

Univerzita Karlova v Praze  
Matematicko-fyzikální fakulta

## BAKALÁŘSKÁ PRÁCE



Jaroslav Vichr

## Expectation-Maximization Algoritmus

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Michal Pešta, Ph.D.

Studijní program: Matematika

Studijní obor: Finanční matematika

Praha 2013

Rád bych poděkoval vedoucímu práce RNDr. Michalovi Peštovi, Ph.D. za množství připomínek, cenných rad, odbornou pomoc a vedení při tvorbě bakalářské práce.

Prohlašuji, že jsem tuto bakalářskou práci vypracoval 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 v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V Praze dne 23.5.2013

Jaroslav Vichr

Název práce: Expectation-Maximization Algoritmus

Autor: Jaroslav Vichr

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Michal Pešta, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: EM (Expectation-Maximization) algoritmus je iterativní metoda sloužící k nalezení odhadu maximální věrohodnosti v případech, kdy buď data obsahují chybějící hodnoty, nebo předpokladem existence dalších skrytých proměnných může dojít ke zjednodušení modelu. Každá jeho iterace se skládá ze dvou částí. V kroku E (expectation) vytváříme očekávání logaritmované věrohodnosti úplných dat, která je podmíněna daty pozorovanými a také současným odhadem zkoumaného parametru. Krok M (maximization) následně hledá nový odhad, který bude maximalizovat funkci získanou v předchozí části a který se následně použije v další iteraci v kroku E. EM algoritmus má významné využití např. v oceňování a řízení rizik portfolia.

Klíčová slova: EM (Expectation-Maximization) algoritmus

Title: Expectation-Maximization Algorithm

Author: Jaroslav Vichr

Department: Department of Probability and Mathematical Statistics

Supervisor: RNDr. Michal Pešta, Ph.D., Department of Probability and Mathematical Statistics

Abstract: EM (Expectation-Maximization) algorithm is an iterative method for finding maximum likelihood estimates in cases, when either complete data include missing values or assuming the existence of additional unobserved data points can lead to more simple formulation of the model. Each of its iterations consists of two parts. During the E step (expectation) we calculate the expected value of the log-likelihood function of the complete data, with respect to the observed data and the current estimate of the parameter. The M step (maximization) then finds new estimate, which will maximize the function obtained in the previous step and which will be used in the next iteration in step E. EM algorithm has important use in e.g. price and manage risk of the portfolio.

Keywords: EM (Expectation-Maximization) algorithm

# Obsah

Úvod	1
<b>1 EM algoritmus</b>	<b>2</b>
1.1 Metoda maximální věrohodnosti . . . . .	2
1.2 Úvod do problému . . . . .	3
1.3 Iterace algoritmu . . . . .	3
1.4 Regulární exponenciální rodiny . . . . .	4
<b>2 Konvergence</b>	<b>6</b>
2.1 Monotonie . . . . .	6
2.2 Míra konvergence . . . . .	7
<b>3 Příklady</b>	<b>11</b>
3.1 Tři mince . . . . .	11
3.2 Normální rozdělení . . . . .	14
3.3 Multinomické rozdělení . . . . .	16
<b>Závěr</b>	<b>20</b>
<b>Tabulky</b>	<b>21</b>
<b>Grafy</b>	<b>25</b>
<b>Literatura</b>	<b>28</b>
<b>A Přílohy</b>	<b>29</b>
A.1 Zdrojové kódy . . . . .	29
A.1.1 Tři mince . . . . .	29
A.1.2 Normální rozdělení . . . . .	30
A.1.3 Multinomické rozdělení . . . . .	32

# Úvod

EM (Expectation-Maximization) algoritmus byl shrnut a pojmenován v článku Dempstera, Lairdové a Rubina (Dempster a kol., 1977). Jedná se o populární iterativní metodu, která se používá k nalezení odhadu maximální věrohodnosti z množiny dat, která je neúplná nebo má chybějící hodnoty.

Existují dva různé přístupy k této metodě. První nastane, pokud v datech máme kvůli problémům s pozorováním skutečně chybějící hodnoty. Druhý případ nastane, pokud optimalizace věrohodnostní funkce je analyticky nezjistitelná. Často lze tento výpočet zjednodušit za předpokladu existence chybějících (skrytých) parametrů a následným použitím EM algoritmu.

Kapitola 1 se zabývá v první části metodou maximální věrohodnosti podle díla Bilmes (viz. Bilmes (1998)), která se poté používá při kroku maximalizace samotného EM algoritmu. V další části se práce opírá o článek Dempstera a kol. (Dempster a kol. (1977)) při představení úplných a pozorovaných dat, dále podle prací Roche (Roche (2012)) a Gupty a Chena (Gupta a Chen (2010)) při sestavení  $Q$ -funkce a popisu jednotlivých iterací. V závěru se kapitola zabývá speciálním případem, kdy pozorovaná data pochází z rozdělení z regulární exponenciální rodiny podle práce Dempstera a kol. (Dempster a kol. (1977)).

Druhá kapitola se věnuje konvergenci algoritmu, nejprve je podle Dempstera a kol. a McLachlana a Krishnana (Dempster a kol. (1977); McLachlan a Krishnan (2008)) ukázána monotonie metody, druhá část se zabývá rychlostí konvergence, mírou konvergence a jejím vyjádřením pomocí matic informací, opírá se při tom o práce McLachlana (McLachlan a Krishnan (2008); McLachlan (1996)).

Poslední kapitola ukazuje použití poznatků z kapitoly první na třech příkladech. Postup je nejprve popsán teoreticky, následně je simulován na náhodně generovaných datech v programu Wolfram Mathematica 9, výsledky jsou zapsány do tabulek v části práce s názvem Tabulky. Zdrojové kódy programu jsou k nahlédnutí v části Přílohy.

# Kapitola 1

## EM algoritmus

### 1.1 Metoda maximální věrohodnosti

Předpokládejme, že máme  $n$  nezávislých stejně rozdělených pozorování  $X = (X_1, X_2, \dots, X_n)^\top$ , o kterých víme pouze to, že jejich rozdělení pochází z modelu  $\mathcal{F} = \{p(x_i; \theta); \theta \in \Theta\}$ , kde  $\theta$  je vektor parametrů. Hledáme neznámou hustotu  $p_0 = p(x_1, x_2, \dots, x_n; \theta_0)$ ,  $\theta_0$  označujeme jako skutečnou hodnotu parametru. Snažíme se najít  $\hat{\theta}$ , nejbližší možný odhad  $\theta_0$ . K tomuto výpočtu určíme sdruženou hustotu náhodného vektoru  $X$ , která je vzhledem k dané nezávislosti:

$$p(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n p_i(x_i; \theta). \quad (1.1)$$

Označme (viz. Bilmes (1998))

$$L_n(\theta) = \prod_{i=1}^n p_i(x_i; \theta),$$

potom  $L_n(\theta)$  nazýváme věrohodnostní funkce. Maximálně věrohodným odhadem parametru  $\theta_0$  rozumíme

$$\hat{\theta} = \arg \max_{\theta} L_n(\theta).$$

V praxi je obvykle jednodušší pracovat s logaritmem věrohodnostní funkce, zvaným logaritmovaná věrohodnost  $L(\theta)$ . Protože  $\log$  je ryze rostoucí funkce, je zaručen stejný výsledek.

$$L(\theta) = \log L_n(\theta) = \sum_{i=1}^n \log p_i(x_i; \theta) \quad (1.2)$$

Hledaný odhad parametru  $\hat{\theta}$  dostaneme, pokud položíme logaritmovanou věrohodnost zderivovanou podle proměnné  $\theta$  rovnu nule:

$$\frac{\partial L(\theta)}{\partial \theta} = \sum_{i=1}^n \frac{\partial}{\partial \theta} \log p(x_i; \theta) = \sum_{i=1}^n \frac{1}{p(x_i; \theta)} \frac{\partial}{\partial \theta} p(x_i; \theta) = 0 \quad (1.3)$$

## 1.2 Úvod do problému

Podle práce Dempstera a kol. (Dempster a kol. (1977)) uvažujme dvě jevová pole  $\mathcal{X}$  a  $\mathcal{Y}$  a zobrazení zobrazující více prvků z  $\mathcal{X}$  do jednoho prvku z  $\mathcal{Y}$ . Přímo pozorujeme data  $y$ , která jsou realizacemi  $\mathcal{Y}$ . Naopak realizace z  $\mathcal{X}$ , označované  $x$ , získáváme až skrz  $y$ . O  $x$  víme jen to, že leží v množině  $\mathcal{X}(y)$ , což je podmnožina  $\mathcal{X}$  splňující rovnici  $y = y(x)$ , za předpokladu, že existuje zobrazení  $x \rightarrow y(x)$ . Realizace  $x$  označujeme jako úplná data a  $y$  jako pozorovaná data.

$\mathcal{X}$  nazveme nosič  $X$ , je to uzávěr množiny  $\{x; p(x|\theta) > 0\}$  a předpokládáme, že nezávisí na  $\theta$ . Jeho podmnožina  $\mathcal{X}(y)$  se nazývá nosič  $X$  podmíněný  $y$  a je to uzávěr množiny  $\{x; p(x|y, \theta) > 0\}$ .

Mějme úplná data modelována jako náhodný vektor  $X$  z rozdělení s hustotou závisící na parametru  $p(x|\theta)$ ,  $\theta \in \Theta$ , kde  $\Theta$  je množina parametrů. Můžeme odvodit odpovídající hustoty  $p(y|\theta)$  pozorovaných dat, které souvisí s úplnými daty následovně:

$$p(y|\theta) = \int_{\mathcal{X}(y)} p(x|\theta) dx.$$

## 1.3 Iterace algoritmu

EM algoritmus je iterativní metoda pro maximalizování logaritmované věrohodnosti  $L(\theta) \equiv \log p(y|\theta)$ . Předpokládejme, že se uskutečnilo  $n$  iterací algoritmu a náš současný odhad parametru označme  $\theta'$ . Neboť naším cílem je maximalizace  $L(\theta)$ , požadujeme, aby nový odhad  $\theta$  splňoval nerovnost

$$L(\theta) > L(\theta'),$$

což ale znamená, že se snažíme maximalizovat rozdíl

$$L(\theta) - L(\theta').$$

Tento rozdíl můžeme dále rozepsat (viz. Roche (2012)) následujícím způsobem:

$$\begin{aligned} L(\theta) - L(\theta') &= \log p(y|\theta) - \log p(y|\theta') = \\ &= \log \frac{p(y|\theta)}{p(y|\theta')} = \\ &= \log \int \frac{p(x, y|\theta)}{p(y|\theta')} dx = \\ &= \log \int \frac{p(x, y|\theta)}{p(x, y|\theta')} p(x|y, \theta') dx = \\ &= \log \int \frac{p(x|\theta)}{p(x|\theta')} p(x|y, \theta') dx \geq \\ &\geq \int \log \frac{p(x|\theta)}{p(x|\theta')} p(x|y, \theta') dx = Q(\theta, \theta'). \end{aligned}$$



Integrál z poslední nerovnosti, která vyplývá z Jensenovy nerovnosti a konkávní vlastnosti logaritmu, můžeme ještě dále rozepsat na

$$\begin{aligned} Q(\theta, \theta') &= \int [\log p(x|\theta) - \log p(x|\theta')] p(x|y, \theta') dx = \\ &= \int \log p(x|\theta) p(x|y, \theta') dx - \int \log p(x|\theta') p(x|y, \theta') dx = \\ &= Q(\theta|\theta') - Q(\theta'|\theta'). \end{aligned}$$

Tedy v každém kroku budeme maximalizovat pomocnou funkci  $Q(\theta|\theta')$ , což je (viz. Gupta a Chen (2010)) podmíněné očekávání logaritmované věrohodnosti úplných dat vzhledem k pozorovaným hodnotám  $y$ . Můžeme ji zapsat ve tvaru

$$Q(\theta|\theta') = E [\log p(x|\theta)|y, \theta'] \quad (1.4)$$

za předpokladu, že odhad  $\theta'$  je správný.

EM algoritmus se skládá ze dvou kroků, kroku očekávání **E** (z anglického expectation) a maximalizace **M** (z anglického maximization). Mějme za sebou  $m$  iterací metody a současný odhad parametru necht' je  $\theta^{(m)}$ . Následující iterace  $\theta^{(m)} \rightarrow \theta^{(m+1)}$  bude vypadat takto:

**E-krok:** Z odhadu  $\theta^{(m)}$  vypočteme podmíněnou hustotu  $p(x|y, \theta^{(m)})$  úplných dat  $x$ , kterou použijeme k sestavení podmíněné očekávané logaritmované věrohodnosti  $Q(\theta|\theta^{(m)})$  podle (1.4).

**M-krok:** Hledáme  $\theta$ , které maximalizuje  $Q(\theta|\theta^{(m)})$ . Tento nový odhad parametru označíme  $\theta^{(m+1)}$ .

Algoritmus v sobě nemá zavedenou podmínku pro zastavení, můžeme ho ukončit např. v případě (viz. McLachlan a Krishnan (2008)), že rozdíl v logaritmovaných věrohodnostech bude menší než nějaké předem dané dostatečně malé  $\varepsilon > 0$ .

$$L(\theta^{(m+1)}) - L(\theta^{(m)}) = \log p(y|\theta^{(m+1)}) - \log p(y|\theta^{(m)}) < \varepsilon$$

V praxi můžeme použít obměnu této metody, kdy v kroku M budeme místo maximalizace  $Q(\theta|\theta^{(m)})$  tuto funkci pouze zvětšovat, což je často o mnoho snazší a monotonní vlastnost algoritmu zůstane zachována. Tomuto postupu se říká GEM (zobecněný EM) algoritmus.

## 1.4 Regulární exponenciální rodiny

Uvažujme speciální případ, kdy  $p(x|\theta)$  patří do exponenciální rodiny, má tedy (viz. Dempster a kol. (1977) a McLachlan a Krishnan (2008)) tvar

$$p(x|\theta) = b(x) \exp(c(\theta) t(x)^\top) / a(\theta), \quad (1.5)$$

kde  $t(x)$  je postačitelná statistika úplných dat v podobě vektoru  $1 \times k$  ( $k \geq d$ ),  $a(\theta)$  a  $b(x)$  jsou skalární funkce parametrů a dat a  $c(\theta)$  o rozměrech  $1 \times k$  je

vektorová funkce vektoru parametrů  $\theta$  o rozměrech  $1 \times d$  z množiny  $\Theta$ , kde  $\Theta$  je  $d$ -dimenzionální konvexní množina taková, že (1.5) definuje hustotu pro každé  $\theta$ , které do ní patří.

$$\Theta = \left\{ \theta; \int_{\mathcal{X}} b(x) \exp(c(\theta) t(x)^\top) dx < \infty \right\}$$

$p(x|\theta)$  patří do regulární exponenciální rodiny, pokud  $k = d$  a Jakobián  $c(\theta)$  má plnou hodnotu.  $c(\theta)$  se pak označuje jako vektor přirozených (kanonických) parametrů a hustota může být zapsána ve tvaru

$$p(x|\theta) = b(x) \exp(\theta t(x)^\top) / a(\theta). \quad (1.6)$$

Očekávání postačitelné statistiky  $t(x)$  v (1.6) je dáno jako

$$\mathbb{E}(t(x)|\theta) = \frac{\partial \log a(\theta)}{\partial \theta}.$$

Předpokládejme, že máme za sebou  $m$  cyklů algoritmu, že současný odhad parametru  $\theta$  je  $\theta^{(m)}$  a že platí vztah (1.6). Příští iterace metody bude vypadat následovně:

**E-krok:** Vyřešením rovnice

$$t^{(m)} = \mathbb{E}(t(x)|y, \theta^{(m)}) \quad (1.7)$$

nalezneme odhad postačitelné statistiky úplných dat  $t(x)$ .

**M-krok:** Nový odhad parametru  $\theta^{(m+1)}$  je řešením rovnice

$$\mathbb{E}(t(x)|\theta) = t^{(m)}. \quad (1.8)$$

Pokud má (1.8) řešení pro  $\theta$  z  $\Theta$  (viz. Dempster a kol. (1977)), potom toto řešení je jedinečné díky konvexní vlastnosti logaritmované věrohodnosti pro regulární exponenciální rodiny. Pokud ale (1.8) není řešitelná pro  $\theta$  z  $\Theta$ , hledané řešení leží na hranici  $\Theta$  a musíme použít obecnější definici.

# Kapitola 2

## Konvergence

### 2.1 Monotonie

Z práce Dempstera a kol. (Dempster a kol. (1977)) víme, že

$$p(x|y, \theta) = \frac{p(x|\theta)}{p(y|\theta)}. \quad (2.1)$$

Aby byla zachována monotonní vlastnost algoritmu, musí pro věrohodnostní funkci pozorovaných dat po každé iteraci platit, že

$$p(y|\theta^{(m+1)}) \geq p(y|\theta^{(m)}), \quad (2.2)$$

kde  $m = 1, 2, 3, \dots$

Pomocí (2.1) můžeme vyjádřit logaritmovanou věrohodnost pozorovaných dat

$$\begin{aligned} L(\theta) &= \log p(y|\theta) = \\ &= \log p(x|\theta) - \log p(x|y, \theta). \end{aligned} \quad (2.3)$$

Nyní můžeme vzít (viz. McLachlan a Krishnan (2008)) očekávání obou stran rovnice (2.3) vzhledem k podmíněnému rozdělení  $X$  při daném  $Y = y$  a tím dostaneme

$$\begin{aligned} L(\theta) &= E[\log p(X|\theta)|y, \theta^{(m)}] - E[\log p(X|y, \theta)|y, \theta^{(m)}] = \\ &= Q(\theta|\theta^{(m)}) - H(\theta|\theta^{(m)}), \end{aligned}$$

kde  $\theta^{(m)}$  je současný odhad parametru  $\theta$  po  $m$  iteracích,  $Q(\theta|\theta^{(m)})$  je funkce  $Q$  definovaná dříve v (1.4) a

$$H(\theta|\theta^{(m)}) = E[\log p(X|y, \theta)|y, \theta^{(m)}].$$

Nerovnici (2.2) můžeme přepsat pomocí funkcí  $Q$  a  $H$  do tvaru

$$Q(\theta^{(m+1)}|\theta^{(m)}) - H(\theta^{(m+1)}|\theta^{(m)}) \geq Q(\theta^{(m)}|\theta^{(m)}) - H(\theta^{(m)}|\theta^{(m)}),$$

odkud následným odečtením pravé strany získáme podmínku

$$\{Q(\theta^{(m+1)}|\theta^{(m)}) - Q(\theta^{(m)}|\theta^{(m)})\} - \{H(\theta^{(m+1)}|\theta^{(m)}) - H(\theta^{(m)}|\theta^{(m)})\} \geq 0. \quad (2.4)$$

Z definice kroku  $\mathbf{M}$  víme, že  $\theta^{(m+1)}$  vybíráme tak, aby maximalizovalo  $Q(\theta|\theta^{(m)})$ , tedy pro každé  $\theta^{(m)}$  musí být splněno

$$Q(\theta^{(m+1)}|\theta^{(m)}) \geq Q(\theta^{(m)}|\theta^{(m)}),$$

a proto menšenec na levé straně (2.4) musí být nezáporný.

Aby byla splněna podmínka (2.4), je nutné, aby menšitel byl nulový nebo záporný. Pro libovolné  $\theta$  můžeme odvodit (viz. McLachlan a Krishnan (2008)) díky konkávní vlastnosti logaritmické funkce a Jensenově nerovnosti

$$\begin{aligned} H(\theta|\theta^{(m)}) - H(\theta^{(m)}|\theta^{(m)}) &= \mathbb{E} [\log p(X|y, \theta)|y, \theta^{(m)}] - \mathbb{E} [\log p(X|y, \theta^{(m)})|y, \theta^{(m)}] = \\ &= \mathbb{E} [\log \frac{p(X|y, \theta)}{p(X|y, \theta^{(m)})}|y, \theta^{(m)}] \leq \\ &\leq \log \mathbb{E} [\frac{p(X|y, \theta)}{p(X|y, \theta^{(m)})}|y, \theta^{(m)}] = \\ &= \log \int_{\mathcal{X}(y)} \frac{p(x|y, \theta)}{p(x|y, \theta^{(m)})} p(x|y, \theta^{(m)}) dx = \\ &= \log \int_{\mathcal{X}(y)} p(x|y, \theta) dx = \\ &= \log 1 = 0. \end{aligned}$$

Nerovnost (2.4) je tedy splněna a tím pádem platí i (2.2), což dokazuje, že po jedné iteraci se věrohodnost  $p(y|\theta)$  nesníží a EM algoritmus má tedy monotonní vlastnost.

## 2.2 Míra konvergence

Na EM algoritmus můžeme nahlížet jako na zobrazení  $\theta \rightarrow M(\theta)$  zobrazující prostor  $\Theta$  sám na sebe tak, že iterace  $\theta^{(m)} \rightarrow \theta^{(m+1)}$  se dá zapsat následujícím způsobem:

$$\theta^{(m+1)} = M(\theta^{(m)})$$

pro  $m = 0, 1, 2, \dots$ , kde  $\theta^{(0)}$  je počáteční odhad parametru  $\theta$ . Předpokládejme, že posloupnost  $\{\theta^{(m)}\}$  konverguje k nějakému  $\theta'$ , které může být sedlovým bodem, lokálním maximem nebo odhadem maximální věrohodnosti. Pokud je zobrazení  $M$  spojitě, potom platí, že

$$\theta' = M(\theta')$$

a říkáme, že  $\theta'$  je pevný bod zobrazení  $M$ .

Použijeme (viz. McLachlan (1996)) Taylorův rozvoj  $\theta^{(m+1)} = M(\theta^{(m)})$  na okolí bodu  $\theta'$  a dostaneme

$$\theta^{(m+1)} - \theta' \approx J(\theta')(\theta^{(m)} - \theta'),$$

kde matice  $J(\theta')$  je Jakobián pro  $M(\theta') = (M_1(\theta'), \dots, M_r(\theta'))^\top$  o rozměrech  $d \times d$  taková, že její prvek o souřadnicích  $(i, j)$  je roven

$$\frac{\partial M_i(\theta')}{\partial \theta'_j}.$$

Na okolí  $\theta'$  je  $J(\theta')$  nemulová (viz. McLachlan a Krishnan (2008)), iterace algoritmu jsou tedy pouze lineární. Z tohoto důvodu je  $J(\theta')$  nazývána mírou konvergence.

Označme

$$r = \lim_{m \rightarrow \infty} \frac{\|\theta^{(m+1)} - \theta'\|}{\|\theta^{(m)} - \theta'\|}$$

globální míru konvergence pro vektor parametrů  $\theta$ , kde  $\|\cdot\|$  je norma na  $d$ -rozměrném Euklidovském prostoru  $\mathbb{R}^d$ . V praxi je však běžnější vyjádření

$$r = \lim_{m \rightarrow \infty} \frac{\|\theta^{(m+1)} - \theta^{(m)}\|}{\|\theta^{(m)} - \theta^{(m-1)}\|}.$$

Zdefinujeme  $r_i \equiv 0$ , pokud  $\theta_i^{(m)} \equiv \theta_i^{(k)}$  pro všechna  $m \geq k$ ,  $k$  pevné. Potom můžeme vyjádřit jednotlivá  $r_i$  jako

$$r_i = \lim_{m \rightarrow \infty} \frac{\|\theta_i^{(m+1)} - \theta_i^{(m)}\|}{\|\theta_i^{(m)} - \theta_i^{(m-1)}\|}.$$

Je možné dokázat, že

$$r = \max_{1 \leq i \leq d} r_i,$$

algoritmus tedy konverguje tehdy a jen tehdy, pokud konverguje každý prvek  $\theta_i$ .

Globální míra konvergence  $r$  je největší vlastní číslo matice  $J(\theta')$ . Podle Roche (2012) zdefinujeme matici

$$K = I - \frac{\partial M}{\partial \theta} \Big|_{\theta'} = I - J(\theta'), \quad (2.5)$$

kde  $I$  je jednotková matice o rozměrech  $d \times d$ .  $K$  se nazývá rychlost konvergence. Označme  $s$  její spektrální poloměr, což je nejmenší vlastní číslo  $K$ . Pak platí, že  $s = 1 - r$  a říká se mu globální rychlost konvergence. Pokud je  $s \neq 1$ , potom je lokální konvergence EM algoritmu pouze lineární.

Míra konvergence může být vyjádřena také pomocí matic informací. Mějme posloupnost  $\{\theta^{(m)}\}$  a necht' pro  $\theta = \theta^{(m+1)}$  je splněno

$$\frac{\partial Q(\theta|\theta^{(m)})}{\partial \theta} = 0. \quad (2.6)$$

Označme (viz. Roche (2012)) druhou derivaci logaritmované věrohodnosti pozorovaných dat

$$I_y(\theta) = I(\theta|y) = - \frac{\partial^2 \log p(y|\theta)}{\partial \theta \partial \theta^\top}$$

a druhou derivaci logaritmované věrohodnosti úplných dat

$$I_x(\theta) = I(\theta|x) = - \frac{\partial^2 \log p(x|\theta)}{\partial \theta \partial \theta^\top}.$$

Pokud posloupnost  $\{\theta^{(m)}\}$  konverguje k  $\theta'$ , potom  $I_y(\theta')$  je matice informací pro pozorovaná data  $y$  a  $I_x(\theta')$  je matice informací úplných dat  $x$ .

Dále označme

$$\begin{aligned}\mathcal{J}_x(\theta^{(m)}|y) &= \mathbf{E} [I_x(\theta^{(m)})|y, \theta^{(m)}] = \\ &= \mathbf{E} \left[ - \frac{\partial^2 \log p(x|\theta^{(m)})}{\partial \theta \partial \theta^\top} \middle| y, \theta^{(m)} \right] = \\ &= - \left[ \frac{\partial^2 Q(\theta|\theta^{(m)})}{\partial \theta \partial \theta^\top} \right]_{\theta^{(m)}} \\ &= - \int_{\mathcal{X}(y)} \frac{\partial^2}{\partial \theta \partial \theta^\top} [\log p(x|\theta^{(m)})] p(x|y, \theta^{(m)}) dx,\end{aligned}$$

kde poslední rovnost platí z vyjádření  $Q$ -funkce (1.4).

Použijme Taylorův rozvoj na  $\frac{\partial}{\partial \theta} Q(\theta|\theta^{(m)})$  okolo bodu  $\theta^{(m)}$  a vyhodnoťme ho v  $\theta^{(m+1)}$  (viz. McLachlan a Krishnan (2008)), máme tedy

$$\begin{aligned}\left[ \frac{\partial Q(\theta|\theta^{(m)})}{\partial \theta} \right]_{\theta^{m+1}} &\approx \left[ \frac{\partial Q(\theta|\theta^{(m)})}{\partial \theta} \right]_{\theta^{(m)}} + \left[ \frac{\partial^2 Q(\theta|\theta^{(m)})}{\partial \theta \partial \theta^\top} \right]_{\theta^{(m)}} (\theta^{(m+1)} - \theta^{(m)}) = \\ &= S(y|\theta^{(m)}) - \mathcal{J}_x(\theta^{(m)}|y) (\theta^{(m+1)} - \theta^{(m)}),\end{aligned}$$

kde  $S(y|\theta^{(m)})$  je vektor gradientu logaritmované věrohodnosti pozorovaných dat  $\log p(y|\theta)$  v bodě  $\theta^{(m)}$ . Z rovnice (2.6) dostaneme

$$S(y|\theta^{(m)}) - \mathcal{J}_x(\theta^{(m)}|y) (\theta^{(m+1)} - \theta^{(m)}) \approx 0,$$

a tedy

$$S(y|\theta^{(m)}) \approx \mathcal{J}_x(\theta^{(m)}|y) (\theta^{(m+1)} - \theta^{(m)}). \quad (2.7)$$

Na vektor  $S(y|\theta)$  použijeme Taylorův rozvoj na okolí  $\theta^{(m)}$ , získáme

$$S(y|\theta) \approx S(y|\theta^{(m)}) - I(\theta^{(m)}|y) (\theta - \theta^{(m)}).$$

Po dosazení  $\theta'$  za  $\theta$  a následnou úpravou zjistíme aproximaci bodu  $\theta'$ .

$$\begin{aligned}S(y|\theta') &\approx S(y|\theta^{(m)}) - I(\theta^{(m)}|y) (\theta' - \theta^{(m)}) \\ S(y|\theta') &\approx S(y|\theta^{(m)}) - I(\theta^{(m)}|y) \theta' + I(\theta^{(m)}|y) \theta^{(m)} \\ I(\theta^{(m)}|y) \theta' &\approx S(y|\theta^{(m)}) - S(y|\theta') + I(\theta^{(m)}|y) \theta^{(m)} \\ \theta' &\approx \theta^{(m)} + I^{-1}(\theta^{(m)}|y) S(y|\theta^{(m)})\end{aligned} \quad (2.8)$$

Aproximace (2.8) platí pouze za předpokladu, že gradient  $S(y|\theta)$  je v bodě  $\theta'$  nulový. Odečtením  $\theta^{(m)}$  od obou stran a následným dosazením za gradient z (2.7) máme

$$\begin{aligned}\theta' - \theta^{(m)} &\approx I^{-1}(\theta^{(m)}|y) S(y|\theta^{(m)}), \\ \theta' - \theta^{(m)} &\approx I^{-1}(\theta^{(m)}|y) \mathcal{J}_x(\theta^{(m)}|y) (\theta^{(m+1)} - \theta^{(m)}).\end{aligned}$$

Udělejme pomocný krok, kdy v posledním činiteli uměle odečteme a následně přičteme  $\theta'$ .

$$\begin{aligned}\theta' - \theta^{(m)} &\approx I^{-1}(\theta^{(m)}|y) \mathcal{J}_x(\theta^{(m)}|y) (\theta^{(m+1)} - \theta' + \theta' - \theta^{(m)}) \\ \theta' - \theta^{(m)} &\approx I^{-1}(\theta^{(m)}|y) \mathcal{J}_x(\theta^{(m)}|y) (\theta^{(m+1)} - \theta') + I^{-1}(\theta^{(m)}|y) \mathcal{J}_x(\theta^{(m)}|y) (\theta' - \theta^{(m)})\end{aligned}$$

Odečtením posledního sčítance a následným vytknutím získáme

$$\begin{aligned}\theta' - \theta^{(m)} - I^{-1}(\theta^{(m)}|y) \mathcal{J}_x(\theta^{(m)}|y)(\theta' - \theta^{(m)}) &\approx I^{-1}(\theta^{(m)}|y) \mathcal{J}_x(\theta^{(m)}|y) (\theta^{(m+1)} - \theta'), \\ (\theta^{(m)} - \theta') (I^{-1}(\theta^{(m)}|y) \mathcal{J}_x(\theta^{(m)}|y) - I_d) &\approx I^{-1}(\theta^{(m)}|y) \mathcal{J}_x(\theta^{(m)}|y) (\theta^{(m+1)} - \theta').\end{aligned}$$

Nyní osamostatníme poslední činitel na pravé straně

$$\begin{aligned}\theta^{(m+1)} - \theta' &\approx (\theta^{(m)} - \theta') (I_d - I(\theta^{(m)}|y) \mathcal{J}_x^{-1}(\theta^{(m)}|y)), \\ \theta^{(m+1)} - \theta' &\approx (\theta^{(m)} - \theta') (I_d - I(\theta'|y) \mathcal{J}_x^{-1}(\theta'|y)), \\ \theta^{(m+1)} - \theta' &\approx J(\theta') (\theta^{(m)} - \theta'),\end{aligned}$$

kde

$$J(\theta') = I_d - I(\theta'|y) \mathcal{J}_x^{-1}(\theta'|y), \quad (2.9)$$

což je vyjádření míry konvergence pomocí matic informací.

Podobným způsobem můžeme vyjádřit také rychlost konvergence z (2.5) a (2.9) a dostaneme

$$K = \mathcal{J}_x^{-1}(\theta'|y) I_y(\theta').$$

Obecně očekáváme (viz. McLachlan (1996)), že druhá derivace logaritmované věrohodnosti pozorovaných dat na okolí  $\theta'$  bude negativně semidefinitní nebo negativně definitní matice, a proto vlastní hodnoty  $J(\theta')$  budou ležet na intervalu  $[0, 1]$  nebo  $[0, 1)$ .

# Kapitola 3

## Příklady

### 3.1 Tři mince

Předpokládejme, že máme k dispozici 3 mince, které označíme  $M_0$ ,  $M_1$  a  $M_2$ . Pravděpodobnost, že na  $M_0$  padne líc, je

$$P(z = L|\theta) = \lambda, \quad (3.1)$$

že na  $M_0$  padne rub, je

$$P(z = R|\theta) = 1 - \lambda. \quad (3.2)$$

Pokud nastane první možnost, dále házíme třikrát mincí  $M_1$ , pokud nastane druhá, házíme třikrát  $M_2$ . Nevíme však, jestli na  $M_0$  padl rub, nebo líc, ani jestli se házelo třikrát mincí  $M_1$ , nebo  $M_2$ . Máme k dispozici pouze výslednou posloupnost představující tři hody na  $M_2$  nebo  $M_3$  tvaru např.

$$y = \{\{LLL\}, \{RRR\}, \{LLL\}, \{RRR\}, \{LLL\}\}. \quad (3.3)$$

Naším úkolem je odhadnout kromě parametru  $\lambda$  také parametry  $p_1$  a  $p_2$  představující pravděpodobnosti padnutí líce na mincích  $M_1$  a  $M_2$ .

Označme úplná data  $X$ . Potom platí, že  $x = (y, z)^\top$ , kde  $y$  jsou pozorovaná neúplná data (např. (3.3)), která jsou doplněna skrytými proměnnými  $z$ , což je v našem případě posloupnost rubů a líců, které padly na  $M_0$ .

Množina všech hodnot, kterých mohou úplná data  $x$  nabývat, je

$$\begin{aligned} \mathcal{X} = & \{(\{LLL\}, L); (\{RRR\}, L); (\{LRR\}, L); (\{RLL\}, L); \\ & (\{LLR\}, L); (\{RRL\}, L); (\{LRL\}, L); (\{RLR\}, L); \\ & (\{LLL\}, R); (\{RRR\}, R); (\{LRR\}, R); (\{RLL\}, R); \\ & (\{LLR\}, R); (\{RRL\}, R); (\{LRL\}, R); (\{RLR\}, R)\}. \end{aligned}$$

Pro pozorovaná data (3.3) označme  $\mathcal{X}(y)$  nosič  $X$  podmíněný  $y$ , což je podmnožina  $\mathcal{X}$  taková, že

$$\mathcal{X}(y) = \{(\{LLL\}, L); (\{LLL\}, R); (\{RRR\}, L); (\{RRR\}, R)\}.$$

Proměnné  $y$  jsou prvky jevového pole  $\mathcal{Y}$  takového, že

$$\mathcal{Y} = \{\{LLL\}, \{RRR\}, \{LRR\}, \{RLL\}, \{LLR\}, \{RRL\}, \{LRL\}, \{RLR\}\},$$



zatímco  $z$  patří do jevového pole  $\mathcal{Z}$ , kde

$$\mathcal{Z} = \{L, R\}.$$

Parametr, který budeme odhadovat, je pravděpodobnost, že na minci padne líc, tedy

$$\theta = \{\lambda, p_1, p_2\}$$

a leží v množině parametrů  $\Theta = [0, 1]$ .

Pro sdružené rozdělení  $P(y, z|\theta)$  platí

$$P(y, z|\theta) = P(y|z, \theta) P(z|\theta). \quad (3.4)$$

Označme  $l$  počet líců a  $r$  počet rubů v pozorovaných datech  $y$ . Potom můžeme vyjádřit podmíněné rozdělení  $y$  jako

$$P(y|z, \theta) = \begin{cases} p_1^l (1 - p_1)^r, & \text{je-li } z = L, \\ p_2^l (1 - p_2)^r, & \text{je-li } z = R. \end{cases} \quad (3.5)$$

Nyní můžeme pomocí (3.1), (3.2), (3.4) a (3.5) spočítat pravděpodobnosti, například

$$\begin{aligned} P(y = LLR, z = L|\theta) &= p_1^2 (1 - p_1) \lambda, \\ P(y = LLR, z = R|\theta) &= p_2^2 (1 - p_2) (1 - \lambda). \end{aligned}$$

Z nich dále

$$\begin{aligned} P(y = LLR|\theta) &= P(y = LLR, z = L|\theta) + P(y = LLR, z = R|\theta) = \\ &= p_1^2 (1 - p_1) \lambda + p_2^2 (1 - p_2) (1 - \lambda), \end{aligned}$$

a také

$$\begin{aligned} P(z = L|y = LLR, \theta) &= \frac{P(y = LLR, z = L|\theta)}{P(y = LLR, z = L|\theta) + P(y = LLR, z = R|\theta)} = \\ &= \frac{p_1^2 (1 - p_1) \lambda}{p_1^2 (1 - p_1) \lambda + p_2^2 (1 - p_2) (1 - \lambda)}. \end{aligned}$$

Předpokládáme, že současná hodnota parametrů je  $\lambda$ ,  $p_1$  a  $p_2$ . Dále uvažujme pozorovaná data  $y$  z (3.3). Můžeme spočítat pravděpodobnosti pro první hodnotu těchto dat  $\{LLL\}$  pro  $z = L$  a  $z = R$  následujícím způsobem (pro další hodnoty analogicky).

$$\begin{aligned} P(z = L|y = \{LLL\}) &= \frac{P(\{LLL\}, L)}{P(\{LLL\}, L) + P(\{LLL\}, R)} = \\ &= \frac{\lambda p_1^3}{\lambda p_1^3 + (1 - \lambda) p_2^3}, \end{aligned}$$

$$\begin{aligned} P(z = R|y = \{LLL\}) &= \frac{P(\{LLL\}, R)}{P(\{LLL\}, R) + P(\{LLL\}, L)} = \\ &= \frac{(1 - \lambda) p_2^3}{(1 - \lambda) p_2^3 + \lambda p_1^3}. \end{aligned}$$

Pro počáteční hodnoty parametrů  $\lambda^{(0)} = 0.2$ ,  $p_1^{(0)} = 0.4$  a  $p_2^{(0)} = 0.6$  vychází:

$$\begin{aligned} P(z = L|y = \{LLL\}) &= 0.068966, \\ P(z = R|y = \{LLL\}) &= 0.931034. \end{aligned}$$

S použitím pozorovaných dat (3.3) a doplněním skrytých proměnných dostáváme

( $\{LLL\}, L$ )	$P(z = L y = \{LLL\})$	=	0.068966,
( $\{LLL\}, R$ )	$P(z = R y = \{LLL\})$	=	0.931034,
( $\{RRR\}, L$ )	$P(z = L y = \{RRR\})$	=	0.457627,
( $\{RRR\}, R$ )	$P(z = R y = \{RRR\})$	=	0.542373,
( $\{LLL\}, L$ )	$P(z = L y = \{LLL\})$	=	0.068966,
( $\{LLL\}, R$ )	$P(z = R y = \{LLL\})$	=	0.931034,
( $\{RRR\}, L$ )	$P(z = L y = \{RRR\})$	=	0.457627,
( $\{RRR\}, R$ )	$P(z = R y = \{RRR\})$	=	0.542373,
( $\{LLL\}, L$ )	$P(z = L y = \{LLL\})$	=	0.068966,
( $\{LLL\}, R$ )	$P(z = R y = \{LLL\})$	=	0.931034.

Z výše spočtených podmíněných pravděpodobností můžeme vypočítat nové odhady parametrů

$$\lambda^{(1)} = \frac{3 * 0.068966 + 2 * 0.457627}{5} = 0.224430,$$

$$p_1^{(1)} = \frac{3 * 3 * 0.068966}{3 * 3 * 0.068966 + 3 * 2 * 0.457627} = 0.184375,$$

$$p_2^{(1)} = \frac{3 * 3 * 0.931034}{3 * 3 * 0.931034 + 2 * 3 * 0.542373} = 0.720271.$$

Tímto jsme dostali  $\theta^{(1)} = (\lambda^{(1)}, p_1^{(1)}, p_2^{(1)})$ , což jsou odhady parametrů po první iteraci. Použijme tyto hodnoty k výpočtu nových pravděpodobností a celý proces opakujme. Algoritmus zastaví v kroku  $m$ , pokud budou splněny všechny podmínky

$$\begin{aligned} |\lambda^{(m)} - \lambda^{(m+1)}| &< 0.0001, \\ |p_1^{(m)} - p_1^{(m+1)}| &< 0.0001, \\ |p_2^{(m)} - p_2^{(m+1)}| &< 0.0001. \end{aligned}$$

Po čtyřech iteracích dostáváme  $\lambda^{(4)} = 0.4$ ,  $p_1^{(4)} = 0$  a  $p_2^{(4)} = 1$  (viz. tabulka (3.1)).

To, že  $\lambda$  konverguje k hodnotě 0.4, znamená, že na minci  $M_0$  padne rub s větší pravděpodobností než líc, tedy budeme v následujících třech hodech používat spíše minci  $M_2$ , proto parametr  $p_2$  konverguje k jedné, zatímco  $p_1$  jde do nuly. Toto platí pouze, pokud je  $p_1^{(0)}$  menší než  $p_2^{(0)}$ . V opačném případě parametr  $\lambda$  konverguje k hodnotě 0.6, a proto se  $p_1$  blíží k jedné a  $p_2$  k nule. Změna počátečního odhadu  $\lambda^{(0)}$  výsledek nezmění.

Přidejme nyní do pozorovaných dat (3.3) ještě jedno pozorování  $\{RRR\}$ , abychom dostali vyvážená data. Nově máme

$$y = \{\{LLL\}, \{RRR\}, \{LLL\}, \{RRR\}, \{LLL\}, \{RRR\}\}. \quad (3.6)$$

Uvažujme opět stejné počáteční odhady  $\lambda^{(0)} = 0.2$ ,  $p_1^{(0)} = 0.4$  a  $p_2^{(0)} = 0.6$ .

Na první pohled je vidět (viz. tabulka (3.2)), že algoritmus vede ke správnému řešení. Na minci  $M_0$  padne buď rub, nebo líc se stejnou pravděpodobností, tedy parametr  $\lambda$  by měl konvergovat k hodnotě 0.5, což je skutečně odhad po třetí iteraci. Pro  $p_1$  platí, že jde k nule, protože jeho počáteční odhad je menší než počáteční odhad parametru  $p_2$ , který jde k jedné.

Vezmeme stejná pozorovaná data (3.6), ale změňme počáteční odhady parametrů tak, že  $\lambda^{(0)} = 0.2$  a  $p_1^{(0)} = p_2^{(0)} = 0.2$ .

V tomto případě je vidět (viz. tabulka (3.3)), že pro stejné výchozí hodnoty parametrů  $p_1$  a  $p_2$  algoritmus pro data (3.6) dosáhne sedlového bodu  $\lambda^{(1)} = 0.5$ ,  $p_1^{(1)} = 0.5$  a  $p_2^{(1)} = 0.5$ .

Stačí ale, abychom nepatrně změnili jednu z těchto pravděpodobností a metoda nalezne hledané globální maximum. Zvýšením počátečního odhadu  $p_1^{(0)}$  o 0.001 dosáhneme po jedenácti iteracích řešení  $\lambda^{(11)} = 0.5$ ,  $p_1^{(11)} = 1$  a  $p_2^{(11)} = 0$  (viz tabulka (3.4)). Při snížení  $p_1^{(0)}$  o 0.001 získáme analogicky  $\lambda^{(11)} = 0.5$ ,  $p_1^{(11)} = 0$  a  $p_2^{(11)} = 1$  (viz tabulka (3.5)).

Zdrojový kód k příkladu Tři mince z programu Wolfram Mathematica 9 je k nahlédnutí v kapitole Přílohy.

## 3.2 Normální rozdělení

Ukažme si použití EM algoritmu na příkladě s náhodným výběrem z normálního rozdělení  $N(\mu, \sigma^2)$  o velikosti  $n$ . Úplná data můžeme rozdělit následujícím způsobem

$$x = (y, z)^\top = (x_1, \dots, x_r, x_{r+1}, \dots, x_n)^\top = (y_1, \dots, y_r, z_1, \dots, z_{n-r})^\top,$$

kde  $y = (y_1, \dots, y_r)^\top$  je vektor pozorovaných dat, která máme k dispozici, délky  $r$  a  $z = (z_1, \dots, z_{n-r})^\top$  je vektor skrytých proměnných délky  $n - r$ .

Hustota normálního rozdělení pro hodnotu  $x_1$  je

$$p(x_1|\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x_1 - \mu)^2}{2\sigma^2}\right\},$$

vektor parametrů, které budeme odhadovat, je v tomto případě  $\theta = (\mu, \sigma^2)^\top$ . Neboť náhodný výběr obsahuje nezávislé stejně rozdělené veličiny, můžeme podle

rovnice (1.1) vyjádřit věrohodnostní funkci úplných dat jako

$$\begin{aligned}
L_n(\theta) &= \prod_{i=1}^n \left( \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(x_i - \mu)^2}{2\sigma^2} \right\} \right) = \\
&= \left( \frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} \exp\left\{ -\frac{(x_1 - \mu)^2}{2\sigma^2} \right\} \exp\left\{ -\frac{(x_2 - \mu)^2}{2\sigma^2} \right\} \cdots \exp\left\{ -\frac{(x_n - \mu)^2}{2\sigma^2} \right\} = \\
&= \left( \frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} \exp\left\{ -\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2} \right\} = \\
&= \left( \frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \left( \sum_{i=1}^n x_i^2 - \sum_{i=1}^n 2\mu x_i + \sum_{i=1}^n \mu^2 \right) \right\} = \\
&= \left( \frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \left( \sum_{i=1}^n x_i^2 - 2\mu \sum_{i=1}^n x_i + n\mu^2 \right) \right\}.
\end{aligned}$$

Odtud dostáváme, že vektor postačitelých statistik pro vektor parametrů  $\theta$  je  $t = (t_1, t_2)^\top = \left( \sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2 \right)^\top$ . Z rovnice (1.2) můžeme vyjádřit logaritmovanou věrohodnost úplných dat

$$\begin{aligned}
L(\theta) &= \log \left( \frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} - \frac{1}{2\sigma^2} \left( \sum_{i=1}^n x_i^2 - 2\mu \sum_{i=1}^n x_i + n\mu^2 \right) = \\
&= -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2 + \frac{\mu}{\sigma^2} \sum_{i=1}^n x_i - \frac{n\mu^2}{2\sigma^2}. \tag{3.7}
\end{aligned}$$

Jelikož se jedná o rozdělení z exponenciální rodiny, můžeme použít verzi algoritmu uvedenou v podkapitole 1.5. V kroku **E** budeme počítat odhad vektoru postačujících statistik v iteraci  $m$ . Použitím rovnice (1.7) máme

$$t^{(m)} = (t_1^{(m)}, t_2^{(m)})^\top = \mathbf{E}(t|y, \mu^{(m)}, \sigma^{2(m)}),$$

kde

$$t_1^{(m)} = \mu^{(m)} = \mathbf{E}\left(\sum_{i=1}^n x_i | y, \mu^{(m)}, \sigma^{2(m)}\right) = \sum_{i=1}^r y_i + (n-r)\mu^{(m)} \tag{3.8}$$

a

$$t_2^{(m)} = \sigma^{2(m)} = \mathbf{E}\left(\sum_{i=1}^n x_i^2 | y, \mu^{(m)}, \sigma^{2(m)}\right) = \sum_{i=1}^r y_i^2 + (n-r)(\sigma^{2(m)} + \mu^{(m)2}). \tag{3.9}$$

V rovnici (3.8) se v poslední rovnosti využilo faktu, že  $\mathbf{E}(x_i | \mu^{(m)}, \sigma^{2(m)}) = \mu^{(m)}$ , v (3.9) poslední rovnost platí díky tomu, že  $\text{var}(x_i^2 | \mu^{(m)}, \sigma^{2(m)}) = \sigma^{2(m)}$  a

$$\text{var}(x_i^2 | \mu^{(m)}, \sigma^{2(m)}) = \mathbf{E}(x_i^2 | \mu^{(m)}, \sigma^{2(m)}) - \mathbf{E}^2(x_i | \mu^{(m)}, \sigma^{2(m)}),$$

proto  $\mathbf{E}(x_i^2 | \mu^{(m)}, \sigma^{2(m)}) = \sigma^{2(m)} + \mu^{(m)2}$ .

Pro provedení kroku **M** musíme spočítat odhady maximální věrohodnosti pro úplná data. Dosazením logaritmované věrohodnosti (3.7) do rovnice (1.3) získáme soustavu

$$\frac{\partial L}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n x_i - \frac{1}{\sigma^2} n\mu = \frac{1}{\sigma^2} \left( \sum_{i=1}^n x_i - n\mu \right) = 0,$$

$$\frac{\partial L}{\partial \sigma^2} = -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2} \frac{1}{(\sigma^2)^2} \sum_{i=1}^n x_i^2 - \frac{\mu}{(\sigma^2)^2} \sum_{i=1}^n x_i + \frac{n\mu^2}{2(\sigma^2)^2} = 0,$$

jejímž řešením je vektor

$$\theta' = (\mu', \sigma'^2)^\top = \left( \frac{\sum_{i=1}^n x_i}{n}, \frac{\sum_{i=1}^n x_i^2}{n} - \left( \frac{\sum_{i=1}^n x_i}{n} \right)^2 \right)^\top. \quad (3.10)$$

Dosazením očekávání pro postačitelné statistiky získané v kroku **E** do (3.10) dostaneme nové odhady parametrů

$$\begin{aligned} \mu^{(m+1)} &= \frac{t_1^{(m)}}{n}, \\ \sigma^{2(m+1)} &= \frac{t_2^{(m)}}{n} - (\mu^{(m+1)})^2. \end{aligned}$$

Ukažme si nyní na náhodně generovaných datech, jak ovlivní podíl pozorovaných dat  $y$  v úplných datech  $x$  rychlost konvergence. Algoritmus zastaví v iteraci  $m$ , pokud budou splněny obě podmínky

$$\begin{aligned} |\mu^{(m)} - \mu^{(m+1)}| &< 0.001, \\ |\sigma^{2(m)} - \sigma^{2(m+1)}| &< 0.001. \end{aligned}$$

Předpokládejme, že skutečná hodnota parametrů je  $\theta = (\mu, \sigma^2)^\top = (0, 1)^\top$  a že úplných dat  $x$  je 1000 (viz. obrázek (3.1)). Pro pozorovaná data  $y$  budeme postupně zkoumat 2 možnosti.

- Pozorujeme velkou část  $x$ , např.  $y$  má 800 prvků.
- Máme k dispozici polovinu úplných dat.

Zvolme počáteční odhady parametrů  $\mu^{(0)} = 10$  a  $\sigma^{2(0)} = 10$  stejné pro obě možnosti.

Je vidět, že algoritmus konverguje rychleji, pokud je poměr pozorovaných a úplných dat blíže k jedné, neboť pro 800 pozorování našel řešení po osmi iteracích (viz. tabulka (3.6)), zatímco pro 500 pozorování bylo potřeba iterací 16 (viz. tabulka (3.7)). Metoda nalezne řešení i pro menší počet pozorovaných dat, rychlost je ale potom velmi nízká (pro 100 pozorování je zapotřebí 89 iterací, pro 20 dokonce 382).

Zdrojový kód k příkladu Normální rozdělení z programu Wolfram Mathematica 9 je k nahlédnutí v kapitole Přílohy.

### 3.3 Multinomické rozdělení

Tento příklad je zpracován podle vzoru Dempster a kol. (1977) a Gupta a Chen (2010).

Mějme pozorovaná data  $y$  velikosti  $n$  rozdělená multinomicky do čtyř kategorií, tedy

$$y = (y_1, y_2, y_3, y_4)^\top.$$

$y_i$  si můžeme představit jako přihrádku, do které každé z  $n$  pozorování padne s pravděpodobností  $p_i$  pro  $i = 1, \dots, 4$  takové, že  $p_i \in \{(0, 1)^4 : \sum_{i=1}^4 p_i = 1\}$ . Můžeme vyjádřit pravděpodobnostní funkci multinomického rozdělení jako

$$p(y|p) = \frac{n!}{y_1! y_2! y_3! y_4!} p_1^{y_1} p_2^{y_2} p_3^{y_3} p_4^{y_4}. \quad (3.11)$$

Předpokládejme, že vektor pravděpodobností  $p = (p_1, p_2, p_3, p_4)^\top$  je parametrizován nějakou skrytou proměnnou  $\theta$  tak, že

$$p_\theta = \left( \frac{1}{2} + \frac{1}{4}\theta, \frac{1}{4}(1 - \theta), \frac{1}{4}(1 - \theta), \frac{1}{4}\theta \right)^\top; \quad \theta \in (0, 1).$$

Dosažením  $p_\theta$  do předpisu pro pravděpodobnostní funkci (3.11) získáme

$$p(y|\theta) = \frac{n!}{y_1! y_2! y_3! y_4!} \left( \frac{1}{2} + \frac{1}{4}\theta \right)^{y_1} \left( \frac{1}{4}(1 - \theta) \right)^{y_2} \left( \frac{1}{4}(1 - \theta) \right)^{y_3} \left( \frac{1}{4}\theta \right)^{y_4}.$$

Zvolme nyní úplná data  $x$  tak, abychom mohli vyjádřit pravděpodobnostní funkci v závislosti pouze na  $\theta$  a  $1 - \theta$ . Toho dosáhneme rozložením  $y_1$  na součet  $x_1 + x_2$ . Úplná data  $x = (x_1, x_2, x_3, x_4, x_5)^\top$ , kde  $x_1 + x_2 = y_1$ ,  $x_3 = y_2$ ,  $x_4 = y_3$  a  $x_5 = y_4$  budou potom také multinomicky rozdělena a jejich vektor pravděpodobností nastání každé možnosti bude

$$q_\theta = \left( \frac{1}{2}, \frac{1}{4}\theta, \frac{1}{4}(1 - \theta), \frac{1}{4}(1 - \theta), \frac{1}{4}\theta \right)^\top; \quad \theta \in (0, 1).$$

Nyní můžeme vyjádřit věrohodnostní funkci úplných dat jako

$$\begin{aligned} p(x|\theta) &= \frac{n!}{x_1! x_2! x_3! x_4! x_5!} \left( \frac{1}{2} \right)^{x_1} \left( \frac{1}{4}\theta \right)^{x_2} \left( \frac{1}{4}(1 - \theta) \right)^{x_3} \left( \frac{1}{4}(1 - \theta) \right)^{x_4} \left( \frac{1}{4}\theta \right)^{x_5} = \\ &= \frac{n!}{x_1! x_2! x_3! x_4! x_5!} \left( \frac{1}{2} \right)^{x_1} \left( \frac{1}{4}\theta \right)^{x_2+x_5} \left( \frac{1}{4}(1 - \theta) \right)^{x_3+x_4}. \end{aligned}$$

Abychom mohli použít **E**-krok algoritmu, musíme nejprve podle rovnice (1.4) vytvořit  $Q$ -funkci

$$\begin{aligned} Q(\theta|\theta^{(m)}) &= \mathbf{E} [\log p(x|\theta)|y, \theta^{(m)}] = \\ &= \mathbf{E} \left\{ \left[ c + x_1 \log \left( \frac{1}{2} \right) + (x_2 + x_5) \log \left( \frac{\theta}{4} \right) + \right. \right. \\ &\quad \left. \left. + (x_3 + x_4) \log \left( \frac{1 - \theta}{4} \right) \right] | y, \theta^{(m)} \right\} = \\ &= \mathbf{E}_{x|y, \theta^{(m)}} [c] + \mathbf{E}_{x|y, \theta^{(m)}} [x_1] \log \left( \frac{1}{2} \right) + \{ \mathbf{E}_{x|y, \theta^{(m)}} [x_2 + x_5] \} \log \left( \frac{\theta}{4} \right) + \\ &\quad + \{ \mathbf{E}_{x|y, \theta^{(m)}} [x_3 + x_4] \} \log \left( \frac{1 - \theta}{4} \right) = \\ &= c + \mathbf{E}_{x|y, \theta^{(m)}} [x_1] \log \left( \frac{1}{2} \right) + \{ \mathbf{E}_{x|y, \theta^{(m)}} [x_2] + \mathbf{E}_{x|y, \theta^{(m)}} [x_5] \} \log \left( \frac{\theta}{4} \right) + \\ &\quad + \{ \mathbf{E}_{x|y, \theta^{(m)}} [x_3] + \mathbf{E}_{x|y, \theta^{(m)}} [x_4] \} \log \left( \frac{1 - \theta}{4} \right), \quad (3.12) \end{aligned}$$

kde

$$c = \log \frac{n!}{x_1! x_2! x_3! x_4! x_5!}.$$

Při sestavení  $Q$ -funkce musíme spočítat očekávání úplných dat  $x$  podmíněné pozorovanými daty  $y$  a současnou hodnotou odhadu parametru  $\theta$  v iteraci  $m$ , což je  $\theta^{(m)}$ .

V našem případě platí, že  $\mathbf{E}_{x|y, \theta^{(m)}}[x_3] = y_2$ ,  $\mathbf{E}_{x|y, \theta^{(m)}}[x_4] = y_3$  a  $\mathbf{E}_{x|y, \theta^{(m)}}[x_5] = y_4$ , zbývá ještě zjistit očekávání  $x_1$  a  $x_2$ . Protože víme, že  $x_1 + x_2 = y_1$ , můžeme při daném  $y_1$  uvažovat, že dvojice  $(x_1, x_2)$  je binomicky rozdělená a  $x_1$  udává počet úspěchů v  $y_1$  pokusech. Platí tedy

$$\begin{aligned} p((x_1, x_2)|y_1, \theta^{(m)}) &= \binom{y_1}{x_1} \left(\frac{q_1}{p_1}\right)^{x_1} \left(1 - \frac{q_1}{p_1}\right)^{y_1 - x_1} = \\ &= \frac{y_1!}{x_1!(y_1 - x_1)!} \left(\frac{\frac{1}{2}}{\frac{1}{2} + \frac{\theta^{(m)}}{4}}\right)^{x_1} \left(\frac{\frac{\theta^{(m)}}{4}}{\frac{1}{2} + \frac{\theta^{(m)}}{4}}\right)^{y_1 - x_1} = \\ &= \frac{y_1!}{x_1!x_2!} \left(\frac{2}{2 + \theta^{(m)}}\right)^{x_1} \left(\frac{\theta^{(m)}}{2 + \theta^{(m)}}\right)^{x_2}. \end{aligned} \quad (3.13)$$

Celkové očekávání úplných dat  $x$  při pevně daném  $y$  a současném odhadu parametru  $\theta^{(m)}$  je

$$\mathbf{E}[x|y, \theta^{(m)}] = \left(\frac{2}{2 + \theta^{(m)}}y_1, \frac{\theta^{(m)}}{2 + \theta^{(m)}}y_1, y_2, y_3, y_4\right)^\top. \quad (3.14)$$

Dosazením očekávání (3.14) do předpisu  $Q$ -funkce (3.12) dostaneme

$$\begin{aligned} Q(\theta|\theta^{(m)}) &= c + \left(\frac{2}{2 + \theta^{(m)}}y_1\right) \log\left(\frac{1}{2}\right) + \left(\frac{\theta^{(m)}}{2 + \theta^{(m)}}y_1 + y_4\right) \log\left(\frac{\theta}{4}\right) + \\ &+ (y_2 + y_3) \log\left(\frac{1 - \theta}{4}\right). \end{aligned}$$

V kroku  $\mathbf{M}$  hledáme  $\theta$ , které bude maximalizovat funkci  $Q(\theta|\theta^{(m)})$ . Takové  $\theta$  označíme  $\theta^{(m+1)}$  a bude to nový odhad parametru, pro který platí

$$\theta^{(m+1)} = \arg \max_{\theta \in (0,1)} Q(\theta|\theta^{(m)}). \quad (3.15)$$

Abychom vyřešili rovnici (3.15), stačí nám pracovat pouze s částmi  $Q$ -funkce, které závisí na  $\theta$ , ostatní členy nebudou mít na maximalizování vliv. Po této úpravě nám zůstane

$$\begin{aligned} \theta^{(m+1)} &= \arg \max_{\theta \in (0,1)} \left( \left(\frac{\theta^{(m)}}{2 + \theta^{(m)}}y_1 + y_4\right) \log(\theta) + (y_2 + y_3) \log(1 - \theta) \right) = \\ &= \frac{\frac{\theta^{(m)}}{2 + \theta^{(m)}}y_1 + y_4}{\frac{\theta^{(m)}}{2 + \theta^{(m)}}y_1 + y_2 + y_3 + y_4}. \end{aligned}$$

Ukažme si nyní tento příklad na konkrétních datech. Předpokládejme, že skutečná hodnota parametru  $\theta$  je 0.2, skutečná hodnota vektoru pravděpodobností tedy bude

$$p_\theta = [0.55, 0.2, 0.2, 0.05]^\top.$$

K dispozici máme pozorovaná data z multinomického rozdělení o velikosti 1000 (viz. obrázek (3.2)), 10000 (viz. obrázek (3.3)) a 100000 (viz. obrázek (3.4)).

Pracujme s počátečním odhadem parametru  $\theta^{(0)} = 0.5$ . Algoritmus zastaví v kroku  $m$ , pokud bude splněna podmínka

$$|\theta^{(m)} - \theta^{(m+1)}| < 0.0001.$$

Pro pozorovaná data velikosti 1000 (viz. tabulka (3.8)) najde metoda řešení po osmi iteracích. Je ale vidět, že odhad  $\theta^{(8)} = 0.160462$  je relativně nepřesný, neboť skutečná hodnota parametru je  $\theta = 0.2$ .

Pro data velikosti 10000 (viz. tabulka (3.9)) i 100000 (viz. tabulka (3.10)) potřebuje metoda k vyřešení problému stejný počet iterací, ale zároveň je vidět, že čím je větší rozsah pozorování, tím je odhad přesnější.

Zdrojový kód k příkladu Multinomické rozdělení z programu Wolfram Mathematica 9 je k nahlédnutí v kapitole Přílohy.



# Závěr

Základní teorie týkající se EM algoritmu zpracovaná v kapitole 1 byla použita při řešení tří příkladů v kapitole 3.

Na motivačním příkladu Tři mince je ukázáno, jak počáteční odhad parametru může změnit výsledný odhad získaný touto metodou. Pro pozorovaná data platí, že výsledné odhady pravděpodobností padnutí líce na mincích  $M_1$  a  $M_2$ ,  $p_1$  a  $p_2$ , nezávisí na  $\lambda^{(0)}$ , ale jsou určeny pouze porovnáním  $p_1^{(0)}$  a  $p_2^{(0)}$ . Pro  $p_1^{(0)} > p_2^{(0)}$  platí, že  $p_1$  jde k jedné a  $p_2$  k nule, pro  $p_1^{(0)} < p_2^{(0)}$  naopak  $p_2$  jde k jedné a  $p_1$  k nule. Může také nastat speciální případ pro  $p_1^{(0)} = p_2^{(0)}$ . V této situaci algoritmus konverguje ke stacionárnímu bodu.

Druhý příklad ukazuje použití metody na datech z normálního rozdělení. Je zde vidět, jak množství pozorovaných dat ovlivňuje rychlost konvergence. Předpokládáme, že úplná data mají velikost 1000 a postupně používáme algoritmus na náhodně generovaná pozorovaná data, jejichž velikosti jsou 800, 500 a 100. Pro 800 metoda nalezne řešení po pouhých osmi iteracích, pro 500 po šestnácti a pro 100 je jich třeba 89. V tabulkách je zaznamenán přehled odhadů pro první dvě možnosti.

V poslední části kapitoly 3 je uvedeno použití algoritmu pro data z multinomického rozdělení. V tomto příkladě je ukázáno, jak velikost pozorovaných dat ovlivňuje přesnost výsledných odhadů parametrů. Konkrétně pro generovaná data velikosti 1000 je odchylka odhadu od skutečné hodnoty parametru 0.0395, pro 10000 je odchylka 0.0023 a pro 100000 už jen 0.00087, to vše pro stejný počáteční odhad.

# Tabulky

Iterace $m$	$\lambda^{(m)}$	$p_1^{(m)}$	$p_2^{(m)}$
0	0.2	0.4	0.6
1	0.22443	0.184375	0.720271
2	0.353959	0.00818793	0.924247
3	0.399675	0.0000005718	0.999459
4	0.4	0	1

Tabulka 3.1: Příklad Tři mince, data (3.3), počáteční odhady  $\lambda^{(0)} = 0.2$ ,  $p_1^{(0)} = 0.4$  a  $p_2^{(0)} = 0.6$ .

Iterace $m$	$\lambda^{(m)}$	$p_1^{(m)}$	$p_2^{(m)}$
0	0.2	0.4	0.6
1	0.263296	0.130966	0.631892
2	0.413907	0.00383162	0.8504
3	0.497614	0.0000000649	0.99525
4	0.5	0	1

Tabulka 3.2: Příklad Tři mince, data (3.6), počáteční odhady  $\lambda^{(0)} = 0.2$ ,  $p_1^{(0)} = 0.4$  a  $p_2^{(0)} = 0.6$ .

Iterace $m$	$\lambda^{(m)}$	$p_1^{(m)}$	$p_2^{(m)}$
0	0.2	0.2	0.2
1	0.2	0.5	0.5
2	0.2	0.5	0.5

Tabulka 3.3: Příklad Tři mince, data (3.6), počáteční odhady  $\lambda^{(0)} = 0.2$ ,  $p_1^{(0)} = 0.2$  a  $p_2^{(0)} = 0.2$ .

Iterace $m$	$\lambda^{(m)}$	$p_1^{(m)}$	$p_2^{(m)}$
0	0.2	0.2001	0.2
1	0.20009	0.500375	0.499906
2	0.20009	0.501125	0.499719
3	0.200093	0.503374	0.499156
4	0.200113	0.510121	0.497468
5	0.200297	0.530328	0.492404
6	0.201948	0.59008	0.477205
7	0.216167	0.74852	0.431463
8	0.306733	0.962001	0.295589
9	0.469269	0.999926	0.0579688
10	0.49989	1	0.000220262
11	0.5	1	0

Tabulka 3.4: Příklad Tři mince, data (3.6), počáteční odhady  $\lambda^{(0)} = 0.2$ ,  $p_1^{(0)} = 0.2001$  a  $p_2^{(0)} = 0.2$ .

Iterace $m$	$\lambda^{(m)}$	$p_1^{(m)}$	$p_2^{(m)}$
0	0.2	0.1999	0.2
1	0.19991	0.499625	0.500094
2	0.19991	0.498875	0.500281
3	0.199913	0.496624	0.500844
4	0.199933	0.489873	0.502531
5	0.200118	0.469654	0.507592
6	0.20177	0.409867	0.522783
7	0.215998	0.251357	0.568503
8	0.306617	0.0379283	0.70433
9	0.469236	0.0000735755	0.941972
10	0.499889	0	0.999779
11	0.5	0	1

Tabulka 3.5: Příklad Tři mince, data (3.6), počáteční odhady  $\lambda^{(0)} = 0.2$ ,  $p_1^{(0)} = 0.1999$  a  $p_2^{(0)} = 0.2$ .

Iterace $m$	$\mu^{(m)}$	$\sigma^{2(m)}$
0	10	10
1	1.98692	18.8393
2	0.384304	5.1969
3	0.0637812	1.85201
4	-0.000323402	1.15838
5	-0.0131443	1.01866
6	-0.0157085	0.990681
7	-0.0162214	0.985083
8	-0.0163239	0.983963

Tabulka 3.6: Příklad Normální rozdělení, počáteční odhady  $\mu^{(0)} = \sigma^{(0)} = 10$ , 1000 úplých dat a 800 pozorování.

Iterace $m$	$\mu^{(m)}$	$\sigma^{2(m)}$
0	10	10
1	5.01121	30.3812
2	2.51681	21.9057
3	1.26961	13.0015
4	0.646016	7.38276
5	0.334217	4.28173
6	0.178317	2.6583
7	0.100367	1.82836
8	0.061392	1.40883
9	0.0419045	1.19793
10	0.0321608	1.09219
11	0.0272889	1.03925
12	0.024853	1.01277
13	0.023635	0.999519
14	0.0230261	0.992894
15	0.0227216	0.989581
16	0.0225693	0.987924

Tabulka 3.7: Příklad Normální rozdělení, počáteční odhady  $\mu^{(0)} = \sigma^{(0)} = 10$ , 1000 úplých dat a 500 pozorování.

Iterace $m$	$\theta^{(m)}$	$E[x_1 y, \theta^{(m)}]$	$E[x_2 y, \theta^{(m)}]$
0	0.5	429	107
1	0.259454	429	107
2	0.195127	474	62
3	0.173256	488	48
4	0.165237	493	43
5	0.162217	495	41
6	0.161068	496	40
7	0.160629	496	40
8	0.160462	496	40

Tabulka 3.8: Příklad Multinomické rozdělení, pozorovaná data  $y$  velikosti 1000, počáteční odhad  $\theta^{(0)} = 0.5$ .

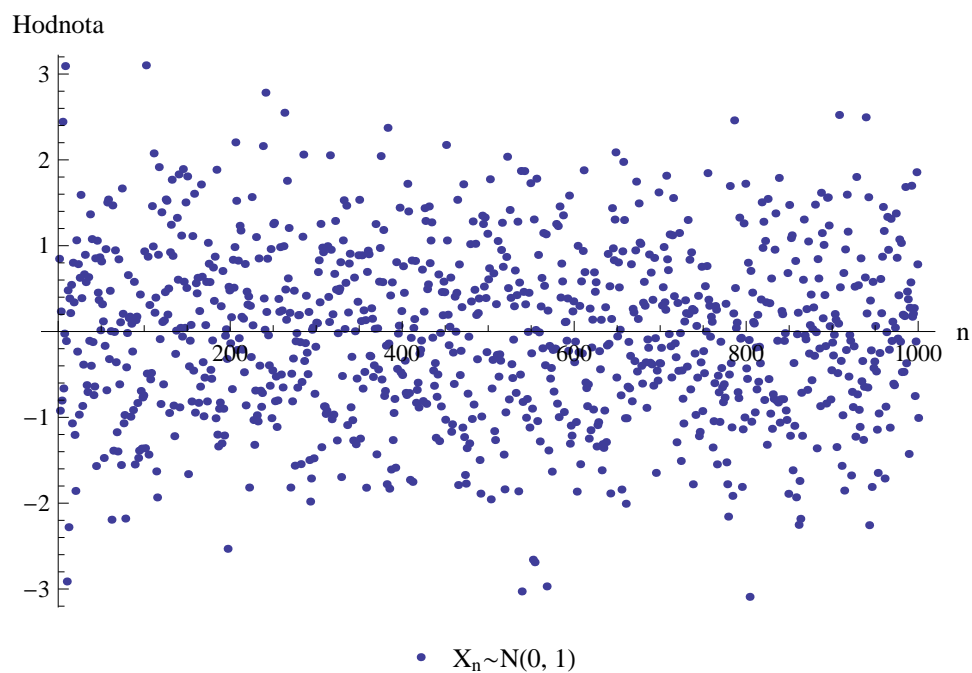
Iterace $m$	$\theta^{(m)}$	$E[x_1 y, \theta^{(m)}]$	$E[x_2 y, \theta^{(m)}]$
0	0.5	4446	1112
1	0.28857	4446	1112
2	0.231744	4857	701
3	0.212814	4981	577
4	0.206074	5023	535
5	0.203617	5039	519
6	0.202715	5044	514
7	0.202382	5046	512
8	0.202259	5047	511

Tabulka 3.9: Příklad Multinomické rozdělení, pozorovaná data  $y$  velikosti 10000, počáteční odhad  $\theta^{(0)} = 0.5$ .

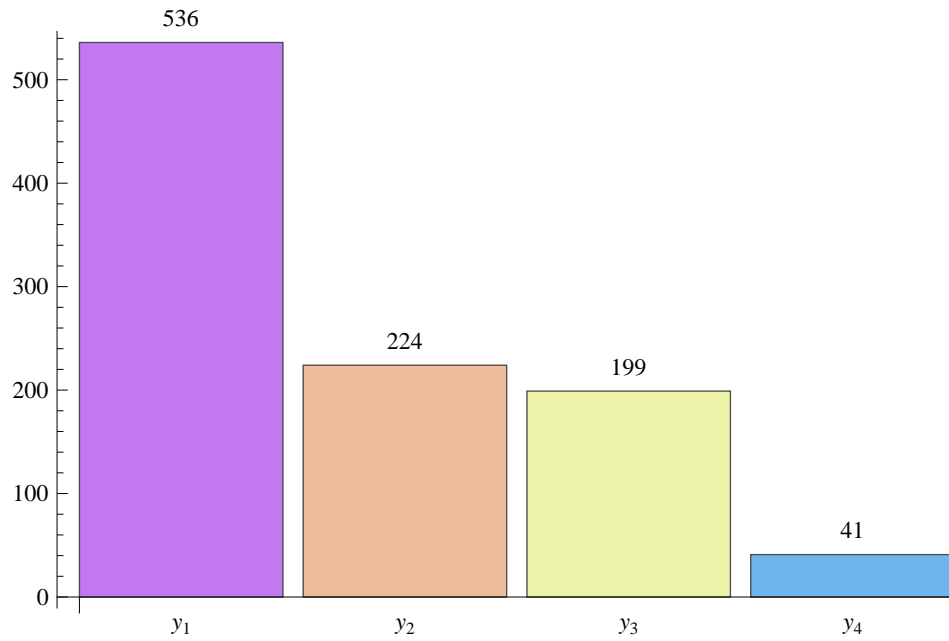
Iterace $m$	$\theta^{(m)}$	$E[x_1 y, \theta^{(m)}]$	$E[x_2 y, \theta^{(m)}]$
0	0.5	44024	11006
1	0.285265	44024	11006
2	0.22823	48161	6869
3	0.20943	49393	5637
4	0.20281	49814	5216
5	0.200424	49963	5067
6	0.199558	50018	5012
7	0.199242	50037	4993
8	0.199127	50045	4985

Tabulka 3.10: Příklad Multinomické rozdělení, pozorovaná data  $y$  velikosti 100000, počáteční odhad  $\theta^{(0)} = 0.5$ .

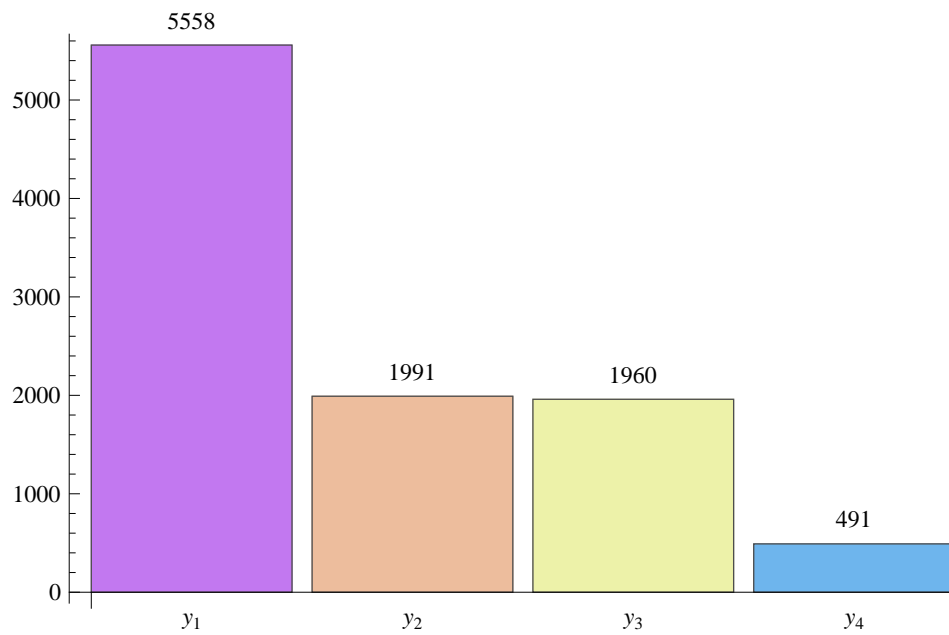
# Grafy



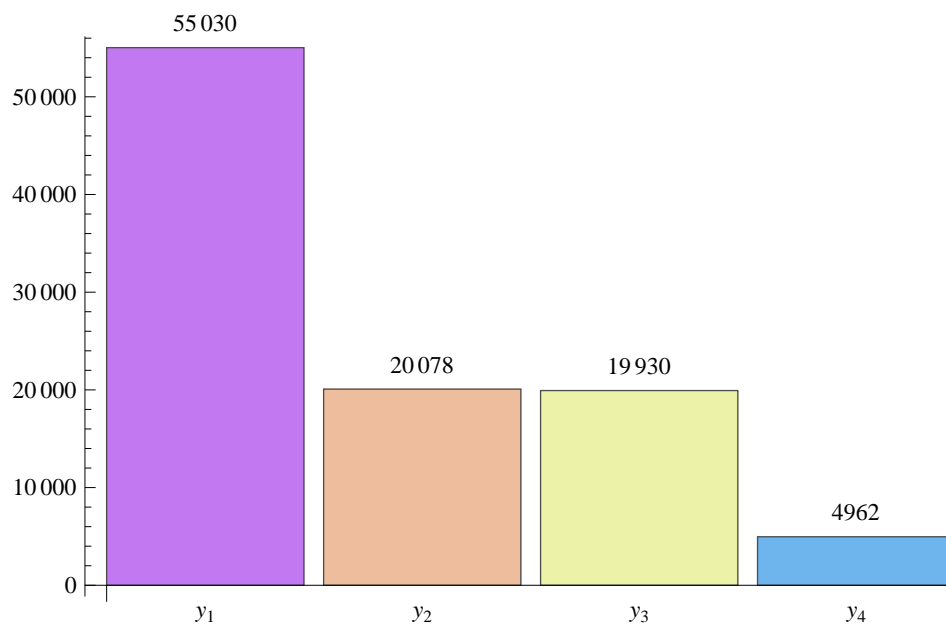
Obrázek 3.1: Příklad Normální rozdělení, úplná data  $x$  o velikosti 1000 z rozdělení  $N(0, 1)$ .



Obrázek 3.2: Příklad Multinomické rozdělení, pozorovaná data  $y$  velikosti 1000.



Obrázek 3.3: Příklad Multinomické rozdělení, pozorovaná data  $y$  velikosti 10000.



Obrázek 3.4: Příklad Multinomické rozdělení, pozorovaná data  $y$  velikosti 100000.



# Literatura

- DEMPSTER, A. P., LAIRD, N. M. a RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **39**(1), 1–38.
- GUPTA, M. R. a CHEN, Y. (1977). Theory and use of the EM algorithm. *Foundations and Trends in Signal Processing*, **4**(3), 223–296.
- MCLACHLAN, G. J. a KRISHNAN, T. (2008). *The EM algorithm and extensions*. Wiley series in probability and statistics. John Wiley and Sons, Inc., Hoboken, New Jersey. ISBN 978-0-471-20170-0.
- BILMES, J. A. (1998). A gentle tutorial of the EM algorithm and its application to parameter estimation for gaussian mixture and Hidden Markov Models. *International Computer Science Institute*, 1–13.
- ROCHE, A. (2012). EM algorithm and variants: An informal tutorial. *Service Hospitalier Frederic Joliot*, 1–17.
- MCLACHLAN, G. J. (1996). On Aitken’s method and other approaches for accelerating convergence of the EM algorithm. *Proceedings of the A.C. Aitken Centenary Conference*, University of Otago, 1995. Dunedin: University of Otago Press, 201–209.

# Příloha A

## Přílohy

### A.1 Zdrojové kódy

#### A.1.1 Tři mince

Funkce pro výpočet pravděpodobností pozorování podmíněných  $z = L$ . Jejich parametry jsou současné hodnoty odhadů parametrů  $\lambda$ ,  $p_1$  a  $p_2$ . První funkce počítá pravděpodobnosti  $P(L|\{LLR\})$ ,  $P(L|\{LRL\})$  a  $P(L|\{RLL\})$ , druhá  $P(L|\{RRR\})$ , třetí  $P(L|\{LLL\})$  a čtvrtá  $P(L|\{RRL\})$ ,  $P(L|\{RLR\})$  a  $P(L|\{LRR\})$ :

```
L2R1l[a_, b_, c_] := (a*b^2*(1 - b))/( a*b^2*(1 - b) + (1 - a)*c^2*(1 - c))
R3l[a_, b_, c_] := (a*(1 - b)^3)/(a*(1 - b)^3 + (1 - a)*(1 - c)^3)
L3l[a_, b_, c_] := (a*b^3)/(a*b^3 + (1 - a)*c^3)
R2L1l[a_, b_, c_] := (a*b*(1 - b)^2)/( a*b*(1 - b)^2 + (1 - a)*c*(1 - c)^2)
```

Analogie funkcí pro výpočet pravděpodobností jednotlivých pozorování, tentokrát ale podmíněných  $z = R$ :

```
L2R1r[a_, b_, c_] := ((1 - a)*c^2*(1 - c))/( a*b^2*(1 - b)
+ (1 - a)*c^2*(1 - c))
R3r[a_, b_, c_] := ((1 - a)*(1 - c)^3)/( a*(1 - b)^3 + (1 - a)*(1 - c)^3)
L3r[a_, b_, c_] := ((1 - a)*c^3)/(a*b^3 + (1 - a)*c^3)
R2L1r[a_, b_, c_] := ((1 - a)*c*(1 - c)^2)/( a*b*(1 - b)^2
+ (1 - a)*c*(1 - c)^2)
```

Odhady nových parametrů pro pozorovaná data (3.3), parametry funkcí jsou opět současné odhady  $\lambda$ ,  $p_1$  a  $p_2$ :

```
odhadlambda[a_, b_, c_] := (2*R3l[a, b, c] + 3*L3l[a, b, c])/5
odhadp1[a_, b_, c_] := (3*3*L3l[a, b, c])/( 3*3*L3l[a, b, c]
+ 2*3*R3l[a, b, c])
odhadp2[a_, b_, c_] := (3*3*L3r[a, b, c])/( 3*3*L3r[a, b, c]
+ 2*3*R3r[a, b, c])
```

Algoritmus pro data (3.3). Vstupními parametry jsou počáteční odhady  $\lambda^{(0)}$ ,  $p_1^{(0)}$  a  $p_2^{(0)}$ , funkce vrací jednotlivé iterace:

```
algoritmus[lambda_, p1_, p2_] :=
Module[{pocet = 1, prvni = {0, lambda, p1, p2}, a = lambda, b = p1,
```

```

    c = p2, novaIterace, pomocna1, pomocna2, pomocna3},
Print[prvni];
While[Abs[a - odhadlambda[a, b, c]] > 0.0001;
Abs[b - odhadp1[a, b, c]] > 0.0001;
Abs[c - odhadp2[a, b, c]] > 0.0001,
novaIterace = {pocet, odhadlambda[a, b, c], odhadp1[a, b, c],
    odhadp2[a, b, c]};
    Print[novaIterace];
    pomocna1 = odhadlambda[a, b, c]; pomocna2 = odhadp1[a, b, c];
    pomocna3 = odhadp2[a, b, c]; pocet++;
    a = pomocna1; b = pomocna2; c = pomocna3;
]
]

```

Odhady nových parametrů pro pozorovaná data (3.6), parametry funkcí jsou současné odhady  $\lambda$ ,  $p_1$  a  $p_2$ :

```

odhad2lambda[a_, b_, c_] := (3*R3l[a, b, c] + 3*L3l[a, b, c])/6
odhad2p1[a_, b_, c_] := (3*3*L3l[a, b, c])/( 3*3*L3l[a, b, c]
+ 3*3*R3l[a, b, c])
odhad2p2[a_, b_, c_] := (3*3*L3r[a, b, c])/( 3*3*L3r[a, b, c]
+ 3*3*R3r[a, b, c])

```

Algoritmus pro data (3.6). Vstupními parametry jsou počáteční odhady  $\lambda^{(0)}$ ,  $p_1^{(0)}$  a  $p_2^{(0)}$ , funkce vrací jednotlivé iterace:

```

algoritmus[lambda_, p1_, p2_] :=
Module[{pocet = 1, prvni = {0, lambda, p1, p2}, a = lambda, b = p1,
    c = p2, novaIterace, pomocna1, pomocna2, pomocna3},
Print[prvni];
While[Abs[a - odhad2lambda[a, b, c]] > 0.0001;
Abs[b - odhad2p1[a, b, c]] > 0.0001;
Abs[c - odhad2p2[a, b, c]] > 0.0001,
novaIterace = {pocet, odhad2lambda[a, b, c], odhad2p1[a, b, c],
    odhad2p2[a, b, c]};
    Print[novaIterace];
    pomocna1 = odhad2lambda[a, b, c]; pomocna2 = odhad2p1[a, b, c];
    pomocna3 = odhad2p2[a, b, c]; pocet++;
    a = pomocna1; b = pomocna2; c = pomocna3;
]
]

```

## A.1.2 Normální rozdělení

Funkce generující data z normálního rozdělení  $N(\mu, \sigma^2)$ , vstupní parametry jsou  $\mu$ ,  $\sigma^2$  a požadovaný počet dat:

```

generovani[u_, s_, pocety_] := (
SeedRandom[13];
RandomVariate[NormalDistribution[u, s], pocety]
)

```

**E**-krok algoritmu, vstupními parametry jsou současné odhady  $\mu^{(m)}$ ,  $\sigma^{2(m)}$ , pozorovaná data  $y$  a celkový počet úplných dat  $x$ . Funkce vrací odhad postačujících statistik  $t_1$  a  $t_2$ :

```
krokE[u_, s_, y_, celkem_] :=
Module[{pocet, odhadt1, odhadt2},
pocet = Length[y];
odhadt1 = Sum[y[[i]], {i, 1, pocet}] + (celkem - pocet)*u;
odhadt2 = Sum[y[[i]]^2, {i, 1, pocet}] + (celkem - pocet)*(s + u^2);
{odhadt1, odhadt2}
]
```

**M**-krok algoritmu, vstupními parametry jsou současné odhady postačujících statistik  $t_1$ ,  $t_2$  a celkový počet úplných dat  $x$ . Funkce vrací nové odhady  $\mu^{(m+1)}$  a  $\sigma^{2(m+1)}$ :

```
krokM[t1_, t2_, celkem_] := Module[{unove, snove},
unove = t1/celkem;
snove = t2/celkem - unove^2;
{unove, snove}
]
```

Funkce představující jednu iteraci algoritmu, spojuje kroky **E** a **M**, vstupními parametry jsou současné odhady  $\mu^{(m)}$ ,  $\sigma^{2(m)}$ , pozorovaná data  $y$  a celkový počet úplných dat  $x$ , vrací nové odhady  $\mu^{(m+1)}$  a  $\sigma^{2(m+1)}$ :

```
iterace[u_, s_, y_, celkem_] :=
Module[{odhadt1, odhadt2, odhadu, odhads},
odhadt1 = krokE[u, s, y, celkem][[1]];
odhadt2 = krokE[u, s, y, celkem][[2]];
odhadu = krokM[odhadt1, odhadt2, celkem][[1]];
odhads = krokM[odhadt1, odhadt2, celkem][[2]];
{odhadu, odhads}
]
```

Algoritmus pro pozorovaná data  $y$ . Vstupní parametry jsou současné odhady  $\mu^{(m)}$ ,  $\sigma^{2(m)}$ ,  $y$  a celkový počet úplných dat  $x$ , funkce vrací jednotlivé iterace:

```
algoritmus[u_, s_, y_, celkem_] :=
Module[{pocet = 0, odhadu = u, odhads = s,
pomocna1, pomocna2, data},
Print[{pocet, odhadu, odhads}];
While[Abs[odhadu - iterace[odhadu, odhads, y, celkem][[1]]] > 0.001;
Abs[odhads - iterace[odhadu, odhads, y, celkem][[2]]] > 0.001,
pocet++;
Print[{pocet, iterace[odhadu, odhads, y, celkem][[1]],
iterace[odhadu, odhads, y, celkem][[2]]}];
pomocna1 = iterace[odhadu, odhads, y, celkem][[1]];
pomocna2 = iterace[odhadu, odhads, y, celkem][[2]];
odhadu = pomocna1;
odhads = pomocna2;
]
]
```

Pomocná funkce, která nagenereuje pozorovaná data  $y$  a následně na ně použije algoritmus. Vstupní parametry jsou skutečné hodnoty  $\mu$ ,  $\sigma^2$ , počáteční odhady  $\mu^{(0)}$ ,  $\sigma^{2(0)}$ , počet  $y$  a celkové množství úplných dat  $x$ , vypisovány jsou jednotlivé iterace:

```
funkce[skutecneu_, skutecnes_, pocatecniu_, pocatecnis_,
pocetPozorovanych_, pocetUplnych_] :=
  algoritmus[pocatecniu, pocatecnis,
  generovani[skutecneu, skutecnes, pocetPozorovanych], pocetUplnych]
```

### A.1.3 Multinomické rozdělení

Funkce generující data z multinomického rozdělení  $Mult_4(n, \{p_1, p_2, p_3, p_4\})$ , vstupními parametry jsou skutečná hodnota  $\theta$  a počet pozorování  $n$ :

```
generovani[t_, pocet_] := (
  SeedRandom[13];
  RandomVariate[MultinomialDistribution[
    pocet, {1/2 + t/4, (1 - t)/4, (1 - t)/4, t/4}]
  ]
)
```

Funkce pro výpočet nového odhadu  $\theta^{(m+1)}$ , vstupními parametry jsou současný odhad parametru  $\theta^{(m)}$  a pozorovaná data  $y$ :

```
novyOdhad[soucasny_, y_] := (soucasny/(2 + soucasny)*y[[1]] +
  y[[4]])/(soucasny/(2 + soucasny)*y[[1]] + y[[2]] + y[[3]] + y[[4]]
)
```

Algoritmus pro pozorovaná data  $y$ , vstupními parametry jsou počáteční odhad  $\theta^{(0)}$  a  $y$ , funkce vrací jednotlivé iterace:

```
algoritmus[pocatecni_, data_] :=
  Module[{pocet = 0, pomocna, odhad = pocatecni},
    Print[{pocet, odhad, 2/(2 + odhad)*data[[1]],
    odhad/(2 + odhad)*data[[1]]};
    While[Abs[odhad - novyOdhad[odhad, data]] > 0.0001,
      pocet++;
      Print[{pocet, novyOdhad[odhad, data], 2/(2 + odhad)*data[[1]],
      odhad/(2 + odhad)*data[[1]]};
      odhad = novyOdhad[odhad, data];]
  ]
```