2.5 Příklady na metodu maximální věrohodnosti a momentovou metodu

Příklad: Z jediné realizace x_1 náhodné veličiny X, o níž víte, že pochází z normálního rozdělení $N(\mu, 1)$, odhadněte parametr μ momentovou metodou a metodou maximální věrohodnosti.

1. Budou odhady nestranné?
2. Jaké budou mít rozptyly?
3. Budou konzistentní?

Kolik a jakých momentů použijete?

Řešení:

1. Momentovou metodou: Stačí jediný moment: $EX = \mu$ odhadneme pomocí x_1: $\hat{\mu} = m_X = \frac{1}{1}\sum_{i=1}^{1} x_i = x_1$. Odhad $\hat{\mu} = x_1$ je nestranný, protože $EX_1 = \mu$. (Všimněte si, že v posledním výrazu vystupuje nikoli realizace x_1, ale náhodná veličina X_1 - pouze z ní má totiž smysl počítat střední hodnotu, protože nás zajímá, jak se odhad bude chovat pro *různé* realizace, nikoli pro jednu jedinou realizaci x_1.)

   Metodou maximální věrohodnosti: věrohodnostní funkce je

   $$L(x_1, \{\mu, \sigma^2\}) = f_X(x_1|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_1-\mu)^2}{2\sigma^2}},$$

   a my bychom ji chtěli maximalizovat vzhledem k hledaným parametrům (tj. maximalizovat věrohodnost za daných dat a získat hledané neznámé parametry). Protože hledat extrém věrohodnostní funkce není jednoduché, zjednodušíme situaci tím, že místo věrohodnostní funkce budeme maximalizovat její logaritmus (protože logaritmus je ryze monotónní, bod, v němž nabývá věrohodnostní funkce maximum, je přesně týž, jako bod, v němž nabývá maxima logaritmovaná věrohodnostní funkce).

   Logaritmická věrohodnostní funkce je

   $$l(x_1, \{\mu, \sigma^2\}) = -\frac{1}{2} ln(2\pi) - \frac{1}{2} ln(\sigma^2) - \frac{(x_1-\mu)^2}{2\sigma^2}. \tag{18}$$

   Její extrém budeme hledat tak, že položíme její derivaci rovnu 0. Protože se jedná o funkci dvou proměnných, mohli bychom postupně derivovat podle jejích jednotlivých parametrů. Nám bude pro tentokrát stačit nalézt hodnotu parametru μ. Tu nalezneme, položíme-li parciální derivaci podle μ rovnu 0:

   $$\frac{\partial l(x_1, \{\mu, \sigma^2\})}{\partial \mu} = -\frac{2(x_1-\mu)(-1)}{2\sigma^2} = \frac{(x_1-\mu)}{\sigma^2} = 0$$

   a tedy

   $$\hat{\mu} = x_1$$

2. Rozptyl odhadu: $var\hat{\mu} = var x_1 = \sigma^2$.
3. Budou odhady konzistentní? Nelze přímo určit, protože ke konzistenci bychom potřebovali vysledovat chování odhadů pro zvětšující se velikosti výběrů, ale my máme k dispozici pouze jediné pozorování.

Příklad: Ze dvou realizací x_1, x_2 náhodné veličiny X, o níž víte, že pochází z normálního rozdělení $N(\mu, 1)$, odhadněte parametr μ momentovou metodou a metodou maximální věrohodnosti.

Řešení:

1. Momentovou metodou: Opět stačí jediný moment: $EX = \mu$ odhadneme pomocí prvního výběrového momentu: $\hat{\mu} = m_X = \frac{1}{2}\sum_{i=1}^{2} x_i = \frac{x_1+x_2}{2}$. Odhad $\hat{\mu}$ je nestranný, protože $E\frac{X_1+X_2}{2} = \frac{EX_1+EX_2}{2} = \frac{\mu+\mu}{2} = \mu$.

   Metodou maximální věrohodnosti: věrohodnostní funkce je
$$L(\{x_1, x_2\}, \{\mu, \sigma^2\}) = \prod_{i=1}^{2} f_X(x_i|\mu, \sigma^2) = \prod_{i=1}^{2} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$
   potom logaritmická věrohodnostní funkce je
$$\begin{aligned} l(\{x_1, x_2\}, \{\mu, \sigma^2\}) &= \sum_{i=1}^{2} \Big(-\frac{1}{2} ln(2\pi) - \frac{1}{2} ln(\sigma^2) - \frac{(x_i-\mu)^2}{2\sigma^2} \Big) \\ &= -\frac{2}{2} ln(2\pi) - \frac{2}{2} ln(\sigma^2) - \sum_{i=1}^{2} \frac{(x_i-\mu)^2}{2\sigma^2} \end{aligned}$$
   a hodnotu parametru μ nalezneme, položíme-li parciální derivaci podle μ rovnu 0:
$$\begin{aligned} \frac{\partial l(\{x_1, x_2\}, \{\mu, \sigma^2\})}{\partial \mu} &= \sum_{i=1}^{2} \frac{2(x_i-\mu)(-1)}{2\sigma^2} = \sum_{i=1}^{2} \frac{(x_i-\mu)}{\sigma^2} = 0 \\ \hat{\mu} = \frac{\sum_{i=1}^{2} x_i}{2} &= \bar{x} \end{aligned}$$

2. Rozptyl odhadu: $var\hat{\mu} = var\frac{X_1+X_2}{2} = \frac{varX_1+varX_2}{4} = \frac{\sigma^2+\sigma^2}{4} = \frac{\sigma^2}{2}$. Tedy tento odhad bude mít rozptyl poloviční ve srovnání s odhadem pořízeným z jediné realizace.

3. Budou odhady konzistentní? Nelze přímo určit, protože ke konzistenci bychom potřebovali vysledovat chování odhadů pro zvětšující se velikosti výběrů, ale my máme k dispozici pouze dvě pozorování.

Příklad: (Pokračování přechozího příkladu.) Co kdybyste pro odhad parametr použili pouze poslední realizaci x_2?

- Jak se změní odhady?
- Budou lepší, než minulé odhady?
- Budou nestranné?
- Jaké budou mít rozptyly?
- Budou konzistentní?

Řešení: Dostali bychom stejný odhad jako v případě, že máme jedinou realizaci. Odhad střední hodnoty bude nestranný, ale ve srovnání s odhadem založeným na dvou realizacích bude mít dvojnásobný rozptyl.

Příklad: (Pokračování přechozího příkladu.) Co kdybyste pro odhad parametrů použili větší z realizací, tj. $max(x_1, x_2)$? Jak se změní odhady?

Řešení: Odhad tentokrát nebude nestranný - tím, že za odhad bereme větší ze dvou realizací, střední hodnota takového odhadu bude větší, než střední hodnota samotné realizace.

Příklad: Z jediné realizace x_1 náhodné veličiny X o níž víte, že pochází z normálního rozdělení $N(\mu, \sigma^2)$, odhadněte parametry μ a σ^2 momentovou metodou a metodou maximální věrohodnosti.

Řešení:
Momentovou metodou oba parametry z jediné relizace odhadnout nelze.

Metodou maximální věrohodnosti: parametr μ odhadneme hodnotou dané relizace x_1 (viz výše). Parametr σ^2 odhadneme maximalizací logaritmické věrohodnostní funkce (18) vzhledem k σ^2, hledáme tedy extrém $l(x_1, \{\mu, \sigma^2\})$ vzhledem k σ^2, tedy parciální derivaci $l(x_1, \{\mu, \sigma^2\})$ podle σ^2 položíme rovnu nule:

$$\frac{\partial l(x_1, \{\mu, \sigma^2\})}{\partial \sigma^2} = -\frac{1}{2}\frac{1}{\sigma^2} + \frac{(x_1 - \mu)^2}{2(\sigma^2)^2} = 0$$

a po úpravě (vynásobení $2(\sigma^2)^2$)

$$\sigma^2 = (x_1 - \mu)^2$$

a vzhledem k tomu, že $\hat{\mu} = x_1$,

$$\sigma^2 = (x_1 - x_1)^2 = 0.$$

Vidíme, že věrohodnost se maximalizuje pro $\widehat{\sigma^2} = 0$, tedy pro degenerované (singulární) rozdělení s nulovým rozptylem, v němž je všechna hustota soustředěna v jediném bodu x_1.

Souvislost s uvíznutím E-M algoritmu v lokálním minimu odpovídajícím singulárním variančním maticím.

Příklad: Odhad parametrů $N(\mu, \sigma^2)$ z více realizací.

Řešení: viz přednáška

Příklad: Odhad parametrů alternativního rozdělení $Alt(p)$ z realizací $0, 0, 1$.

Řešení: Alternativního rozdělení je popsáno pravděpodobnostmi

$$f_{Alt}(X = 1) = p$$

a

$$f_{Alt}(X = 0) = 1 - p,$$

kde p je pravděpodobnost úspěchu (jevu 1).

Věrohodnostní funkce pak je součin pravděpodobností jednotlivých realizací za dané hodnoty parametru p:

$$\begin{aligned} L(\{0,0,1\}, p) &= \prod_{i=1}^{3} f_{Alt}(x_i|p) \\ &= (1-p) \cdot (1-p) \cdot p \\ &= (1-p)^2 p. \end{aligned}$$

Tuto funkci bychom mohli již přímo maximalizovat (to by vedlo ke kvadratické rovnici), nebo můžeme podobně jako dříve tuto funkci logaritmovat a úlohu si zjednodušit.

Logaritmická věrohodnostní funkce pak je

$$l(\{0,0,1\}, p) = 2ln(1-p) + ln(p)$$

a její derivace podle p položíme rovnu 0:

$$\frac{l(\{0,0,1\}, p)}{\partial p} = 2\frac{-1}{1-p} + \frac{1}{p} = 0$$

a po úpravě (vynásobení $p(1-p)$)

$$\begin{aligned} 2\hat{p} &= (1-\hat{p}) \\ 3\hat{p} &= 1 \\ \hat{p} &= \frac{1}{3} \end{aligned}$$

Odhadem parametru p metodou max. věrohodnosti je tedy $\hat{p} = 1/3$.

Poznámka: ke stejnému výsledku by vedlo i to, pokud bychom daný počet jedniček ve sledu tří nezávislých realizací náhodné alternativní veličiny považovali za jedinou realizaci náhodné veličiny mající binomické rozdělení. Věrohodnostní funkce by pak byla

$$L(\{0,0,1\},p) = \binom{3}{1} p(1-p)^2,$$

což by vedlo ke stejnému výsledku.

Příklad: Házíme mincí, u níž se obáváme, že je falešná: panna údajně padá 2x častěji než orel. Spočtěte věrohodnost tohoto tvrzení na základě pozorování, že padl 2x orel. Spočtěte rovněž věrohodnost pozorovaných dat pro případ, že mince není falešná. Dále z dat odhadněte pravděpodobnost, že padá panna, metodou maximální věrohodnosti a momentovou metodou.

Řešení: Náhodný jev modelujeme alternativním rozdělením s parametrem p vyjadřujícím pravděpodobnost, že padne panna. Pak věrohodnost pozorovaného sledu dvou orlů je:

$$L(x,p) = L(\{0,0\},p) = (1-p)^2.$$

Pro očekávanou falešnou minci, kde $p = \frac{2}{3}$ dostáváme $L(\{0,0\},\frac{2}{3}) = \frac{1}{9}$. Pro férovou minci, kde $p = \frac{1}{2}$ dostáváme $L(\{0,0\},\frac{1}{2}) = \frac{1}{4}$. Věrohodnější tedy je, že mince je férová.

Odhad p metodou maximální věrohodnosti: věrohodnost $L(x,p) = L(\{0,0\},p) = (1-p)^2$ se maximalizuje pro $\hat{p} = 0$.

Odhad p momentovou metodou: První obecný (necentrální) moment alternativního rozdělení

$$M_1^{'} = EX = p$$

položíme roven odpovídajícímu výběrovému obecnému momentu $m_1^{'} = \bar{x}$, tedy:

$$p = M_1^{'} = m_1^{'} = \bar{x}$$

a odtud přímo dostáváme odhad $\hat{p}$:

$$\hat{p} = \bar{x} = 0.$$

Oba odhady si jsou tedy rovny.

Příklad: Metodou maximální věrohodnosti a momentovou metodou odhadněte na základě pozorovaných $x_1, \ldots, x_n$ neznámý parametr λ Poissonova rozdělení:

$$P(X=k) = \frac{\lambda^k e^{-\lambda}}{k!}.$$

Řešení:

Metodou maximální věrohodnosti:

$$\begin{aligned}
L(\{x_1,\dots,x_n\},\lambda) &= \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} \\
\ell(\{x_1,\dots,x_n\},\lambda) &= \sum_{i=1}^{n} (x_i \ln\lambda - \lambda - \ln x_i!) \\
\ell(\{x_1,\dots,x_n\},\lambda) &= \ln\lambda \sum_{i=1}^{n} x_i - n\lambda - \sum_{i=1}^{n} \ln x_i! \\
\frac{\partial \ell(\{x_1,\dots,x_n\},\lambda)}{\partial\lambda} &= \frac{1}{\lambda}\sum_{i=1}^{n} x_i - n = 0 \\
\hat{\lambda} &= \frac{1}{n}\sum_{i=1}^{n} x_i.
\end{aligned}$$

Momentovou metodou: První obecný moment Poissonova rozdělení

$$M_1^{'} = EX = \lambda$$

položíme roven odpovídajícímu výběrovému obecnému momentu $m_1^{'} = \bar{x}$, tedy:

$$\lambda = M_1^{'} = m_1^{'} = \bar{x}$$

a odtud přímo dostáváme odhad

$$\hat{\lambda} = \bar{x}.$$

Oba odhady si jsou tedy rovny.

Příklad: Metodou maximální věrohodnosti a momentovou metodou odhadněte na základě pozorovaných $x_1,\dots,x_m$ neznámý parametr p binomického rozdělení $Bi(n=10,p)$:

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Řešení: Metodou maximální věrohodnosti:

$$\begin{aligned}
L(\{x_1,\dots,x_m\},n=10,p) &= \prod_{i=1}^{m} \binom{n}{x_i} p^{x_i}(1-p)^{n-x_i} \\
\ell(\{x_1,\dots,x_m\},n=10,p) &= \sum_{i=1}^{m} \left(\ln\binom{n}{x_i} + x_i \ln p + (n-x_i)\ln(1-p) \right) \\
\frac{\partial \ell(\{x_1,\dots,x_m\},\lambda)}{\partial\lambda} &= \frac{\sum_{i=1}^{m} x_i}{p} + \frac{\sum_{i=1}^{m}(n-x_i)}{1-p} \\
\frac{\partial \ell(\{x_1,\dots,x_m\},\lambda)}{\partial\lambda} &= \frac{\sum_{i=1}^{m} x_i}{p} + \frac{mn - \sum_{i=1}^{m} x_i}{1-p} = 0 \\
\sum_{i=1}^{m} x_i - p\sum_{i=1}^{m} x_i &= mnp - p\sum_{i=1}^{m} x_i \\
\sum_{i=1}^{m} x_i &= mnp \\
\hat{p} &= \frac{\sum_{i=1}^{m} x_i}{mn}
\end{aligned}$$

Momentovou metodou: První obecný moment binomického rozdělení

$$M_1^{'} = EX = np$$

položíme roven odpovídajícímu výběrovému obecnému momentu $m_1^{'} = \bar{x}$, tedy:

$$np = M_1^{'} = m_1^{'} = \bar{x} = \frac{\sum_{i=1}^{m} x_i}{m}$$

a odtud přímo dostáváme odhad

$$\hat{p} = \frac{\bar{x}}{n} = \frac{\sum_{i=1}^{m} x_i}{mn}$$

Oba odhady si jsou tedy opět rovny.

Příklad: Metodou maximální věrohodnosti a momentovou metodou odhadněte na základě pozorovaných $x_1, \ldots, x_n$ neznámé parametry μ, h rovnoměrného rozdělení $R(\mu, h)$ na intervalu $(\mu - h, \mu + h)$. (Poznamejme, že standardně se rovnoměrné rozdělení značí $Uni(a, b)$ a parametrizuje se nikoli střední hodnotou a rozsahem, ale mezemi intervalu, na kterém je definováno.)

Řešení: Metodou maximální věrohodnosti:

$$L(\{x_1, \ldots, x_n\}, \mu, h) = \prod_{i=1}^{n} \frac{1}{2h} = \left(\frac{1}{2h}\right)^n$$

$$\ell(\{x_1, \ldots, x_n\}, n = 10, p) = n \ln \frac{1}{2h} = -n \ln(2h).$$

Tato věrohodnost se maximalizuje pro co nejmenší h. Ovšem h musí být natolik veliké, aby interval $(\mu - h, \mu + h)$ pokryl všechny pozorování x_i. Tedy vychází, že

$$\hat{\mu} = \frac{1}{2}(max_{i=1}^{n} x_i + min_{i=1}^{n} x_i)$$

$$\hat{h} = \frac{1}{2}(max_{i=1}^{n} x_i - min_{i=1}^{n} x_i)$$

Momentovou metodou: První obecný moment rovnoměrného rozdělení

$$M_1^{'} = EX = \mu$$

položíme roven odpovídajícímu výběrovému obecnému momentu $m_1^{'} = \bar{x}$, tedy:

$$\mu = M_1^{'} = m_1^{'} = \bar{x}.$$

a dostáváme

$$\hat{\mu} = \bar{x}.$$

Víme (nebo vypočítáme), že druhý obecný moment rovnoměrného rozdělení je

$$M_2^{'} = EX^2 = E(X - EX)^2 + (EX)^2 = varX + (EX)^2 = \frac{h^2}{3} + (M_1^{'})^2 = \frac{h^2}{3} + \mu^2.$$

Přitom $varX$ se spočítá přímo z definice jako:

$$varX = E(X - EX)^2 = \int_{-\infty}^{\infty} f(x)(x - EX)^2 dx = \int_{\mu-h}^{\mu+h} f(x)(x - EX)^2 dx = \int_{\mu-h}^{\mu+h} \frac{1}{2h}(x - \mu)^2 dx$$

$$\overset{y=x-\mu}{=} \int_{-h}^{+h} \frac{1}{2h} y^2 dy = \frac{1}{2h}\left[\frac{y^3}{3}\right]_{-h}^{h} = \frac{1}{2h}\left[\frac{h^3}{3} - \frac{(-h)^3}{3}\right] = \frac{1}{2h}\left[\frac{2h^3}{3}\right] = \frac{h^2}{3}.$$

Druhý obecný moment rovnoměrného rozdělení pak položíme roven odpovídajícímu výběrovému obecnému momentu $m_2^{'} = \sum_{i=1}^{n} x_i^2$, tedy:

$$\frac{h^2}{3} + \mu^2 = \sum_{i=1}^{n} x_i^2$$

$$\hat{h} = \sqrt{3 \sum_{i=1}^{n} x_i^2 - \hat{\mu}^2}$$

V tomto případě se tedy odhady momentovou metodou a metodou maximální věrohodnosti liší.