.67 (1) [32]	4.97 (5) [97]	8.95 (9) [55]
			8	15	3	3	3	3	3.62 {0.66}	3.63 (4) [49]	4.45 (5) [47]	1.00 (1) [100]	5.00 (5) [59]	11.43 (11) [50]	
50	16		10	10	10	10	10	5.06 {0.09}	5.01 (5) [99]	4.55 (5) [89]	4.89 (5) [97]	5.00 (5) [100]	12.62 (13) [48]		
			16	13	10	7	4		4.96 {0.25}	4.93 (5) [93]	4.94 (5) [97]	2.79 (1) [31]	4.98 (5) [98]	12.55 (13) [41]	
		13	25	3	3	3	3	3.45 {0.62}	3.41 (3) [55]	3.71 (4) [42]	1.00 (1) [100]	3.96 (4) [44]	15.70 (16) [53]		

¹ Počet měření $j \in \{1, \dots, m\}$ takových, že δ_j je rovno k .

² Metoda, pomocí níž bylo z markovského řetězce délky $M = 20000$ odhadnuto K .

³ Metoda optimálních souřadnic (*optimal coordinates*).

⁴ Metoda faktorů akcelerace (*acceleration factor*).

⁵ Metoda paralelní analýzy (*parallel analysis*).

⁶ Kaiserovo pravidlo (*Kaiser rule*).

Tabulka 5.2: Druhá situace – model s dominantním faktorem. Odhady skutečného počtu faktorů K při různém počtu celkových měření m a různém rozvržení daném indikátorovou maticí Δ . Počet pozorování $n = 100$, počet opakování simulace 100, délka zážehové fáze $B = 20000$, délka markovského řetězce použitého pro inferenci $M = 20000$. Průměr přes všechna opakování {směrodatná odchylka} (nejčastější navrhaná hodnota) [počet případů navržení nejčastější hodnoty ze 100 možných].

K	m	K ^{max}	Rozvržení ($n_k(\Delta)$) ¹					Bayesovské odhady ²		Klasické odhady				
			0	1	2	3	4	5	Průměr	Modus	OC ³	AF ⁴	P ⁵	Kaiser ⁶
2	6	2	3	3				1.81 {0.32}	1.81 (2) [81]	0.94 (0) [53]	2.00 (2) [98]	2.01 (2) [99]	2.01 (2) [99]	
			6	6					2.35 {0.34}	2.25 (2) [75]	1.68 (2) [84]	2.00 (2) [100]	2.00 (2) [100]	2.32 (2) [68]
	12	4	7	5				2.12 {0.16}	2.04 (2) [96]	1.98 (2) [99]	2.00 (2) [100]	2.00 (2) [100]	2.32 (2) [68]	
			3	6	3				2.27 {0.53}	2.26 (2) [53]	2.35 (2) [67]	1.79 (2) [79]	2.37 (2) [66]	3.68 (4) [64]
	24	8	12	12				3.24 {0.54}	3.10 (3) [59]	1.88 (2) [94]	2.00 (2) [100]	2.00 (2) [100]	4.35 (4) [38]	
			14	10					2.88 {0.51}	2.73 (3) [50]	2.00 (2) [100]	1.98 (2) [98]	2.00 (2) [100]	4.32 (4) [44]
		6	15	3			2.95 {0.65}	2.77 (3) [52]	2.03 (2) [97]	1.00 (1) [100]	2.03 (2) [97]	6.52 (7) [49]		
3	15	5	5	5	5			3.10 {0.23}	3.06 (3) [92]	2.50 (3) [78]	2.97 (3) [98]	3.01 (3) [99]	3.41 (3) [62]	
			7	5	3				3.39 {0.53}	3.38 (3) [47]	2.88 (3) [96]	2.72 (3) [79]	3.00 (3) [100]	3.46 (3) [57]
			3	6	3	3			2.78 {0.43}	2.72 (3) [64]	3.14 (3) [74]	1.34 (1) [80]	3.24 (3) [74]	4.54 (4) [49]
	21	7	7	7	7			4.02 {0.51}	3.94 (4) [58]	2.79 (3) [92]	2.96 (3) [98]	3.00 (3) [100]	4.39 (4) [57]	
			10	7	4				3.77 {0.47}	3.72 (4) [60]	3.00 (3) [100]	2.18 (3) [56]	3.00 (3) [100]	4.34 (4) [55]
			6	9	3	3			3.08 {0.64}	3.04 (3) [41]	3.10 (3) [72]	1.02 (1) [98]	3.26 (3) [76]	6.47 (6) [46]
30	10	10	10	10			4.71 {0.69}	4.66 (5) [43]	2.90 (3) [96]	3.00 (3) [100]	3.00 (3) [100]	6.08 (6) [56]		
		15	10	5				4.83 {0.54}	4.82 (5) [57]	3.00 (3) [100]	2.69 (3) [78]	3.00 (3) [100]	5.92 (6) [47]	
		9	15	3	3			4.06 {0.80}	3.95 (5) [34]	3.02 (3) [93]	1.00 (1) [100]	3.05 (3) [93]	8.98 (9) [50]	
25	8	5	5	5	5	5	5.62 {0.47}	5.61 (6) [54]	4.51 (5) [86]	3.46 (5) [50]	4.98 (5) [98]	6.11 (6) [59]		
		7	6	5	4	3		5.07 {0.64}	5.04 (5) [58]	4.63 (5) [84]	3.58 (5) [50]	4.93 (5) [93]	6.25 (6) [59]	
		3	9	3	3	3	3	4.46 {0.78}	4.42 (5) [41]	4.40 (5) [59]	1.34 (1) [75]	4.70 (5) [67]	7.21 (7) [58]	
5	35	11	7	7	7	7	7	6.74 {0.66}	6.73 (7) [43]	4.70 (5) [91]	4.62 (5) [89]	5.00 (5) [100]	8.08 (8) [62]	
			11	9	7	5	3		6.33 {0.83}	6.28 (6) [45]	4.86 (5) [93]	2.53 (1) [36]	4.96 (5) [96]	8.09 (8) [57]
			8	15	3	3	3	3	5.72 {0.94}	5.69 (6) [41]	4.73 (5) [81]	1.18 (1) [83]	4.90 (5) [87]	9.96 (10) [49]
50	16	10	10	10	10	10	9.41 {1.04}	9.43 (9) [37]	4.72 (5) [92]	4.96 (5) [99]	5.00 (5) [100]	11.56 (11) [43]		
		16	13	10	7	4		8.73 {1.17}	8.70 (8) [29]	4.92 (5) [98]	3.18 (5) [44]	5.00 (5) [100]	11.36 (11) [49]	
		13	25	3	3	3	7.23 {1.15}	7.21 (7) [38]	4.74 (5) [80]	1.00 (1) [100]	4.83 (5) [83]	14.21 (14) [37]		

¹ Počet měření $j \in \{1, \dots, m\}$ takových, že δ_j je rovno k .

² Metoda, pomocí níž bylo z markovského řetězce délky $M = 20000$ odhadnuto K .

³ Metoda optimálních souřadnic (*optimal coordinates*).

⁴ Metoda faktoru akcelerace (*acceleration factor*).

⁵ Metoda paralelní analýzy (*parallel analysis*).

⁶ Kaiserovo pravidlo (*Kaiser rule*).

Tabulka 5.3: Třetí situace – obecný model faktorové analýzy. Odhady skutečného počtu faktorů K při různém počtu celkových měření m . Počet pozorování $n = 100$, počet opakování simulace 100, délka zážehové fáze $B = 20000$, délka markovského řetězce použitého pro inferenci $M = 20000$. Průměr přes všechna opakování {směrodatná odchylka} (nejčastější navrhovaná hodnota) [počet případů navržení nejčastější hodnoty ze 100 možných].

K	m	K^{\max}	Bayesovské odhady ¹		Klasické odhady			
			Průměr	Modus	OC ²	AF ³	P ⁴	Kaiser ⁵
	6	2	1.04 {0.05}	1.00 (1) [100]	1.03 (1) [97]	1.00 (1) [100]	1.03 (1) [97]	1.32 (1) [68]
2	12	4	2.83 {0.41}	2.84 (3) [81]	2.01 (2) [99]	1.58 (2) [58]	2.01 (2) [99]	2.85 (3) [81]
	24	8	4.90 {0.36}	4.93 (5) [82]	2.01 (2) [93]	2.00 (2) [100]	2.05 (2) [95]	6.38 (6) [49]
	15	5	4.05 {0.43}	4.08 (4) [76]	2.94 (3) [94]	1.97 (2) [53]	2.94 (3) [94]	3.00 (3) [100]
3	21	7	6.02 {0.40}	6.03 (6) [83]	2.92 (3) [96]	1.76 (1) [62]	3.00 (3) [100]	3.33 (3) [67]
	30	10	8.15 {0.68}	8.18 (8) [48]	2.94 (3) [98]	3.00 (3) [100]	3.00 (3) [100]	3.49 (3) [53]
	25	8	7.12 {0.39}	7.13 (7) [81]	4.43 (5) [50]	1.83 (1) [52]	4.51 (5) [52]	5.00 (5) [100]
5	35	11	10.38 {0.62}	10.38 (10) [49]	4.88 (5) [97]	4.68 (5) [91]	5.00 (5) [100]	5.02 (5) [98]
	50	16	13.60 {1.37}	13.57 (14) [32]	4.76 (5) [93]	4.97 (5) [99]	5.00 (5) [100]	5.46 (5) [54]

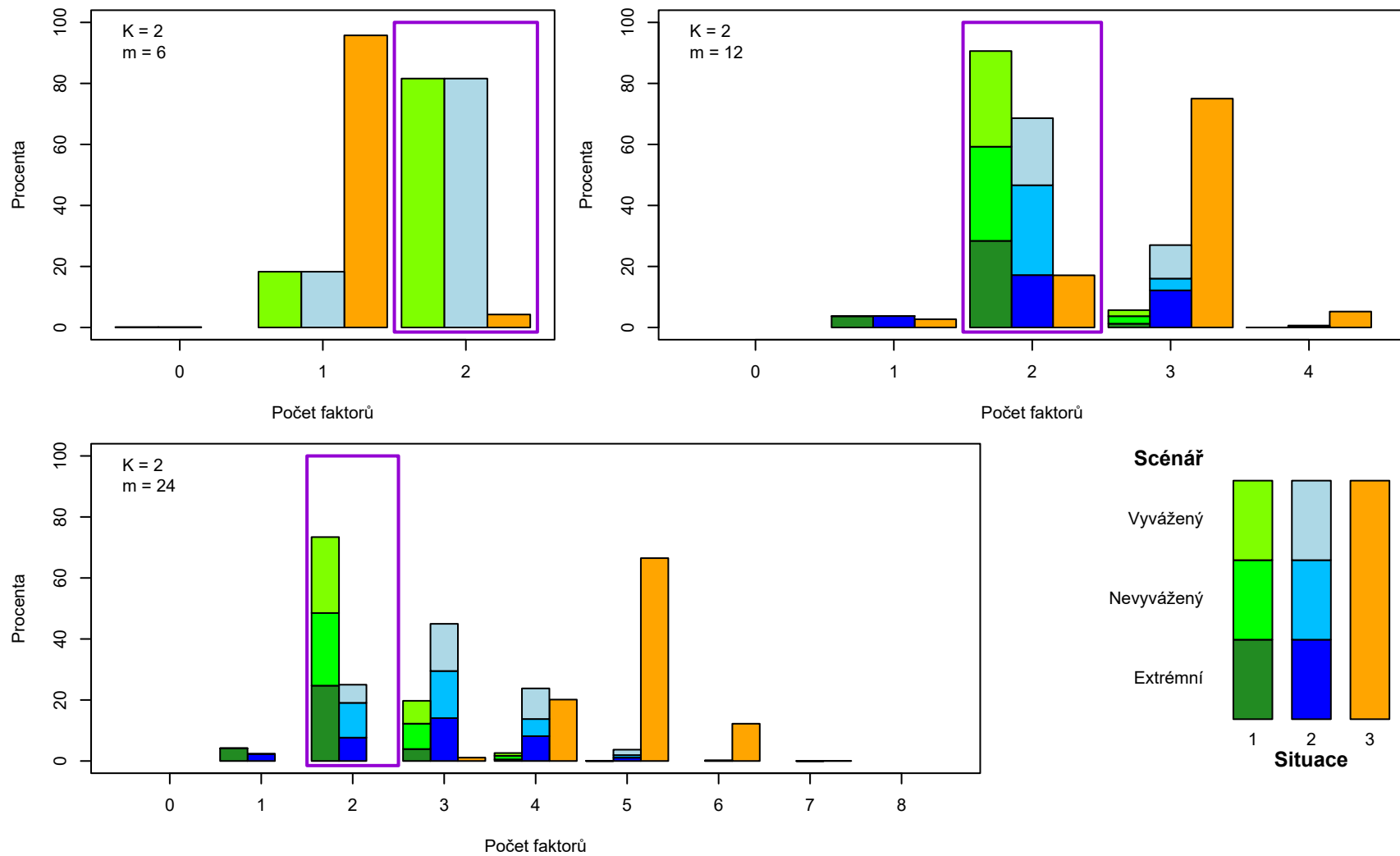
¹ Metoda, pomocí níž bylo z markovského řetězce délky $M = 20000$ odhadnuto K .

² Metoda optimálních souřadnic (*optimal coordinates*).

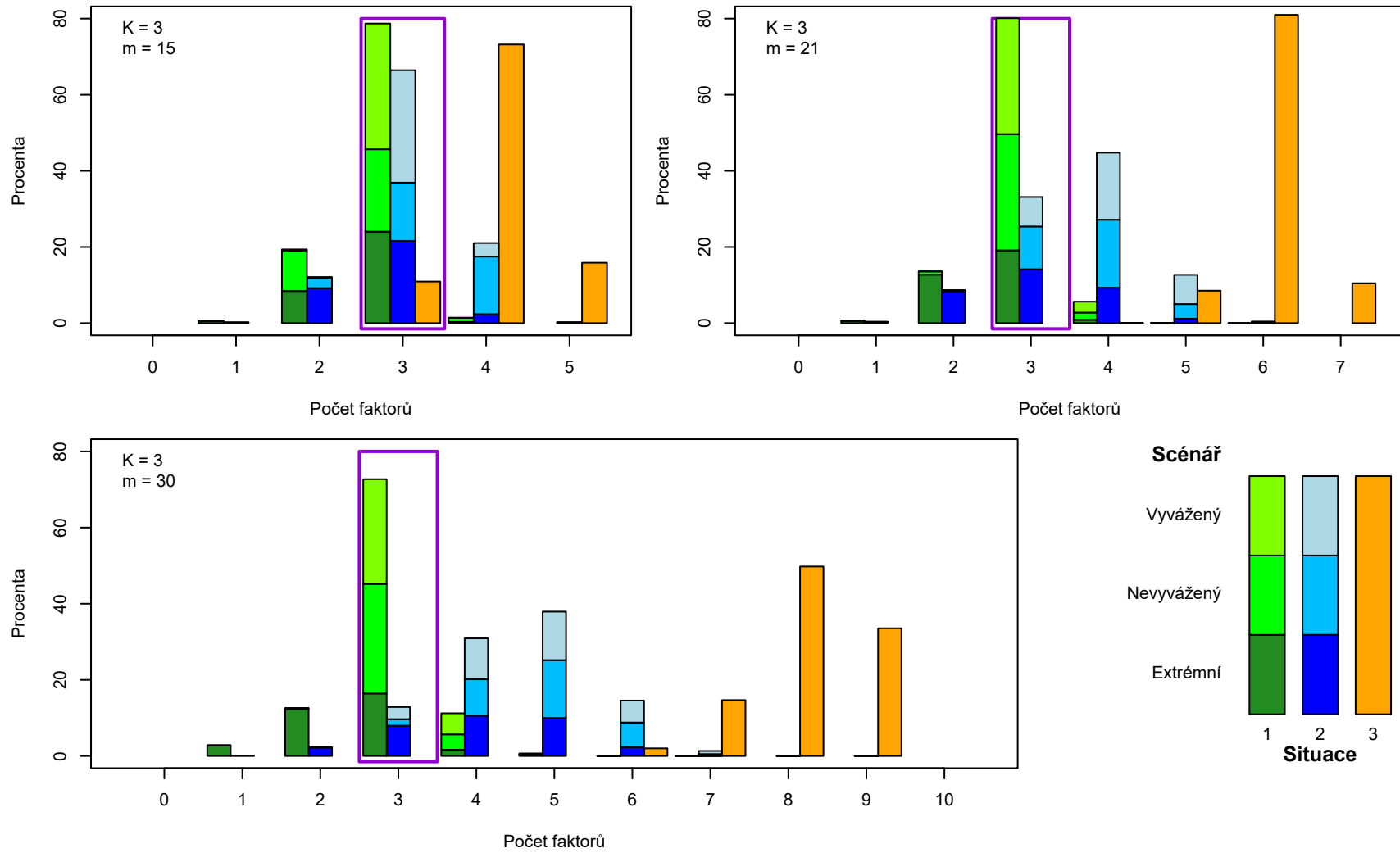
³ Metoda faktoru akcelerace (*acceleration factor*).

⁴ Metoda paralelní analýzy (*parallel analysis*).

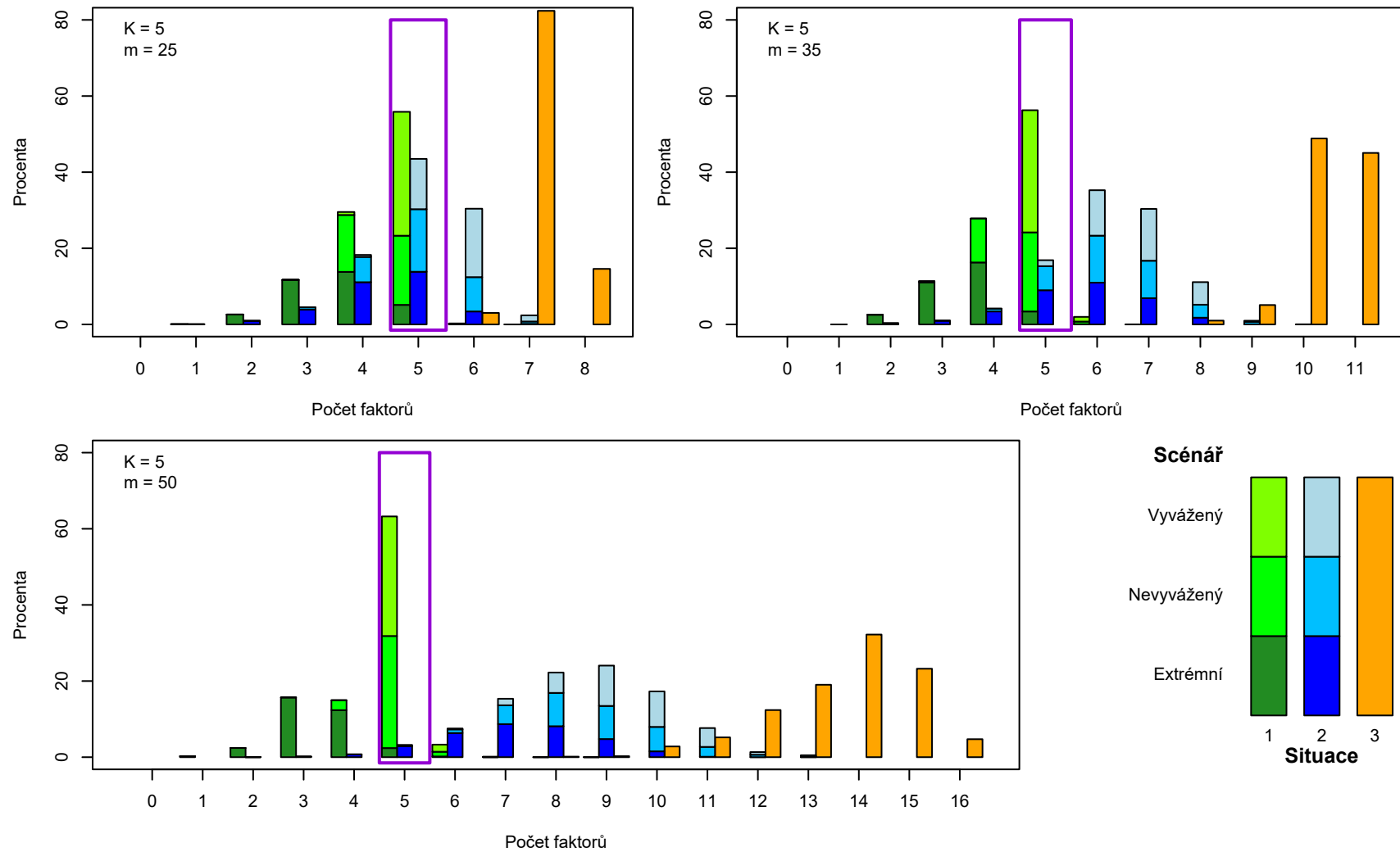
⁵ Kaiserovo pravidlo (*Kaiser rule*).



Obrázek 5.1: Rozložení počtu aktivních faktorů ve všech generovaných řetězcích za rozdílných situacích, je-li skutečný počet faktorů $K = 2$.



Obrázek 5.2: Rozložení počtu aktivních faktorů ve všech generovaných řetězcích za rozdílných situací, je-li skutečný počet faktorů $K = 3$.



Obrázek 5.3: Rozložení počtu aktivních faktorů ve všech generovaných řetězcích za rozdílných situacích, je-li skutečný počet faktorů $K = 5$.

Závěr

V modelu faktorové analýzy vzhledem k problému neznámého počtu faktorů či problému identifikovatelnosti existuje celá řada metod, které lze využít pro samotný odhad. Bayesovský pohled je jen další z možných přístupů mající své výhody i nevýhody.

Bylo zapotřebí si ukázat, jakým způsobem se v bayesovské statistice pracuje s neznámými parametry a jak získat jejich odhady. Vše spočívá v aposteriorním rozdělení, které vzniká složením věrohodnosti modelu a apriorního rozdělení parametrů. Za tímto účelem jsme si představili hustoty základních pravděpodobnostních rozdělení, abychom viděli, jaká rozdělení spolu konjugují a dají se tak podle Bayesovy věty dát dohromady s výslednou hustotou opět známého tvaru. Model mnohorozměrné regrese, lišící se od modelu faktorové analýzy ve znalosti regresorů (faktorů ve faktorové analýze), nám posloužil jako příkladná ukázka použití bayesovských metod. Řádně jsme odvodili marginální aposteriorní rozdělení parametrů tohoto modelu, na jehož základě je již jednoduché sestavit odhady.

Ovšem jakmile jsme přešli k faktorové analýze, tak jsme zjistili, že neznámý parametr odpovídající jednotlivým faktorům aposteriorní rozdělení velmi zkomplikoval. Ukázali jsme si, že přes veškerou snahu vyintegrovat neznámé parametry z aposteriorní hustoty, nedostáváme hustotu tvaru běžně používaných rozdělení. Dokázali jsme odvodit pouze tvar marginální aposteriorní hustoty rozdělení faktorů \mathbb{F} (tvrzení 7) či dvojice parametrů $(\mathbf{\Lambda}, \mathbf{\Sigma})$ (tvrzení 8). To je důvodem, proč jsme se museli spokojit s odvozením plně podmíněných rozdělení (tvrzení 6), která již díky našemu předpokladu vhodného apriorního rozdělení patří mezi běžně používaná. Často jsme využívali metody doplnění na čtverec, kterou zmiňujeme v příloženém appendixu.

Mohli jsme tak těchto výsledků použít k sestavení Gibbsova algoritmu založeného na Monte Carlo principu a teorie markovských řetězců. Tento algoritmus nám po dostatečném počtu iterací generuje přibližně z aposteriorního rozdělení. Nagenované hodnoty tak lze použít k odhadu marginálních aposteriorních rozdělení parametrů modelu. Jako alternativu k tomuto algoritmu jsme navíc představili ICM algoritmus, který hledá modus aposteriorního rozdělení užitím modů plně podmíněných rozdělení. Věnovali jsme několik stránek tomu, jaké by byly výsledky za předpokladu jiného apriorního rozdělení. Dokonce jsme si ukázali, jak se dá model faktorové analýzy zobecnit, a přitom lze aplikovat stále stejný postup a dostat obdobné výsledky.

Ani tyto modely však neřeší všechny problémy, které se mohou naskytnout. Ukázali jsme si tak moderní přístup ve faktorové analýze, který se snaží model stavět tak, aby řešil hned několik věcí najednou. Náhodným určením nejvýše jednoho faktoru (z předem určeného maximálního množství) ke každému měření dosahujeme nejen velmi řídké matice faktorových zátěží, ale zároveň i náhodného počtu aktivních faktorů. Zároveň jsme si zavedli prosté identifikační podmínky, díky nimž (až na permutaci faktorů a případné změny znaménka) dostaneme jednoznačnost parametrů. Po zavedení apriorního rozdělení dodržující hierarchický model jsme podobnými metodami jako ve třetí kapitole podrobně odvodili plně podmíněná rozdělení skupinek parametrů (tvrzení 17, 18 a 19). Znalosti těchto rozdělení jsme využili pro sestavení algoritmu, který velmi sofistikovaně od jistého

okamžiku generuje přibližně z aposteriorního rozdělení. Nejčastější počet aktivních faktorů potom lze považovat za odhad počtu skrytých faktorů. Díky řídkosti matice faktorových zátěží pak jsme navíc schopni tento model jednoduše interpretovat sloučením měření do skupin podle toho, s jakým faktorem jsou spojeny.

Z teoretického pohledu se tento model nádherně vypořádává s klasickými problémy modelu faktorové analýzy. Bohužel nám zde však vyvstávají jiné problémy spojené s bayesovským pohledem využívající MCMC metod. Předem totiž nevíme, od jakého okamžiku již dosáhneme limitního (aposteriorního) rozdělení a zda již můžeme generované parametry považovat za reprezentanty z tohoto rozdělení. Může být zapotřebí velmi dlouhý řetězec, což s sebou nese i mnohem delší výpočetní čas. V provedené simulační studii jsme ověřili, že za dodržení předpokladů modelu tato metoda dává poměrně přesný odhad počtu aktivních faktorů (tabulka 5.1). Ovšem za nedodržení předpokladů tohoto modelu, tj. za platnosti obecného modelu faktorové analýzy, kde každé měření je ovlivněno více faktory, již tato metoda selhává. Nezachytitelné vztahy jsou nahrazeny přidáním dalších faktorů, což vede k značně vyšším odhadům počtu faktorů, než jaký je skutečný počet aktivních faktorů (tabulka 5.3). Dokonce i v obecném faktorovém modelu, který má své dominantní faktorové zátěže a je tak svou strukturou velmi podobný modelu určujících faktorů, je odhad počtu aktivních faktorů vychýlený (tabulka 5.2). Je proto nutné tento model rozšířit o možnost toho, aby měření mohlo být ovlivněno i více faktory najednou. Tomuto nesnadnému zobecnění se nyní věnují autoři článku Conti a kol. (2014), z něhož bylo čerpáno.

Apendix

A Tvrzení z lineární algebry

Aby čtenář nemusel hledat v jiných zdrojích přesná znění tvrzení používaných v této práci, uvádíme je zde v přílohách. Jedná se o pojmy a tvrzení jednoduše dohledatelné v mnoha učebnicích lineární algebry.

Nejprve několik užitečných tvrzení pro práci s determinanty.

Tvrzení A1 (Sylvestrovo pravidlo). *Jsou-li $\mathbb{A} \in \mathbb{R}^{n \times m}$ a $\mathbb{B} \in \mathbb{R}^{m \times n}$ obecné obdélníkové matice a $\mathbb{X} \in \mathbb{R}^{n \times n}$ invertibilní čtvercová matice, pak*

$$i) |\mathbb{I}_n + \mathbb{A}\mathbb{B}| = |\mathbb{I}_m + \mathbb{B}\mathbb{A}|,$$

$$ii) |\mathbb{X} + \mathbb{A}\mathbb{B}| = |\mathbb{X}| \cdot |\mathbb{I}_m + \mathbb{B}\mathbb{X}^{-1}\mathbb{A}|.$$

Speciálně pro n -rozměrné vektory \mathbf{a} a \mathbf{b} platí $|\mathbb{I}_n + \mathbf{a}\mathbf{b}^\top| = 1 + \mathbf{b}^\top \mathbf{a}$.

Stopou čtvercové matice \mathbb{A} řádu n rozumíme součet jejích diagonálních prvků, tedy

$$\text{Tr } \mathbb{A} = \sum_{i=1}^n a_{ii}.$$

Jedná se o lineární funkcionál s velmi pěknou vlastností.

Tvrzení A2. *Budte $\mathbb{A} \in \mathbb{R}^{j \times k}$, $\mathbb{B} \in \mathbb{R}^{k \times l}$, $\mathbb{C} \in \mathbb{R}^{l \times m}$, $\mathbb{D} \in \mathbb{R}^{m \times n}$. Potom*

$$\text{Tr } \mathbb{A}\mathbb{B}\mathbb{C}\mathbb{D} = \text{Tr } \mathbb{B}\mathbb{C}\mathbb{D}\mathbb{A} = \text{Tr } \mathbb{C}\mathbb{D}\mathbb{A}\mathbb{B} = \text{Tr } \mathbb{D}\mathbb{A}\mathbb{B}\mathbb{C}.$$

Obecně platí, že stopa součinu matic je invariantní vůči cyklické permutaci jednotlivých činitelů.

Operace vektorizace matice je definována jako seřazení jednotlivých sloupců matice $\mathbb{A} = (\mathbf{a}_1, \dots, \mathbf{a}_n) \in \mathbb{R}^{m \times n}$ pod sebe v jeden nm -složkový vektor, tedy

$$\text{vec } \mathbb{A} = \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_n \end{pmatrix} = (\mathbf{a}_1^\top, \dots, \mathbf{a}_n^\top)^\top.$$

Transpozíčně-permutační matici $\mathbb{P}_{\text{vec } \mathbb{A}}$ matice \mathbb{A} definujeme jako matici s vlastností $\text{vec } (\mathbb{A}^\top) = \mathbb{P}_{\text{vec } \mathbb{A}} \text{vec } \mathbb{A}$.

Kroneckerův součin matic $\mathbb{A} \in \mathbb{R}^{m \times n}$ a $\mathbb{B} \in \mathbb{R}^{k \times l}$ je definován jako matice rozměrů $mk \times nl$ tvaru

$$\mathbb{A} \otimes \mathbb{B} = \begin{pmatrix} a_{11}\mathbb{B} & a_{12}\mathbb{B} & \cdots & a_{1n}\mathbb{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}\mathbb{B} & a_{m2}\mathbb{B} & \cdots & a_{mn}\mathbb{B} \end{pmatrix}.$$

Uvádíme zde základní pravidla pro počítání s Kroneckerovým součinem.

Tvrzení A3. Budte $\alpha \in \mathbb{R}$ a $\mathbb{A}, \mathbb{B}, \mathbb{C}$ a \mathbb{D} matice takových rozměrů, aby následující výrazy v jednotlivých řádcích dávaly smysl. Platí

- i) $\alpha (\mathbb{A} \otimes \mathbb{B}) = (\alpha \mathbb{A}) \otimes \mathbb{B} = \mathbb{A} \otimes (\alpha \mathbb{B}),$
- ii) $\mathbb{A} \otimes (\mathbb{B} \otimes \mathbb{C}) = (\mathbb{A} \otimes \mathbb{B}) \otimes \mathbb{C},$
- iii) $(\mathbb{A} \otimes \mathbb{B})^\top = \mathbb{A}^\top \otimes \mathbb{B}^\top,$
- iv) $(\mathbb{A} \otimes \mathbb{B})^{-1} = \mathbb{A}^{-1} \otimes \mathbb{B}^{-1},$
- v) $(\mathbb{A} \otimes \mathbb{B})(\mathbb{C} \otimes \mathbb{D}) = (\mathbb{A}\mathbb{C}) \otimes (\mathbb{B}\mathbb{D}),$
- vi) $(\mathbb{A} + \mathbb{B}) \otimes \mathbb{C} = \mathbb{A} \otimes \mathbb{C} + \mathbb{B} \otimes \mathbb{C},$ podobně $\mathbb{A} \otimes (\mathbb{B} + \mathbb{C}) = \mathbb{A} \otimes \mathbb{B} + \mathbb{A} \otimes \mathbb{C}.$

Nyní si uvedeme užitečná pravidla, která dávají dohromady vlastnosti determinantu, stopy matice, vektorizace matice a Kroneckerova součinu.

Tvrzení A4. Platí

- i) $|\mathbb{A} \otimes \mathbb{B}| = |\mathbb{A}|^n |\mathbb{B}|^m \quad \text{pro } \mathbb{A} \in \mathbb{R}^{m \times m}, \mathbb{B} \in \mathbb{R}^{n \times n},$
- ii) $\text{vec}(\mathbb{A}\mathbb{X}\mathbb{B}) = (\mathbb{B}^\top \otimes \mathbb{A}) \text{vec } \mathbb{X} \quad \text{pro } \mathbb{A} \in \mathbb{R}^{m \times n}, \mathbb{X} \in \mathbb{R}^{n \times k}, \mathbb{B} \in \mathbb{R}^{k \times l},$
- iii) $\text{Tr}(\mathbb{A} \otimes \mathbb{B}) = \text{Tr } \mathbb{A} \cdot \text{Tr } \mathbb{B} \quad \text{pro } \mathbb{A} \in \mathbb{R}^{m \times m}, \mathbb{B} \in \mathbb{R}^{n \times n},$
- iv) $\text{Tr}(\mathbb{A}^\top \mathbb{B}) = \text{Tr}(\mathbb{B}\mathbb{A}^\top) = (\text{vec } \mathbb{A})^\top \text{vec } \mathbb{B} \quad \text{pro } \mathbb{A}, \mathbb{B} \in \mathbb{R}^{m \times n},$
- v) $\text{Tr}(\mathbb{A}^\top \Sigma^{-1} \mathbb{A}) = (\text{vec } \mathbb{A})^\top [\mathbb{I}_n \otimes \Sigma^{-1}] (\text{vec } \mathbb{A}) \quad \text{pro } \mathbb{A} \in \mathbb{R}^{k \times n}, 0 < \Sigma \in \mathbb{R}^{k \times k}.$

Dále následují tvrzení o derivacích podle vektoru a podle (symetrické) matice.

Tvrzení A5. Platí

- i) $\frac{\partial \mathbf{a}^\top \mathbf{x}}{\partial \mathbf{x}} = \frac{\partial \mathbf{x}^\top \mathbf{a}}{\partial \mathbf{x}} = \mathbf{a} \quad \text{pro } \mathbf{a} \text{ a } \mathbf{x} \text{ } n\text{-rozměrné vektory},$
- ii) $\frac{\mathbf{x}^\top \mathbb{A} \mathbf{x}}{\partial \mathbf{x}} = (\mathbb{A} + \mathbb{A}^\top) \mathbf{x} \stackrel{\mathbb{A}=\mathbb{A}^\top}{=} 2\mathbb{A} \mathbf{x} \quad \text{pro } \mathbb{A} \in \mathbb{R}^{n \times n}, \mathbf{x} \in \mathbb{R}^n,$
- iii) $\frac{\partial |\mathbb{X}|}{\partial \mathbb{X}} = \begin{cases} \mathbb{X} & \text{pro } \mathbb{X} \in \mathbb{R}^{n \times n} \text{ s různými prvky,} \\ 2\mathbb{X} - \text{diag } \mathbb{X} & \text{pro } \mathbb{X} \in \mathbb{R}^{n \times n} \text{ symetrickou,} \end{cases}$
- iv) $\frac{\partial \log |\mathbb{X}|}{\partial \mathbb{X}} = \begin{cases} (\mathbb{X}^{-1})^\top & \text{pro } \mathbb{X} \in \mathbb{R}^{n \times n} \text{ s různými prvky, } |\mathbb{X}| > 0, \\ 2\mathbb{X}^{-1} - \text{diag}(\mathbb{X}^{-1}) & \text{pro } \mathbb{X} \in \mathbb{R}^{n \times n} \text{ symetrickou, } |\mathbb{X}| > 0, \end{cases}$
- v) $\frac{\partial \text{Tr } \mathbb{X}\mathbb{A}}{\partial \mathbb{X}} = \begin{cases} \mathbb{A}^\top & \text{pro } \mathbb{A}^\top, \mathbb{X} \in \mathbb{R}^{m \times n} \text{ s různými prvky,} \\ \mathbb{A} + \mathbb{A}^\top - \text{diag } \mathbb{A} & \text{pro } \mathbb{A}, \mathbb{X} \in \mathbb{R}^{n \times n}, \mathbb{X} = \mathbb{X}^\top. \end{cases}$
- vi) Bud $\mathbb{A} \in \mathbb{R}^{n \times n}$ symetrická a $\mathbb{X} \in \mathbb{R}^{n \times n}, |\mathbb{X}| > 0,$ potom

$$\frac{\partial \text{Tr } \mathbb{X}^{-1} \mathbb{A}}{\partial \mathbb{X}} = \begin{cases} -\mathbb{X}^{-1} \mathbb{A} \mathbb{X}^{-1} & \text{pro } \mathbb{X} \text{ s různými prvky,} \\ -[2\mathbb{X}^{-1} \mathbb{A} \mathbb{X}^{-1} - \text{diag}(\mathbb{X}^{-1} \mathbb{A} \mathbb{X}^{-1})] & \text{pro } \mathbb{X} \text{ symetrickou.} \end{cases}$$

vii) Pro symetrické matice $\mathbb{A} \in \mathbb{R}^{k \times k}$ a $\mathbb{B} \in \mathbb{R}^{n \times n}$ a obecné $\mathbb{X}, \mathbb{X}_0 \in \mathbb{R}^{n \times m}$ platí

$$\frac{\partial}{\partial \mathbb{X}} \text{Tr} [\mathbb{A} (\mathbb{X} - \mathbb{X}_0)^\top \mathbb{B} (\mathbb{X} - \mathbb{X}_0)] = 2\mathbb{B} (\mathbb{X} - \mathbb{X}_0) \mathbb{A}.$$

B Doplnění na čtverec

Při odvozování aposteriorního rozdělení v bayesovské statistice velice často potřebujeme umět sloučit součin dvou (i více) hustot v jedinou funkci a umět ji opět vyjádřit v jednom společném tvaru. Při předpokladu normálního rozdělení musíme často v exponentu sloučit dva (i více) kvadratických útvarů. Základem je jednorozměrný případ, kdy potřebujeme výraz tvaru

$$c(x - a)^2 + d(x - b)^2 \text{ převést na tvar } \tilde{c}(x - \mu)^2 + \tilde{d}.$$

Postup je takový, že se zaměříme na kvadratický člen, koeficient u něj vytkneme a *doplníme na čtverec* tak, aby vymizel i lineární člen, zbytek výrazu by již žádné x v sobě neměl obsahovat. Platí tedy

$$\tilde{c} = c + d, \quad \mu = \frac{ac + bd}{c + d}, \quad \tilde{d} = ca^2 + db^2 - (c + d)\mu^2 = \frac{cd}{c + d}(a - b)^2.$$

Toto však nelze provést, pokud $c + d = 0$. Potom totiž kvadratický člen úplně zmizí. Na μ lze pohlížet jako na vážený součet čísel a a b , váha je určena oněmi koeficienty c a d . Tento proces si zobecníme pro vektory a matice.

Tvrzení B6. *Budte $\mathbb{X}, \mathbb{A}, \mathbb{B}$ obdélníkové matice stejného typu $n \times k$ a \mathbb{C}, \mathbb{D} budte symetrické pozitivně definitní matice stejného řádu n . Potom*

$$\begin{aligned} & (\mathbb{X} - \mathbb{A})^\top \mathbb{C} (\mathbb{X} - \mathbb{A}) + (\mathbb{X} - \mathbb{B})^\top \mathbb{D} (\mathbb{X} - \mathbb{B}) = \\ & = (\mathbb{X} - \mathbb{M})^\top [\mathbb{C} + \mathbb{D}] (\mathbb{X} - \mathbb{M}) + \mathbb{A}^\top \mathbb{C} \mathbb{A} + \mathbb{B}^\top \mathbb{D} \mathbb{B} - \mathbb{M}^\top [\mathbb{C} + \mathbb{D}] \mathbb{M}, \end{aligned} \quad (5.1)$$

kde $\mathbb{M} = [\mathbb{C} + \mathbb{D}]^{-1} (\mathbb{C} \mathbb{A} + \mathbb{D} \mathbb{B})$.

Důkaz. Provedeme jen několik algebraických úprav, kdy budeme mít na paměti, že dle našeho předpokladu je matice $\mathbb{C} + \mathbb{D}$ také symetrická pozitivně definitní, a tedy také invertovatelná. Ve znění tvrzení by právě stačilo předpokládat pouze invertibilitu matice $\mathbb{C} + \mathbb{D}$, a to za účelem platnosti následujících úprav

$$\begin{aligned} & (\mathbb{X} - \mathbb{A})^\top \mathbb{C} (\mathbb{X} - \mathbb{A}) + (\mathbb{X} - \mathbb{B})^\top \mathbb{D} (\mathbb{X} - \mathbb{B}) = \\ & = \mathbb{X}^\top [\mathbb{C} + \mathbb{D}] \mathbb{X} - \mathbb{X}^\top (\mathbb{C} \mathbb{A} + \mathbb{D} \mathbb{B}) - (\mathbb{A}^\top \mathbb{C} + \mathbb{B}^\top \mathbb{D}) \mathbb{X} + \mathbb{A}^\top \mathbb{C} \mathbb{A} + \mathbb{B}^\top \mathbb{D} \mathbb{B} \\ & = \mathbb{X}^\top [\mathbb{C} + \mathbb{D}] (\mathbb{X} - [\mathbb{C} + \mathbb{D}]^{-1} (\mathbb{C} \mathbb{A} + \mathbb{D} \mathbb{B})) - (\mathbb{C} \mathbb{A} + \mathbb{D} \mathbb{B})^\top \mathbb{X} + \mathbb{A}^\top \mathbb{C} \mathbb{A} + \mathbb{B}^\top \mathbb{D} \mathbb{B} \\ & = (\mathbb{X} - \mathbb{M} + \mathbb{M})^\top [\mathbb{C} + \mathbb{D}] (\mathbb{X} - \mathbb{M}) - (\mathbb{C} \mathbb{A} + \mathbb{D} \mathbb{B})^\top \mathbb{X} + \mathbb{A}^\top \mathbb{C} \mathbb{A} + \mathbb{B}^\top \mathbb{D} \mathbb{B} \\ & = (\mathbb{X} - \mathbb{M})^\top [\mathbb{C} + \mathbb{D}] (\mathbb{X} - \mathbb{M}) + \mathbb{A}^\top \mathbb{C} \mathbb{A} + \mathbb{B}^\top \mathbb{D} \mathbb{B} - \mathbb{M}^\top [\mathbb{C} + \mathbb{D}] \mathbb{M}, \end{aligned}$$

kde jsme využili toho, že platí $\mathbb{M}^\top [\mathbb{C} + \mathbb{D}] = (\mathbb{C} \mathbb{A} + \mathbb{D} \mathbb{B})^\top$. □

Důsledek (Tvrzení B6 pro $k = 1$). Budte $\mathbf{x}, \mathbf{a}, \mathbf{b}$ vektory délky n a \mathbb{C}, \mathbb{D} budte symetrické pozitivně definitní matice stejného řádu n . Potom

$$\begin{aligned} & (\mathbf{x} - \mathbf{a})^\top \mathbb{C} (\mathbf{x} - \mathbf{a}) + (\mathbf{x} - \mathbf{b})^\top \mathbb{D} (\mathbf{x} - \mathbf{b}) = \\ & = (\mathbf{x} - \boldsymbol{\mu})^\top [\mathbb{C} + \mathbb{D}] (\mathbf{x} - \boldsymbol{\mu}) + \mathbf{a}^\top \mathbb{C} \mathbf{a} + \mathbf{b}^\top \mathbb{D} \mathbf{b} - \boldsymbol{\mu}^\top [\mathbb{C} + \mathbb{D}] \boldsymbol{\mu}, \end{aligned} \quad (5.2)$$

kde $\boldsymbol{\mu} = [\mathbb{C} + \mathbb{D}]^{-1} (\mathbb{C} \mathbf{a} + \mathbb{D} \mathbf{b})$.

Často je však cílový parametr spjatý ještě s nějakou jinou maticí. Doplnění na čtverec v takovéto situaci nám pomůže vyřešit následující tvrzení.

Tvrzení B7. *Budte $\mathbb{X}, \mathbb{B} \in \mathbb{R}^{n \times k}$, $\mathbb{A} \in \mathbb{R}^{n \times m}$, $\mathbb{C} \in \mathbb{R}^{k \times m}$, $0 < \mathbb{D} \in \mathbb{R}^{k \times k}$ a $0 < \Psi \in \mathbb{R}^{m \times m}$. Potom*

$$\begin{aligned} & (\mathbb{X}\mathbb{C} - \mathbb{A}) \Psi (\mathbb{X}\mathbb{C} - \mathbb{A})^\top + (\mathbb{X} - \mathbb{B}) \mathbb{D} (\mathbb{X} - \mathbb{B})^\top = \\ & = (\mathbb{X} - \mathbb{M}) [\mathbb{C}\Psi\mathbb{C}^\top + \mathbb{D}] (\mathbb{X} - \mathbb{M})^\top - \mathbb{M} [\mathbb{C}\Psi\mathbb{C}^\top + \mathbb{D}] \mathbb{M}^\top + \mathbb{A}\Psi\mathbb{A}^\top + \mathbb{B}\mathbb{D}\mathbb{B}^\top \\ & \quad \text{kde } \mathbb{M} = (\mathbb{A}\Psi\mathbb{C}^\top + \mathbb{B}\mathbb{D}) [\mathbb{C}\Psi\mathbb{C}^\top + \mathbb{D}]^{-1}. \end{aligned}$$

Důkaz. Jelikož z předpokladu $\Psi > 0$ plyne, že $\mathbb{C}\Psi\mathbb{C}^\top \geq 0$, a $\mathbb{D} > 0$, tak matice $\mathbb{C}\Psi\mathbb{C}^\top + \mathbb{D}$ je symetrická pozitivně definitní, a tedy také invertovatelná. Pro platnost tohoto tvrzení však stačí rovnou předpokládat, že je tato matice invertovatelná. Platí

$$\begin{aligned} & (\mathbb{X}\mathbb{C} - \mathbb{A}) \Psi (\mathbb{X}\mathbb{C} - \mathbb{A})^\top + (\mathbb{X} - \mathbb{B}) \mathbb{D} (\mathbb{X} - \mathbb{B})^\top = \\ & = \mathbb{X} [\mathbb{C}\Psi\mathbb{C}^\top + \mathbb{D}] \mathbb{X}^\top - (\mathbb{A}\Psi\mathbb{C}^\top + \mathbb{B}\mathbb{D}) \mathbb{X}^\top - \mathbb{X} (\mathbb{C}\Psi\mathbb{A}^\top + \mathbb{D}\mathbb{B}^\top) + \mathbb{A}\Psi\mathbb{A}^\top + \mathbb{B}\mathbb{D}\mathbb{B}^\top \\ & = \left(\mathbb{X} - (\mathbb{A}\Psi\mathbb{C}^\top + \mathbb{B}\mathbb{D}) [\mathbb{C}\Psi\mathbb{C}^\top + \mathbb{D}]^{-1} \right) [\mathbb{C}\Psi\mathbb{C}^\top + \mathbb{D}] \mathbb{X}^\top - \\ & \quad - \mathbb{X} (\mathbb{C}\Psi\mathbb{A}^\top + \mathbb{D}\mathbb{B}^\top) + \mathbb{A}\Psi\mathbb{A}^\top + \mathbb{B}\mathbb{D}\mathbb{B}^\top \\ & = (\mathbb{X} - \mathbb{M}) [\mathbb{C}\Psi\mathbb{C}^\top + \mathbb{D}] (\mathbb{X} - \mathbb{M} + \mathbb{M})^\top - \mathbb{X} (\mathbb{C}\Psi\mathbb{A}^\top + \mathbb{D}\mathbb{B}^\top) + \mathbb{A}\Psi\mathbb{A}^\top + \mathbb{B}\mathbb{D}\mathbb{B}^\top \\ & = (\mathbb{X} - \mathbb{M}) [\mathbb{C}\Psi\mathbb{C}^\top + \mathbb{D}] (\mathbb{X} - \mathbb{M})^\top - \mathbb{M} [\mathbb{C}\Psi\mathbb{C}^\top + \mathbb{D}] \mathbb{M}^\top + \mathbb{A}\Psi\mathbb{A}^\top + \mathbb{B}\mathbb{D}\mathbb{B}^\top, \end{aligned}$$

kde jsme využili toho, že platí $[\mathbb{C}\Psi\mathbb{C}^\top + \mathbb{D}] \mathbb{M}^\top = (\mathbb{C}\Psi\mathbb{A}^\top + \mathbb{D}\mathbb{B}^\top)$. □

Důsledek. Budte $\mathbb{Y} \in \mathbb{R}^{n \times m}$, $\mathbb{X} \in \mathbb{R}^{n \times k}$, $\mathbb{B}, \mathbb{B}_0 \in \mathbb{R}^{m \times k}$ a $0 < \mathbb{D} \in \mathbb{R}^{k \times k}$. Nechť \mathbb{X} má plnou sloupcovou hodnotu, potom

$$\begin{aligned} & (\mathbb{Y} - \mathbb{X}\mathbb{B}^\top)^\top (\mathbb{Y} - \mathbb{X}\mathbb{B}^\top) + (\mathbb{B} - \mathbb{B}_0) \mathbb{D}^{-1} (\mathbb{B} - \mathbb{B}_0)^\top = \\ & = (\mathbb{B} - \tilde{\mathbb{B}}) [\mathbb{X}^\top \mathbb{X} + \mathbb{D}^{-1}] (\mathbb{B} - \tilde{\mathbb{B}})^\top - \tilde{\mathbb{B}} [\mathbb{X}^\top \mathbb{X} + \mathbb{D}^{-1}] \tilde{\mathbb{B}}^\top + \mathbb{B}_0 \mathbb{D}^{-1} \mathbb{B}_0^\top + \mathbb{Y}^\top \mathbb{Y}, \end{aligned}$$

kde $\tilde{\mathbb{B}} = (\mathbb{Y}^\top \mathbb{X} + \mathbb{B}_0 \mathbb{D}^{-1}) [\mathbb{X}^\top \mathbb{X} + \mathbb{D}^{-1}]^{-1} = (\hat{\mathbb{B}} \mathbb{X}^\top \mathbb{X} + \mathbb{B}_0 \mathbb{D}^{-1}) [\mathbb{X}^\top \mathbb{X} + \mathbb{D}^{-1}]^{-1}$
a $\hat{\mathbb{B}} = \mathbb{Y}^\top \mathbb{X} (\mathbb{X}^\top \mathbb{X})^{-1}$.

Důkaz. Plyne z předchozího tvrzení B7 volbou $\mathbb{A}^* = \mathbb{Y}^\top$, $\mathbb{B}^* = \mathbb{B}_0$, $\mathbb{C}^* = \mathbb{X}^\top$, $\mathbb{D}^* = \mathbb{D}^{-1}$, $\Psi^* = \mathbb{I}_k$, $\mathbb{X}^* = \mathbb{B}$ a $\mathbb{M}^* = \tilde{\mathbb{B}}$. Díky předpokladu plné sloupcové hodnoty matice \mathbb{X} je $\mathbb{X}^\top \mathbb{X}$ invertibilní, a tedy lze definovat pomocnou matici $\hat{\mathbb{B}}^\top = (\mathbb{X}^\top \mathbb{X})^{-1} \mathbb{X}^\top \mathbb{Y}$. Navíc lze také rozložit první část výrazu z tvrzení na dva výrazy, z nichž jeden již neobsahuje \mathbb{B} , ale pouze $\hat{\mathbb{B}}$:

$$\begin{aligned} & (\mathbb{Y} - \mathbb{X}\mathbb{B}^\top)^\top (\mathbb{Y} - \mathbb{X}\mathbb{B}^\top) = (\mathbb{Y} \mp \mathbb{X}\hat{\mathbb{B}}^\top - \mathbb{X}\mathbb{B}^\top)^\top (\mathbb{Y} \mp \mathbb{X}\hat{\mathbb{B}}^\top - \mathbb{X}\mathbb{B}^\top) \\ & = (\mathbb{Y} - \mathbb{X}\hat{\mathbb{B}}^\top)^\top (\mathbb{Y} - \mathbb{X}\hat{\mathbb{B}}^\top) + (\hat{\mathbb{B}} - \mathbb{B}) \mathbb{X}^\top \mathbb{X} (\hat{\mathbb{B}} - \mathbb{B})^\top, \end{aligned}$$

kde jsme využili, že $(\mathbb{Y} - \mathbb{X}\hat{\mathbb{B}}^\top)^\top \mathbb{X} (\hat{\mathbb{B}} - \mathbb{B})^\top = \mathbb{O}_{k \times k}$. □

Další tvrzení nám pomůže vypořádat se s případem, kdy máme dvojce sady matic, které chceme opět doplnit na čtverec.

Tvrzení B8. *Budte $\mathbb{Y} \in \mathbb{R}^{n \times k}$, $\mathbb{X} \in \mathbb{R}^{n \times m}$, $\mathbb{B}, \mathbb{B}_0 \in \mathbb{R}^{k \times m}$, $\mathbb{Z} \in \mathbb{R}^{n \times l}$, $\mathbb{C}, \mathbb{C}_0 \in \mathbb{R}^{k \times l}$, $0 < \Phi \in \mathbb{R}^{m \times m}$ a $0 < \Psi \in \mathbb{R}^{l \times l}$. Potom*

$$\begin{aligned} & (\mathbb{Y} - \mathbb{X}\mathbb{B}^\top - \mathbb{Z}\mathbb{C}^\top)^\top (\mathbb{Y} - \mathbb{X}\mathbb{B}^\top - \mathbb{Z}\mathbb{C}^\top) + \\ & \quad + (\mathbb{B} - \mathbb{B}_0) \Phi^{-1} (\mathbb{B} - \mathbb{B}_0)^\top + (\mathbb{C} - \mathbb{C}_0) \Psi^{-1} (\mathbb{C} - \mathbb{C}_0)^\top = \\ & = \mathbb{Y}^\top \mathbb{Y} + (\mathbb{B} - \tilde{\mathbb{B}}(\mathbb{C})) [\mathbb{X}^\top \mathbb{X} + \Phi^{-1}] (\mathbb{B} - \tilde{\mathbb{B}}(\mathbb{C}))^\top + \mathbb{B}_0 \Phi^{-1} \mathbb{B}_0^\top + \\ & + (\mathbb{C} - \tilde{\mathbb{C}}) [\mathbb{Z}^\top \mathbb{A}_\Phi \mathbb{Z} + \Psi^{-1}] (\mathbb{C} - \tilde{\mathbb{C}})^\top + \mathbb{C}_0 \Psi^{-1} \mathbb{C}_0^\top - \tilde{\mathbb{C}} [\mathbb{Z}^\top \mathbb{A}_\Phi \mathbb{Z} + \Psi^{-1}] \tilde{\mathbb{C}}^\top - \\ & \quad - (\mathbb{Y}^\top \mathbb{X} - \mathbb{B}_0 \Phi^{-1}) [\mathbb{X}^\top \mathbb{X} + \Phi^{-1}]^{-1} (\mathbb{Y}^\top \mathbb{X} - \mathbb{B}_0 \Phi^{-1})^\top, \end{aligned}$$

kde

$$\begin{aligned} \tilde{\mathbb{B}}(\mathbb{C}) &= \left((\mathbb{Y} - \mathbb{Z}\mathbb{C}^\top)^\top \mathbb{X} + \mathbb{B}_0 \Phi^{-1} \right) [\mathbb{X}^\top \mathbb{X} + \Phi^{-1}]^{-1}, \\ \tilde{\mathbb{C}} &= \left[\mathbb{Y}^\top \mathbb{A}_\Phi \mathbb{Z} + \mathbb{C}_0 \Psi^{-1} + \mathbb{B}_0 \Phi^{-1} (\mathbb{X}^\top \mathbb{X} + \Phi^{-1})^{-1} \mathbb{X}^\top \mathbb{Z} \right] \cdot [\mathbb{Z}^\top \mathbb{A}_\Phi \mathbb{Z} + \Psi^{-1}]^{-1}, \\ \mathbb{A}_\Phi &= \mathbb{I}_n - \mathbb{X} [\mathbb{X}^\top \mathbb{X} + \Phi^{-1}]^{-1} \mathbb{X}^\top. \end{aligned}$$

Důkaz. Začneme tím, že se na tento výraz podíváme jako na funkci v \mathbb{B} a doplníme na čtverec. Zavedme si pomocnou matici $\mathbb{Y}_\mathbb{C}^* = \mathbb{Y} - \mathbb{Z}\mathbb{C}^\top$, pak lze použít důsledku předchozího tvrzení B7 pro $\mathbb{Y}_\mathbb{C}^*$ namísto \mathbb{Y} a obdržet tak

$$\begin{aligned} & (\mathbb{Y}_\mathbb{C}^* - \mathbb{X}\mathbb{B}^\top)^\top (\mathbb{Y}_\mathbb{C}^* - \mathbb{X}\mathbb{B}^\top) + (\mathbb{B} - \mathbb{B}_0) \Phi^{-1} (\mathbb{B} - \mathbb{B}_0)^\top = \mathbb{B}_0 \Phi^{-1} \mathbb{B}_0^\top + (\mathbb{Y}_\mathbb{C}^*)^\top \mathbb{Y}_\mathbb{C}^* + \\ & \quad + (\mathbb{B} - \tilde{\mathbb{B}}(\mathbb{C})) [\mathbb{X}^\top \mathbb{X} + \Phi^{-1}] (\mathbb{B} - \tilde{\mathbb{B}}(\mathbb{C}))^\top - \tilde{\mathbb{B}}(\mathbb{C}) [\mathbb{X}^\top \mathbb{X} + \Phi^{-1}] \tilde{\mathbb{B}}(\mathbb{C})^\top, \\ & \quad \text{kde } \tilde{\mathbb{B}}(\mathbb{C}) = \left((\mathbb{Y}_\mathbb{C}^*)^\top \mathbb{X} + \mathbb{B}_0 \Phi^{-1} \right) [\mathbb{X}^\top \mathbb{X} + \Phi^{-1}]^{-1}. \end{aligned}$$

Upravme

$$\begin{aligned} & \tilde{\mathbb{B}}(\mathbb{C}) [\mathbb{X}^\top \mathbb{X} + \Phi^{-1}] \tilde{\mathbb{B}}(\mathbb{C})^\top = \\ & = \left((\mathbb{Y} - \mathbb{Z}\mathbb{C}^\top)^\top \mathbb{X} + \mathbb{B}_0 \Phi^{-1} \right) [\mathbb{X}^\top \mathbb{X} + \Phi^{-1}]^{-1} \left((\mathbb{Y} - \mathbb{Z}\mathbb{C}^\top)^\top \mathbb{X} + \mathbb{B}_0 \Phi^{-1} \right)^\top \\ & = (\mathbb{C}\mathbb{Z}^\top \mathbb{X} - (\mathbb{Y}^\top \mathbb{X} - \mathbb{B}_0 \Phi^{-1})) [\mathbb{X}^\top \mathbb{X} + \Phi^{-1}]^{-1} (\mathbb{C}\mathbb{Z}^\top \mathbb{X} - (\mathbb{Y}^\top \mathbb{X} - \mathbb{B}_0 \Phi^{-1}))^\top. \end{aligned}$$

Vidíme, že zde \mathbb{C} vystupuje opět v kvadratické formě, a proto bude možné opět doplnit na čtverec. Matice \mathbb{C} se dále vyskytuje ještě ve dvou výrazech. Doplníme na čtverec tedy nejprve tyto dva sčítance

$$(\mathbb{Y} - \mathbb{Z}\mathbb{C}^\top)^\top (\mathbb{Y} - \mathbb{Z}\mathbb{C}^\top) + (\mathbb{C} - \mathbb{C}_0) \Psi^{-1} (\mathbb{C} - \mathbb{C}_0)^\top.$$

Na tento součet můžeme opět aplikovat důsledek tvrzení B7, kde za k , \mathbb{B} , \mathbb{B}_0 , \mathbb{X} a \mathbb{D} budeme považovat l , \mathbb{C} , \mathbb{C}_0 , \mathbb{Z} a Ψ . Dostaneme tak, že

$$\begin{aligned} & (\mathbb{Y} - \mathbb{Z}\mathbb{C}^\top)^\top (\mathbb{Y} - \mathbb{Z}\mathbb{C}^\top) + (\mathbb{C} - \mathbb{C}_0) \Psi^{-1} (\mathbb{C} - \mathbb{C}_0)^\top = \\ & = (\mathbb{C} - \tilde{\mathbb{C}}) [\mathbb{Z}^\top \mathbb{Z} + \Psi^{-1}] (\mathbb{C} - \tilde{\mathbb{C}})^\top - \tilde{\mathbb{C}} [\mathbb{Z}^\top \mathbb{Z} + \Psi^{-1}] \tilde{\mathbb{C}}^\top + \mathbb{C}_0 \Psi^{-1} \mathbb{C}_0^\top + \mathbb{Y}^\top \mathbb{Y}, \end{aligned}$$

$$\text{kde } \tilde{\mathbf{C}} = \left(\mathbf{Y}^\top \mathbf{Z} + \mathbf{C}_0 \boldsymbol{\Psi}^{-1} \right) \left[\mathbf{Z}^\top \mathbf{Z} + \boldsymbol{\Psi}^{-1} \right]^{-1}.$$

Zbývá sečíst zbylé dva výrazy, kde vystupuje \mathbf{C}

$$\begin{aligned} & - \left(\mathbf{CZ}^\top \mathbf{X} - \left(\mathbf{Y}^\top \mathbf{X} - \mathbb{B}_0 \boldsymbol{\Phi}^{-1} \right) \right) \left[\mathbf{X}^\top \mathbf{X} + \boldsymbol{\Phi}^{-1} \right]^{-1} \left(\mathbf{CZ}^\top \mathbf{X} - \left(\mathbf{Y}^\top \mathbf{X} - \mathbb{B}_0 \boldsymbol{\Phi}^{-1} \right) \right)^\top + \\ & \quad + \left(\mathbf{C} - \tilde{\mathbf{C}} \right) \left[\mathbf{Z}^\top \mathbf{Z} + \boldsymbol{\Psi}^{-1} \right] \left(\mathbf{C} - \tilde{\mathbf{C}} \right)^\top. \end{aligned}$$

Použijme tvrzení B7, kde klíčovou proměnnou je $\mathbf{X}^* = \mathbf{C}$ a zbytek proměnných je dáno vztahy

$$\begin{aligned} \mathbf{C}^* &= \mathbf{Z}^\top \mathbf{X}, \quad \mathbf{A}^* = \mathbf{Y}^\top \mathbf{X} - \mathbb{B}_0 \boldsymbol{\Phi}^{-1}, \quad \boldsymbol{\Psi}^* = - \left[\mathbf{X}^\top \mathbf{X} + \boldsymbol{\Phi}^{-1} \right]^{-1}, \\ \mathbf{B}^* &= \tilde{\mathbf{C}}, \quad \mathbf{D}^* = \mathbf{Z}^\top \mathbf{Z} + \boldsymbol{\Psi}^{-1}. \end{aligned}$$

Problémem je zde ovšem fakt, že takto definované $\boldsymbol{\Psi}^*$ nesplňuje předpoklad $\boldsymbol{\Psi}^* > 0$ ve znění tvrzení B7. Ovšem, jak je v samotném důkazu tohoto tvrzení napsáno, stačí zaručit, že matice

$$\mathbf{C} \boldsymbol{\Psi}^* \mathbf{C}^\top + \mathbf{D} = \mathbf{Z}^\top \left[\mathbb{I}_n - \mathbf{X} \left(\mathbf{X}^\top \mathbf{X} + \boldsymbol{\Phi}^{-1} \right)^{-1} \mathbf{X}^\top \right] \mathbf{Z} + \boldsymbol{\Psi}^{-1}$$

je invertibilní. My ukážeme, že za našich předpokladů je pozitivně definitní. Jelikož i pro obecně invertovatelné matice \mathbf{U} a \mathbf{V} platí, že pokud $\mathbf{V} \geq \mathbf{U} > 0$, musí $\mathbf{U}^{-1} \geq \mathbf{V}^{-1} > 0$, tak i $\mathbf{X}^\top \mathbf{X} + \boldsymbol{\Phi}^{-1} \geq \mathbf{X}^\top \mathbf{X}$ musí implikovat, že

$$\mathbf{A}_\Phi := \mathbb{I}_n - \mathbf{X} \left(\mathbf{X}^\top \mathbf{X} + \boldsymbol{\Phi}^{-1} \right)^{-1} \mathbf{X}^\top \geq \mathbb{I}_n - \mathbf{X} \left(\mathbf{X}^\top \mathbf{X} \right)^{-1} \mathbf{X}^\top =: \mathbf{A}_n.$$

Matice \mathbf{A}_n je symetrická a idempotentní, z čehož plyne, že je také pozitivně semidefinitní, tj. $\mathbf{A}_n = \mathbf{A}_n^\top \mathbf{A}_n \geq 0$. Proto i $\mathbf{A}_\Phi \geq 0$, tedy i $\mathbf{Z}^\top \mathbf{A}_\Phi \mathbf{Z} \geq 0$. Jelikož je dle našeho předpokladu $\boldsymbol{\Psi}^{-1} > 0$, tak musí i součet $\mathbf{Z}^\top \mathbf{A}_\Phi \mathbf{Z} + \boldsymbol{\Psi}^{-1}$ být pozitivně definitní matice a tedy invertovatelná.

Tvrzení B7 nám tedy dává, že

$$\begin{aligned} & - \left(\mathbf{CZ}^\top \mathbf{X} - \left(\mathbf{Y}^\top \mathbf{X} - \mathbb{B}_0 \boldsymbol{\Phi}^{-1} \right) \right) \left[\mathbf{X}^\top \mathbf{X} + \boldsymbol{\Phi}^{-1} \right]^{-1} \left(\mathbf{CZ}^\top \mathbf{X} - \left(\mathbf{Y}^\top \mathbf{X} - \mathbb{B}_0 \boldsymbol{\Phi}^{-1} \right) \right)^\top + \\ & \quad + \left(\mathbf{C} - \tilde{\mathbf{C}} \right) \left[\mathbf{Z}^\top \mathbf{Z} + \boldsymbol{\Psi}^{-1} \right] \left(\mathbf{C} - \tilde{\mathbf{C}} \right)^\top = \\ & = \left(\mathbf{C} - \tilde{\mathbf{C}} \right) \left[\mathbf{Z}^\top \left[\mathbb{I}_n - \mathbf{X} \left(\mathbf{X}^\top \mathbf{X} + \boldsymbol{\Phi}^{-1} \right)^{-1} \mathbf{X}^\top \right] \mathbf{Z} + \boldsymbol{\Psi}^{-1} \right] \left(\mathbf{C} - \tilde{\mathbf{C}} \right)^\top - \\ & \quad - \tilde{\mathbf{C}} \left[\mathbf{Z}^\top \left[\mathbb{I}_n - \mathbf{X} \left(\mathbf{X}^\top \mathbf{X} + \boldsymbol{\Phi}^{-1} \right)^{-1} \mathbf{X}^\top \right] \mathbf{Z} + \boldsymbol{\Psi}^{-1} \right] \tilde{\mathbf{C}}^\top - \\ & \quad - \left(\mathbf{Y}^\top \mathbf{X} - \mathbb{B}_0 \boldsymbol{\Phi}^{-1} \right) \left[\mathbf{X}^\top \mathbf{X} + \boldsymbol{\Phi}^{-1} \right]^{-1} \left(\mathbf{Y}^\top \mathbf{X} - \mathbb{B}_0 \boldsymbol{\Phi}^{-1} \right)^\top + \\ & \quad \quad \quad + \tilde{\mathbf{C}} \left[\mathbf{Z}^\top \mathbf{Z} + \boldsymbol{\Psi}^{-1} \right] \tilde{\mathbf{C}}^\top, \end{aligned}$$

kde

$$\begin{aligned} \bar{\mathbf{C}} &= \left[\tilde{\mathbf{C}} \left(\mathbf{Z}^\top \mathbf{Z} + \boldsymbol{\Psi}^{-1} \right) - \left(\mathbf{Y}^\top \mathbf{X} - \mathbb{B}_0 \boldsymbol{\Phi}^{-1} \right) \left[\mathbf{X}^\top \mathbf{X} + \boldsymbol{\Phi}^{-1} \right]^{-1} \mathbf{X}^\top \mathbf{Z} \right] \cdot \\ & \quad \cdot \left[\mathbf{Z}^\top \left[\mathbb{I}_n - \mathbf{X} \left(\mathbf{X}^\top \mathbf{X} + \boldsymbol{\Phi}^{-1} \right)^{-1} \mathbf{X}^\top \right] \mathbf{Z} + \boldsymbol{\Psi}^{-1} \right]^{-1} \\ & = \left[\mathbf{Y}^\top \mathbf{A}_\Phi \mathbf{Z} + \mathbf{C}_0 \boldsymbol{\Psi}^{-1} + \mathbb{B}_0 \boldsymbol{\Phi}^{-1} \left(\mathbf{X}^\top \mathbf{X} + \boldsymbol{\Phi}^{-1} \right)^{-1} \mathbf{X}^\top \mathbf{Z} \right] \cdot \left[\mathbf{Z}^\top \mathbf{A}_\Phi \mathbf{Z} + \boldsymbol{\Psi}^{-1} \right]^{-1}. \end{aligned}$$

Nyní už jen stačí dát výsledky dohromady a dostaneme tak výsledné tvrzení. Všimněme si, že člen $\tilde{\mathbf{C}} [\mathbf{Z}^\top \mathbf{Z} + \mathbf{\Psi}^{-1}] \tilde{\mathbf{C}}^\top$ zmizí. □

Přidáme-li předpoklad plných sloupcových hodnotí matic \mathbb{X} a \mathbb{Z} , tak navíc z důsledku tvrzení B7 můžeme využít též vzorce pro $\tilde{\mathbb{B}}(\mathbf{C})$ a $\tilde{\mathbf{C}}$, ze kterého lze vidět, že se jedná o váženou kombinaci klasického regresního odhadu a \mathbb{B}_0 či \mathbf{C}_0 :

$$\begin{aligned}\tilde{\mathbb{B}}(\mathbf{C}) &= (\hat{\mathbb{B}}(\mathbf{C}) \mathbf{X}^\top \mathbf{X} + \mathbb{B}_0 \mathbf{\Phi}^{-1}) [\mathbf{X}^\top \mathbf{X} + \mathbf{\Phi}^{-1}]^{-1}, \\ \hat{\mathbb{B}}(\mathbf{C}) &= (\mathbf{Y} - \mathbf{Z} \mathbf{C}^\top)^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1}, \\ \tilde{\mathbf{C}} &= (\hat{\mathbf{C}} \mathbf{Z}^\top \mathbf{Z} + \mathbf{C}_0 \mathbf{\Psi}^{-1}) [\mathbf{Z}^\top \mathbf{Z} + \mathbf{\Psi}^{-1}]^{-1}, \\ \hat{\mathbf{C}} &= \mathbf{Y}^\top \mathbf{Z} (\mathbf{Z}^\top \mathbf{Z})^{-1}.\end{aligned}$$

C Věta o transformaci


Věta C9 (Věta o transformaci náhodného vektoru). *Nechť k -rozměrný náhodný vektor \mathbf{X} má hustotu p vzhledem k Lebesgueově míře na \mathbb{R}^k . Nechť \mathbf{t} je zobrazení z \mathbb{R}^k do \mathbb{R}^k , které je regulární a prosté na takové otevřené množině G , pro niž platí $\int_G p(\mathbf{x}) d\mathbf{x} = 1$. Označme $\boldsymbol{\tau}$ inverzní zobrazení k $\mathbf{t} : G \rightarrow \mathbf{t}(G)$ a $D_{\boldsymbol{\tau}}$ nechť značí Jakobiův determinant (tzv. jakobián) tohoto zobrazení. Pak náhodný vektor $\mathbf{Y} = \mathbf{t}(\mathbf{X})$ má hustotu vzhledem k Lebesgueově míře a tato hustota je rovna*

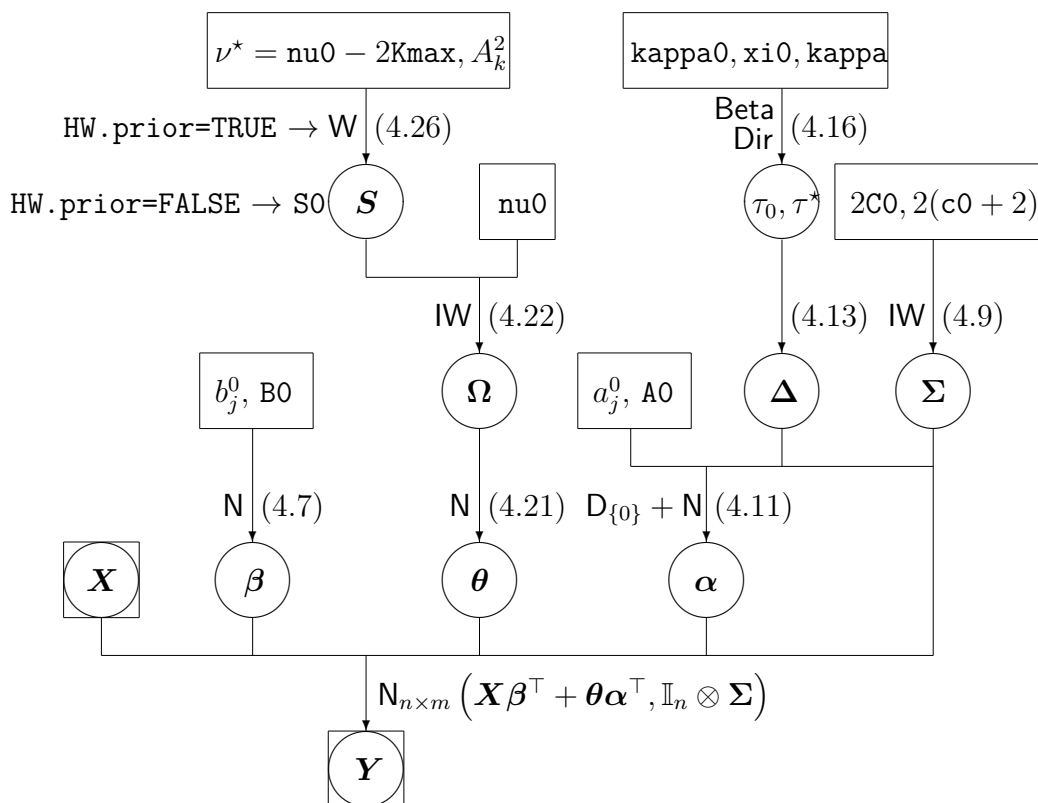
$$q(\mathbf{y}) = p(\boldsymbol{\tau}(\mathbf{y})) \cdot |D_{\boldsymbol{\tau}}| \cdot \mathbb{1}_{\mathbf{t}(G)}(\mathbf{y}).$$

Důkaz. Viz Anděl (2007, Věta 3.7). □

Tuto větu lze zobecnit i pro maticové rozdělení pomocí vektorizace.

D Značení parametrů v modelu určujících faktorů ve funkci befa

Následující diagram D1 má stejné uspořádání parametrů i jejich význam jako diagram 4.1. Ovšem je zde uvedeno takové označení parametrů, jaké je používáno buď ve funkci `befa` z balíčku `BayesFM` softwaru  (strojovým písmem) nebo v samotném článku (Conti a kol., 2014).



Obrázek D1: Diagram popisující strukturu modelu určujících faktorů ve značení používaném v článku (Conti a kol., 2014) a s označením vstupních parametrů funkce `befa` z balíčku `BayesFM`.

Seznam použité literatury

- ANDĚL, J. (2007). *Základy matematické statistiky*. Druhé opravené vydání. Matfyzpress, Praha. ISBN 80-7378-001-1.
- ANDERSON, T. W. a RUBIN, H. (1956). Statistical Inference in Factor Analysis. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, University of Carolina Press*, **5**(6), 111–150.
- BEKKER, P. A. a TEN BERGE, J., M. F. (1997). Generic Global Identification in Factor Analysis. *Linear Algebra and its Applications*, **264**, 255–263.
- BROOKS, S., GELMAN, A., JONES, G., L. a MENG, X. (2011). *Handbook of Markov Chain Monte Carlo*. First edition. Chapman & Hall/CRC, Boca Raton, FL. ISBN 978-1-4200-7941-8.
- CONTI, G., FRÜHWIRTH-SCHNATTER, S., HECKMAN, J., J. a PIATEK, R. (2014). Bayesian exploratory factor analysis. *Journal of Econometrics*, **183** (1), 31–57.
- FRÜHWIRTH-SCHNATTER, S. a LOPES, H. F. (2012). Parsimonious Bayesian Factor Analysis when the Number of Factors is Unknown. *Inspere Working Paper*.
- GEMAN, S. a GEMAN, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**(6), 721–741.
- HARMANN, H. H. (1976). *The Bayesian Choice*. Third edition. The University of Chicago Press, Chicago. ISBN 0-226-31652-1.
- HUŠKOVÁ, M. (1985). *Bayesovské metody*. První vydání. Státní pedagogické nakladatelství, Praha.
- HUANG, A. a WAND, M., P. (2013). Simple Marginally Noninformative Prior Distributions for Covariance Matrices. *Bayesian Analysis*, **8**(2), 439–452.
- LINDLEY, D., V. a SMITH, A., F. M. (1972). Bayes estimates for the linear model. *Journal of The Royal Statistical Society*, **34**, 1–41.
- O’HAGAN, A. a FORSTER, J. (2004). *Kendall’s Advanced Theory of Statistics, Volume 2B Bayesian Inference*. Second edition. Arnold, London. ISBN 0-340-807520.
- PIATEK, R. (2017). *BayesFM: Bayesian Inference for Factor Modeling*. URL <https://CRAN.R-project.org/package=BayesFM>. R package version 0.1.2.
- RAICHE, G. (2010). *an R package for parallel analysis and non graphical solutions to the Cattell scree test*. URL <http://CRAN.R-project.org/package=nFactors>. R package version 2.3.3.
- ROBERT, C. P. (2001). *The Bayesian Choice*. Second edition. Springer-Verlag, New York. ISBN 0-387-95231-4.

ROBERT, C. P. a CASELLA, G. (2004). *Monte Carlo statistical methods*. Second edition. Springer-Verlag, New York. ISBN 0-387-21239-6.

ROWE, D. B. (2003). *Multivariate Bayesian Statistics: Models for Source Separation and Signal Unmixing*. Boca Raton: Chapman & Hall/CRC. ISBN 978-1-58488-318-0.

Seznam obrázků

3.1	Míry ekonomické a osobní svobody prvních tří prezidentů České republiky dle politického kompasu pro volby na podzim roku 2017 (zdroj internet).	32
4.1	Diagram popisující strukturu modelu určujících faktorů.	70
5.1	Rozložení počtu aktivních faktorů ve všech generovaných řetězcích za rozdílných situací, je-li skutečný počet faktorů $K = 2$	90
5.2	Rozložení počtu aktivních faktorů ve všech generovaných řetězcích za rozdílných situací, je-li skutečný počet faktorů $K = 3$	91
5.3	Rozložení počtu aktivních faktorů ve všech generovaných řetězcích za rozdílných situací, je-li skutečný počet faktorů $K = 5$	92
D1	Diagram popisující strukturu modelu určujících faktorů ve značení používaném v článku (Conti a kol., 2014) a s označením vstupních parametrů funkce <code>befa</code> z balíčku <code>BayesFM</code>	102

Seznam tabulek

1.1	Základní charakteristiky pravděpodobnostních rozdělení představených v kapitole 1.	13
3.1	Hodnoty Ledermannovy meze $\phi(m)$ pro počet faktorů k pro různá m	36
5.1	První situace – model určujících faktorů. Odhady skutečného počtu faktorů K při různém počtu celkových měření m a různém rozvržení daném indikátorovou maticí Δ . Počet pozorování $n = 100$, počet opakování simulace 100, délka zážehové fáze $B = 20000$, délka markovského řetězce použitého pro inferenci $M = 20000$. Průměr přes všechna opakování {směrodatná odchylka} (nejčastější navrhaná hodnota) [počet případů navržení nejčastější hodnoty ze 100 možných].	87
5.2	Druhá situace – model s dominantním faktorem. Odhady skutečného počtu faktorů K při různém počtu celkových měření m a různém rozvržení daném indikátorovou maticí Δ . Počet pozorování $n = 100$, počet opakování simulace 100, délka zážehové fáze $B = 20000$, délka markovského řetězce použitého pro inferenci $M = 20000$. Průměr přes všechna opakování {směrodatná odchylka} (nejčastější navrhaná hodnota) [počet případů navržení nejčastější hodnoty ze 100 možných].	88
5.3	Třetí situace – obecný model faktorové analýzy. Odhady skutečného počtu faktorů K při různém počtu celkových měření m . Počet pozorování $n = 100$, počet opakování simulace 100, délka zážehové fáze $B = 20000$, délka markovského řetězce použitého pro inferenci $M = 20000$. Průměr přes všechna opakování {směrodatná odchylka} (nejčastější navrhaná hodnota) [počet případů navržení nejčastější hodnoty ze 100 možných].	89

