Metody odhadování parametrů.
Metoda momentů. Maximálně věrohodné odhady.

Petr Pošík

Části dokumentu jsou převzaty (i doslovně)
z *Mirko Navara: Pravděpodobnost a matematická statistika*,
`https://cw.felk.cvut.cz/lib/exe/fetch.php/courses/a6m33ssl/pms_print.pdf`
s laskavým svolením autora.

Teorie odhadů 2
Teorie odhadů . . . 3
Odhad parametrů . . . 4

Metoda momentů 5
Metoda momentů . . . 6
Př: MME/Alt . . . 7
Př: MME/Norm . . . 8
Př: MME/Unif . . . 9
Použitelnost MME . . . 10

MLE 11
Motivace . . . 12
Pr. vs L . . . 13
Věrohodnost . . . 14
MLE . . . 15
Př: MLE/Alt . . . 16
Př: MLE/Norm . . . 17
Př: MLE/Unif . . . 18
Př: MLE/Disc . . . 19
Použitelnost MLE . . . 20

Teorie odhadů

Teorie odhadů je jedna z větví statistiky: *na základě pozorovaných dat (náhodného výběru)* se snaží *odhadnout hodnoty parametrů* procesu, z něhož data pocházejí.

- V teorii řízení se obdobné úloze hledání parametrů dynamických systémů říká *identifikace systému*.
- V dalším předpokládáme, že proces můžeme popsat pravděpodobnostním modelem daného typu s několika „málo“ parametry.
- **Cíl:** najít v jistém smyslu *optimální odhad* (statistiku, postup odhadu, *estimator*), ideálně snadno implementovatelný, který z naměřených dat poskytne realizaci odhadu (odhad parametru, *estimate*) jistého parametru modelu.

Příklady:

- Odhad volebního výsledku jisté politické strany (tj. procento aktivních voličů, které ji bude volit). Volební výsledek je neznámý parametr, odhad je založen na malém náhodném vzorku voličů.
- Radar počítá vzdálenost k objektům (letadla, lodě) na základě časového intervalu, který uplyne mezi vysláním signálu a detekcí jeho odrazu.
  - V zašuměném signálu, který radar přijímá, je ale těžké rozpoznat, zda se jedná o odraz nebo ne. Potřebujeme nějakou metodu odhadu pro detekci této události.
  - Když už odraz detekujeme, změřená délka intervalu je náhodně rozdělená, i ji je třeba nějak odhadnout.

Úloha odhadu parametrů

Náhodná veličina X má *rozdělení parametrizované vektorem parametrů* $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_i)$:

- Pravděpodobostní funkci značíme

  $$p_X(x) = p_X(x|\boldsymbol{\theta}),$$

  abychom zdůraznili závislost na parametru (parametrech). Příklady:
  - $X \sim Ber(q)$: $p_X(x) = p_X(x|q)$
  - $X \sim Unif(a,b)$: $p_X(x) = p_X(x|a,b)$
  - $X \sim N(\mu,\sigma^2)$: $p_X(x) = p_X(x|\mu,\sigma^2)$
- **Parametrický prostor** $\Pi \subseteq \mathbb{R}^i$ je množina všech přípustných hodnot parametrů, tj. $\boldsymbol{\theta} \in \Pi$:
  - $X \sim Ber(q)$: $\boldsymbol{\theta} = q$, $\boldsymbol{\theta} \in \Pi = \langle 0,1 \rangle$
  - $X \sim Unif(a,b)$: $\boldsymbol{\theta} = (a,b)$, $\boldsymbol{\theta} \in \Pi = \mathbb{R}^2$
  - $X \sim N(\mu,\sigma^2)$: $\boldsymbol{\theta} = (\mu,\sigma^2)$, $\boldsymbol{\theta} \in \Pi = \mathbb{R} \times \mathbb{R}^+$
- Hledáme odhad $\widehat{\boldsymbol{\Theta}} = (\widehat{\Theta}_1, \ldots, \widehat{\Theta}_i)$, resp. realizaci odhadu $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \ldots, \widehat{\theta}_i)$ pomocí realizace $\boldsymbol{x} = (x_1, \ldots, x_n)$, který je z jistého hlediska optimální.

Metody odhadování:

- metoda momentů (MME), metoda maximální věrohodnosti (MLE), metoda maximální aposteriorní pravděpodobností (MAPE), Bayesovské odhady, …
- Související: metoda nejmenších čtverců (LSE), metoda nejmenších absolutních odchylek, …

Metoda momentů

Pro náhodnou veličinu X je **k-tý obecný moment**, $k \in \mathbb{N}$, definován jako $\mathrm{E}\, X^k$.

- Připomeňme: k výpočtu střední hodnoty potřebujeme pravděpodobnostní funkci nebo hustotu, zde parametrizovanou.
- k-tý obecný moment n.v. X je tak funkcí parametrů pravděpodobnostního modelu:

$$\mathrm{E}\, X^k = \mathrm{E}\, X^k(\boldsymbol{\theta}).$$

Z náhodného výběru lze spočítat **výběrový k-tý obecný moment**:

$$m_{X^k} = \frac{1}{n}\sum_{j=1}^{n} x_j^k.$$

Metoda momentů doporučuje realizaci odhadu $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \ldots, \widehat{\theta}_i)$ takovou, že

$$\mathrm{E}\, X^k(\widehat{\theta}_1, \ldots, \widehat{\theta}_i) = m_{X^k}, \qquad k = 1, 2, \ldots \quad .$$

K jednoznačnému určení i proměnných (parametrů) obvykle potřebujeme (prvních) i rovnic pro $k = 1, 2, \ldots, i$.

Příklad: MME pro alternativní rozdělení

Příklad: Odhadněte prevalenci obezity v populaci v ČR (obezita: BMI > 30) z realizace náhodného výběru $\boldsymbol{x} = (x_1, \ldots, x_n)$, kde z $n = 50$ testovaných jedinců bylo prvních $k = 14$ obézních.

Řešení metodou momentů: Náhodná veličina X vyjadřuje, zda je osoba obézní nebo ne (nabývá hodnot 0 – "normální" a 1 – "obézní"). Má alternativní (Bernoulliho) rozdělení $X \sim Ber(q)$, kde q je hledaná prevalence. První centrální moment proměnné s alternativním rozdělením je

$$\mathrm{E}\, X = q$$

Máme k dispozici realizaci náhodného výběru $\boldsymbol{x} = (x_1, \ldots, x_n), n = 50$. Výběrový první obecný moment je

$$m_X = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Dle metody momentů máme najít takový odhad $\widehat{q}$, který zajistí rovnost momentu odhadovaného rozdělení s výběrovým momentem:

$$\mathrm{E}\, X = m_X \qquad \Longrightarrow \qquad \widehat{q} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{k}{n}$$

$$\widehat{q} = \frac{k}{n} = \frac{14}{50} = 0.28$$

Příklad: MME pro normální rozdělení

Příklad: Z realizace náhodného výběru $\boldsymbol{x} = (x_1, \ldots, x_n)$ z normálního rozdělení $\mathrm{N}(\mu, \sigma^2)$ odhadněte parametry μ a σ^2.

Řešení metodou momentů: Použijeme první 2 centrální momenty normálního rozdělení.

$$\begin{aligned} \mathrm{E}\, X &= \mu \\ \mathrm{E}\, X^2 &= (\mathrm{E}\, X)^2 + DX = \mu^2 + \sigma^2 \end{aligned}$$

Dle metody momentů máme najít takové odhady, které zajistí rovnost momentů odhadovaného rozdělení s výběrovými momenty:

$$\mathrm{E}\, X = m_X \quad \Longrightarrow \quad \widehat{\mu} = \frac{1}{n}\sum_{j=1}^{n} x_i$$

$$\mathrm{E}\, X^2 = m_{X^2} \quad \Longrightarrow \quad \widehat{\mu}^2 + \widehat{\sigma^2} = \frac{1}{n}\sum_{j=1}^{n} x_i^2$$

Výsledek:

$$\begin{aligned} \widehat{\mu} &= \bar{\boldsymbol{x}}, \\ \widehat{\sigma^2} &= \frac{1}{n}\sum_{j=1}^{n} x_i^2 - \widehat{\mu}^2 = \frac{1}{n}\sum_{j=1}^{n} x_i^2 - \bar{\boldsymbol{x}}^2, \end{aligned}$$

což je alternativní (vychýlený konzistentní) odhad rozptylu.

Příklad: MME pro rovnoměrné diskrétní rozdělení

Problém německých tanků (wikipedia): během 2. světové války spojenci používali statistické metody pro odhad velikosti německého arzenálu. Odhadněte, kolik raket bylo vyrobeno, víte-li, že existují rakety se sériovými čísly $\boldsymbol{x} = (5, 10, 12, 27, 28, 30, 39, 49, 73, 77)$ (byly již odpáleny, nebo byly zabaveny spojenci).

Řešení metodou momentů: Předpokládáme, že existují sériová čísla $1, \ldots, N$, kde N je hledaná velikost arzenálu. Střední hodnota $\mathrm{E}\, X$ proměnné s rovnoměrným diskrétním rozdělením s minimem a a maximem b je

$$\mathrm{E}\, X = \frac{a+b}{2} = \frac{1+N}{2}$$

Dle metody momentů:

$$\mathrm{E}\, X = m_X \quad \Longrightarrow \quad \begin{aligned} \frac{1+\widehat{N}}{2} &= \bar{\boldsymbol{x}} \\ \widehat{N} &= 2\bar{\boldsymbol{x}} - 1 = 2 \cdot \frac{350}{10} - 1 = 69 \end{aligned}$$

Co můžete říct o tomto odhadu???

Použitelnost metody momentů

Možné problémy:

1. Řešení soustavy rovnic neexistuje $\Rightarrow$ zkusme ubrat rovnice.
2. Řešení je nekonečně mnoho $\Rightarrow$ zkusme přidat další rovnice.
3. Je více než jedno řešení (např. soustavy kvadratických rovnic).
4. Je jediné řešení, ale je obtížné je nalézt.
5. Soustava je špatně podmíněná (typicky pro velký počet parametrů).
6. Našli jsme jediné řešení, které však *nesplňuje předpoklady*, $\widehat{\boldsymbol{\theta}} \notin \Pi$ (např. parametry nemohou být libovolná čísla) $\Rightarrow$ *NELZE! Vždy kontrolujte řešení!*
7. Všem rovnicím je přikládána stejná důležitost, což bývá nežádoucí (typicky pro velký počet parametrů).
8. MME nelze použít pro nenumerická data (pokud je nelze smysluplně očíslovat).

Výhody:

- Lze použít pro diskrétní, spojité i smíšené rozdělení beze změn.

Metoda maximální věrohodnosti

Věrohodnost: Motivace

Vraťme se k příkladu s obezitou:

- Náhodná veličina $X \sim Ber(q)$; q je prevalence obezity v populaci.
- Výsledek konkrétního experimentu:
  realizace náhodného výběru $\boldsymbol{x} = (1, 0, 0)$, v němž je z $n = 3$ lidí $k = 1$ obézní.
- Odhadněte, zda byl experiment proveden v Japonsku, České republice nebo v USA, pokud je prevalence obezity $q_{\text{Jap}} = 0.03$, $q_{\text{CR}} = 0.28$ a $q_{\text{USA}} = 0.32$.

Jaká je pravděpodobnost, že by experiment měl výše zmíněný výsledek v jednotlivých zemích? Jednotlivé X_i jsou nezávislé, lze psát $p_X(\boldsymbol{x}|q) = \prod_{i=1}^{n} p_{Ber(q)}(x_i) = q^k(1-q)^{n-k}$.

- $p_X(\boldsymbol{x}|q_{\text{Jap}}) = q_{\text{Jap}}^k(1-q_{\text{Jap}})^{n-k} \doteq 0.028$
- $p_X(\boldsymbol{x}|q_{\text{CR}}) = q_{\text{CR}}^k(1-q_{\text{CR}})^{n-k} \doteq 0.145$
- $p_X(\boldsymbol{x}|q_{\text{USA}}) = q_{\text{USA}}^k(1-q_{\text{USA}})^{n-k} \doteq 0.148$

Zdůrazněme: Výše uvedené hodnoty p_X znamenají

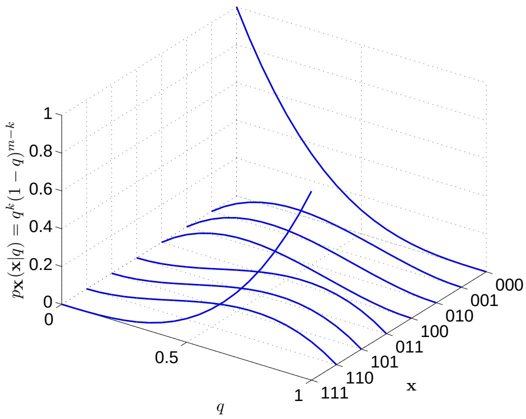
- *pravděpodobnost daného výsledku experimentu za předpokladu, že experiment byl proveden v dané zemi*, nikoli
- ~~pravděpodobnost, že experiment byl proveden v dané zemi za předpokladu, že pozorujeme daný výsledek experimentu~~.

Na základě výsledků se zdá *nejvěrohodnější*, že experiment byl proveden v USA. (Nikoli nejpravděpodobnější: k tomu bychom potřebovali ještě znát počty takových experimentů provedených v jednotlivých zemích.)

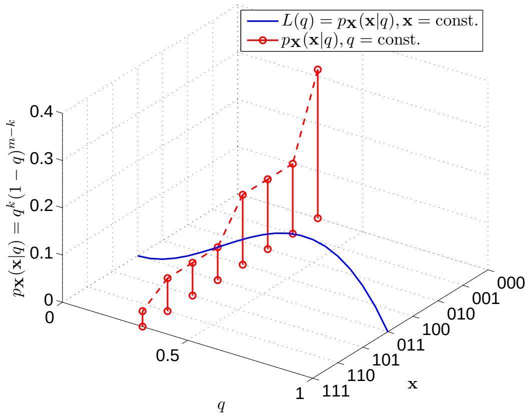
Rozdělení pravděpodobnosti vs. věrohodnostní funkce

Parametrizovanou funkci $p_{\boldsymbol{X}}(\boldsymbol{x}|q) = \prod_{j=1}^{m} p_{Ber(q)}(x_j)$ lze interpretovat různě:

Jako funkci dvou proměnných
($\boldsymbol{x}$ je diskrétní, q je spojitá):



Jako 2 funkce jedné proměnné, druhá je vždy zafixovaná na jisté hodnotě:



- Na ose $\boldsymbol{x}$ jsou vyneseny všechny možné realizace náhodného výběru rozsahu 3 (diskrétní proměnná).
- Na ose q jsou vyneseny všechny možné hodnoty parametru q (spojitá proměnná na intervalu $\langle 0, 1 \rangle$).
- Funkce $p_{\boldsymbol{X}}(\boldsymbol{x}|q)$ popisuje, jak dobře odpovídá realizace n. v. $\boldsymbol{x}$ rozdělení s parametrem q.
- Každý řez pro $q =$ konst. (červeně) je rozdělení pravděpodobnosti $p_{\boldsymbol{X}}(\boldsymbol{x})$. (Pstní dedukce.)
- Každý řez pro $\boldsymbol{x} =$ konst. (modře) je funkce věrohodnosti parametru q, $L(q)$. (Stat. indukce.)

Věrohodnost

Věrohodnost parametrů *diskrétního* rozdělení (vzhledem k realizaci náhodného výběru) je funkce $L\colon \Pi \to \langle 0, 1 \rangle$, parametrů $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_i)$ definovaná jako

$$L(\boldsymbol{\theta}) = p_{\boldsymbol{X}}(\boldsymbol{x}|\boldsymbol{\theta}) = P[X_1 = x_1 \wedge \ldots \wedge X_n = x_n|\boldsymbol{\theta}] =$$
$$= \prod_{j=1}^{n} P[X_j = x_j|\boldsymbol{\theta}] = \prod_{j=1}^{n} p_X(x_j|\boldsymbol{\theta})$$

a je to tedy *pravděpodobnost realizace $\boldsymbol{x}$ náhodného výběru* z diskrétního rozdělení *při hodnotách parametrů $\boldsymbol{\theta}$*.

Věrohodnost parametrů *spojitého* rozdělení (vzhledem k realizaci náhodného výběru) je zcela jiný pojem: každá realizace má nulovou pravděpodobnost, proto místo ní použijeme hustotu pravděpodobnosti. Je to funkce $\Lambda\colon \Pi \to \langle 0, \infty), \Pi \subseteq \mathbb{R}^i$, definovaná jako

$$\Lambda(\boldsymbol{\theta}) = f_{\boldsymbol{X}}(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{j=1}^{n} f_X(x_j|\boldsymbol{\theta})$$

- Hustota f_X musí být spojitá (alespoň na oboru hodnot, jichž náhodná veličina nabývá).
- Věrohodnost $\Lambda(\boldsymbol{\theta})$ může mít rozměr (jednotky), pokud má rozměr i hustota pravděpodobnosti f_X.

Metoda maximální věrohodnosti

Metoda maximální věrohodnosti doporučuje takovou realizaci odhadů $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \ldots, \widehat{\theta}_i)$, která maximalizuje věrohodnost, tj.

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\Pi}{\operatorname{argmax}}\, L(\boldsymbol{\theta}) = \underset{\boldsymbol{\theta}\in\Pi}{\operatorname{argmax}} \prod_{j=1}^{n} p_X(x_j|\boldsymbol{\theta}) \text{ pro diskrétní rozdělení a}$$

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\Pi}{\operatorname{argmax}}\, \Lambda(\boldsymbol{\theta}) = \underset{\boldsymbol{\theta}\in\Pi}{\operatorname{argmax}} \prod_{j=1}^{n} f_X(x_j|\boldsymbol{\theta}) \text{ pro spojitá rozdělení.}$$

Lze maximalizovat buď věrohodnost přímo, nebo její logaritmus (*log-likelihood*), což často vede na snazší výpočet, tj.

$$l(\boldsymbol{\theta}) = \ln L(\boldsymbol{\theta}) = \sum_{j=1}^{n} \ln p_X(x_j|\boldsymbol{\theta}) \text{ pro diskrétní rozdělení a}$$

$$\lambda(\boldsymbol{\theta}) = \ln \Lambda(\boldsymbol{\theta}) = \sum_{j=1}^{n} \ln f_X(x_j|\boldsymbol{\theta}) \text{ pro spojitá rozdělení.}$$

Je třeba vyloučit případy, kdy $p_X(x_j|\boldsymbol{\theta}) = 0$, resp. $f_X(x_j|\boldsymbol{\theta}) = 0$, ty však nevedou na maximum.

Příklad: MLE pro alternativní rozdělení

Příklad: Odhadněte prevalenci obezity v populaci v ČR (obezita: BMI > 30). K dispozici máme realizaci náhodného výběru $\boldsymbol{x} = (x_1, \ldots, x_n)$, kde z $n = 50$ testovaných jedinců bylo prvních $k = 14$ obézních.

Řešení metodou max. věrohodnosti: $X \sim Ber(q)$, kde q je hledaná prevalence. Pravděpodobnost realizace $\boldsymbol{x}$ náhodného výběru v závislosti na parametru q:

$$p_{\boldsymbol{X}}(\boldsymbol{x}|q) = \prod_{j=1}^{n} p_X(x_j|q) = q^k(1-q)^{n-k}$$

Metoda maximální věrohodnosti říká, že máme maximalizovat funkci

$$L(p) = p_{\boldsymbol{X}}(\boldsymbol{x}|q) = q^k(1-q)^{n-k}, \quad \text{nebo} \quad l(p) = \ln L(p) = k \ln q + (n-k)\ln(1-q)$$

kde n a k jsou známé z experimentu (z realizace náh. výběru). Spočtěme derivaci $l(q)$

$$\frac{\partial l(q)}{\partial q} = \frac{k}{q} - \frac{n-k}{1-q}$$

položme ji rovnou nule a vyřešme pro $\widehat{q}$:

$$\frac{k}{\widehat{q}} - \frac{n-k}{1-\widehat{q}} = 0 \qquad \Rightarrow \qquad \widehat{q} = \frac{k}{n}.$$

Maximálně věrohodný odhad populační pravděpodobnosti je shodný s odhadem metodou momentů a se střední hodnotou empirického rozdělení.

Příklad: MLE pro normální rozdělení

Příklad: Z realizace náhodného výběru $\boldsymbol{x} = (x_1, \ldots, x_n)$ z normálního rozdělení $\mathrm{N}(\mu, \sigma^2)$ odhadněte parametry μ a $r = \sigma^2$.

Řešení metodou max. věrohodnosti: Chceme maximalizovat funkci

$$\Lambda(\mu, r) = \prod_j f_{\mathrm{N}(\mu,r)}(x_j) = \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi r}} \exp\left(\frac{-(x_j - \mu)^2}{2r}\right),$$

což je totéž, jako maximalizovat její logaritmus:

$$\lambda(\mu, r) = \ln \Lambda(\mu, r) = \frac{-1}{2r} \sum_{j=1}^{n} (x_j - \mu)^2 - \frac{n}{2} \ln r - \frac{n}{2} \ln 2\pi .$$

Položme derivace rovné nule a vyřešme pro parametry μ a r:

$$\frac{\partial}{\partial \mu} \lambda(\mu, r)\Big|_{\mu=\widehat{\mu}, r=\widehat{r}} = \frac{1}{\widehat{r}} \sum_{j=1}^{n} (x_j - \widehat{\mu}) = \frac{1}{\widehat{r}} \left(\sum_{j=1}^{n} x_j - n\widehat{\mu} \right) = \frac{n}{\widehat{r}} (\bar{x} - \widehat{\mu}) = 0,$$

$$\frac{\partial}{\partial r} \lambda(\mu, r)\Big|_{\mu=\widehat{\mu}, r=\widehat{r}} = \frac{1}{2\widehat{r}^2} \sum_{j=1}^{n} (x_j - \widehat{\mu})^2 - \frac{n}{2\widehat{r}} = \frac{n}{2\widehat{r}^2} \left(\frac{1}{n} \sum_{j=1}^{n} (x_j - \widehat{\mu})^2 - \widehat{r} \right) = 0 .$$

Maxima nabývá funkce věrohodnosti pro hodnoty shodné s odhady metodou momentů:

$$\widehat{\mu} = \bar{x}, \qquad \widehat{r} = \frac{1}{n} \sum_{j=1}^{n} (x_j - \mu)^2 = \frac{1}{n} \sum_{j=1}^{n} (x_j - \bar{x})^2 = \widehat{DX},$$

Příklad: MLE pro rovnoměrné diskrétní rozdělení

Problém německých tanků (wikipedia): během 2. světové války spojenci používali statistické metody pro odhad velikosti německého arzenálu. Odhadněte, kolik raket bylo vyrobeno, víte-li, že existují rakety se sériovými čísly $\boldsymbol{x} = (5, 10, 12, 27, 28, 30, 39, 49, 73, 77)$ (byly již odpáleny, nebo byly zabaveny spojenci).

Řešení metodou max. věrohodnosti: Předpokládáme, že existují sériová čísla $1, \ldots, N$, kde N je hledaná velikost arzenálu. Rovnoměrné rozdělení pravděpodobnosti je konstantní pro X nabývající hodnot $a, a+1, \ldots, b-1, b$.

$$L(a, b) = \prod_{j=1}^{n} \frac{1}{b-a} = \left(\frac{1}{b-a}\right)^n,$$

pokud $\forall x_j : x_j \in \langle a, b \rangle$; jinak je nulová.

Věrohodnost je maximální, pokud $b - a$ je minimální, tj.

$$a = \min_j x_j, \qquad b = \max_j x_j .$$

Pro úlohu německých tanků je $a = 1$ (vyplývá z úlohy), odhadujeme jen horní mez rozdělení, tedy $\widehat{N} = \max_j x_j = 77$.

To je lepší odhad, než poskytuje metoda momentů. Nicméně je to odhad vychýlený, velikost arzenálu většinou podhodnocuje. Nejlepším nestranným odhadem je v tomto případě $\widehat{N} = m + \frac{m-k}{k}$, kde m je výběrové maximum a k je velikost výběru.

Příklad: MLE pro složitější diskrétní rozdělení

Příklad (obdoba bude v DÚ): Komunikační kanál produkuje 3 různé symboly, "A", "B"a "C", s pravděpodobnostmi p_a, p_b a p_c, které *neznáme*. Z výstupu jsme přečetli sekvenci **x** dlouhou $n = a + b + c$ symbolů (a výskytů "A", atd.). Odhadněte p_a, p_b, p_c pomocí MLE.

Řešení metodou max. věrohodnosti: Pravděpodobnost $p(\boldsymbol{x}|p_a, p_b, p_c)$ lze vypočítat jako

$$p(\boldsymbol{x}|p_a, p_b, p_c) = \binom{n}{a} p_a^a \binom{n-a}{b} p_b^b \binom{n-a-b}{c} p_c^c = \frac{n!}{a!b!c!} p_a^a p_b^b p_c^c$$

Protože p_a, p_b a p_c mají definovat rozdělení psti, musí např. $p_c = 1 - p_a - p_b$, takže:

$$p(\boldsymbol{x}|p_a, p_b) = \frac{n!}{a!b!c!} p_a^a p_b^b (1 - p_a - p_b)^c = L(p_a, p_b)$$
$$l(p_a, p_b) = \ln L(p_a, p_b) = \text{konst.} + a \ln p_a + b \ln p_b + c \ln(1 - p_a - p_b)$$

Pro ML odhady musí platit, že derivace l jsou rovné nule:

$$\left.\frac{\partial l(p_a, p_b)}{\partial p_a}\right|_{p_a=\widehat{p}_a, p_b=\widehat{p}_b} = \frac{a}{\widehat{p}_a} - \frac{c}{1 - \widehat{p}_a - \widehat{p}_b} = 0$$
$$\left.\frac{\partial l(p_a, p_b)}{\partial p_b}\right|_{p_a=\widehat{p}_a, p_b=\widehat{p}_b} = \frac{b}{\widehat{p}_b} - \frac{c}{1 - \widehat{p}_a - \widehat{p}_b} = 0$$

Řešením této soustavy rovnic jsou odhady

$$\widehat{p}_a = \frac{a}{n} \qquad \widehat{p}_b = \frac{b}{n} \qquad \widehat{p}_c = \frac{c}{n}$$

Použitelnost metody maximální věrohodnosti

Možné problémy:

1. Je více než jedno řešení. (Může se stát, že různé hodnoty parametrů popisují totéž rozdělení. Vadí to?)
2. Řešení neexistuje. (Věrohodnostní funkce je nespojitá nebo parametrický prostor není uzavřený.)
3. Je jediné řešení, ale je obtížné ho najít. (Lokální extrémy nemusí být globální. Používají se numerické optimalizační algoritmy, obvykle spouštěné z různých úvodních odhadů.)
4. Hodnoty věrohodnosti mohou být velmi malé.
5. *Nelze použít pro smíšené rozdělení!*

Výhody:

- Hledání optima bývá snazší než řešení soustavy rovnic.
- Různým datům je dán společný (srovnatelný) význam.
- Lze použít i na nenumerická data.