

Errata k bakalářské práci Odhadování na principu věrohodnosti

Eva Březinová

6. září 2023

Tento dokument řeší některé důležité připomínky z oponentského posudku k bakalářské práci *Odhadování na principu věrohodnosti*. Obsahuje úpravy formulací, výpočtů a některá doplnění či upřesnění informací.

1. Maximálně věrohodný odhad (str. 5, řádky 5 a 20)

Původní formulace: Položíme-li skórovou funkci rovnu 0, získáme tzv. *věrohodnostní rovnici*, jejímž řešením je maximálně věrohodný odhad.

Oprava: Za splnění podmínek regularity je logaritmická věrohodnostní funkce diferencovatelná jako funkce proměnné θ . Derivováním podle θ získáme skórovou funkci. Tu položíme rovnu 0. Maximálně věrohodný odhad je řešením této rovnice.

Původní formulace: Můžeme-li soustavu věrohodnostních rovnic zapsat, pak jakýkoliv MLE $\hat{\theta}_{\text{ML}}$ parametru $\theta \in \Theta \subseteq \mathbb{R}^p$, $p \in \mathbb{N}$, je jejím řešením. Často se proto MLE hledá právě jako řešení soustavy věrohodnostních rovnic.

Oprava: Jsou-li splněny předpoklady regularity a věrohodnostní logaritmická funkce je tedy diferencovatelná jako funkce proměnné $\theta = (\theta_1, \dots, \theta_p)^\top$, pak je MLE $\hat{\theta}_{\text{ML}}$ parametru $\theta \in \Theta \subseteq \mathbb{R}^p$, řešením soustavy věrohodnostních rovnic.

2. Tvar logaritmické věrohodnostní funkce (str. 7)

Pro náhodný výběr $\mathbf{X} = (X_1, \dots, X_n)^\top$ z normálního rozdělení s hustotou $f(x; \theta)$, která závisí na parametru $\theta = (\mu, \sigma^2)^\top$ získáme zlogaritmováním věrohodnostní funkce $L(\mathbf{X}; \theta) = \prod_{i=1}^n f(X_i; \theta)$ logaritmickou věrohodnostní funkci

$$\ell(\mathbf{X}; \theta) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{\sum_{i=1}^n (X_i - \mu)^2}{2\sigma^2}.$$

Skórová funkce (vektor parciálních derivací podle proměnných μ a σ^2) má tvar

$$S(\mathbf{X}; \theta) = \left(\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu), -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (X_i - \mu)^2 \right)^\top.$$

Obě složky položíme rovny nule. Získaná soustava věrohodnostních rovnic má řešení $(\mu, \sigma^2)^\top = \left(\overline{X}_n, \frac{\sum_{i=1}^n (X_i - \overline{X}_n)^2}{n} \right)^\top$. Dostáváme tedy maximálně věrohodný odhad parametru $\boldsymbol{\theta} \in \mathbb{R} \times \mathbb{R}^+$ tvaru

$$\hat{\boldsymbol{\theta}}_{\text{ML}} = \left(\overline{X}_n, \frac{n-1}{n} S_n^2 \right)^\top,$$

kde \overline{X}_n značí výběrový průměr a $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X}_n)^2$ značí výběrový rozptyl.

3. Řešitelnost pseudo-věrohodnostních rovnic (str.13)

V definici 9 zavádíme pro náhodný vektor $\mathbf{X} = (X_1, \dots, X_n)^\top$ jednorozměrnou logaritmickou pseudo-věrohodnostní funkci $\ell_1(\mathbf{X}; \theta)$ a dvourozměrnou logaritmickou pseudo-věrohodnostní funkci $\ell_2(\mathbf{X}; \theta)$. V obou případech jde o funkce neznámého parametru $\theta \in \Theta \subseteq \mathbb{R}$. Jednorozměrnou pseudo-skórovou funkci $U_1(\mathbf{X}; \theta)$ a dvourozměrnou pseudo-skórovou funkci $U_2(\mathbf{X}; \theta)$ definujeme jako derivace příslušných logaritmických pseudo-věrohodnostních funkcí podle proměnné θ a pseudo-věrohodným odhadem rozumíme řešení soustavy $U_1(\mathbf{X}; \theta) = 0$ a $U_2(\mathbf{X}; \theta) = 0$. Jde o soustavu dvou rovnic o jedné neznámé a existence řešení tak není zaručena, soustava rovnic je přeurčená. K řešení přeurčených nelineárních rovnic se využívá například Gauss-Newtonova metoda, věnuje se jí článek (Pratt a kol., 2010).

Speciálním případem je situace, kdy složky náhodného vektoru \mathbf{X} tvoří náhodný výběr. Díky nezávislosti platí:

$$\begin{aligned} \ell_1(\mathbf{X}; \theta) &= \ell(\mathbf{X}; \theta), \\ \ell_2(\mathbf{X}; \theta) &= [(1-a)n - 1]\ell(\mathbf{X}; \theta), \end{aligned}$$

kde $\ell(\mathbf{X}; \theta)$ značí logaritmickou věrohodnostní funkci, jež je v práci zavedena v definici 3. Pseudo-skórové funkce budou v tomto případě tvaru

$$\begin{aligned} U_1(\mathbf{X}; \theta) &= \frac{\partial \ell(\mathbf{X}; \theta)}{\partial \theta}, \\ U_2(\mathbf{X}; \theta) &= [(1-a)n - 1] \frac{\partial \ell(\mathbf{X}; \theta)}{\partial \theta}. \end{aligned}$$

V případech, kdy $a \neq 1 - \frac{1}{n}$, jsou pseudo-věrohodnostní rovnice $U_1(\mathbf{X}; \theta) = 0$ a $U_2(\mathbf{X}; \theta) = 0$ ekvivalentní. Pseudo-věrohodný odhad tak hledáme jako řešení libovolné z nich.

V simulační studii v kapitole 2.3, kde hledáme odhad neznámého parametru $\mu \in \mathbb{R}$, na kterém závisí sdružené mnohorozměrné normální rozdělení náhodného vektoru \mathbf{X} , bychom řešením rovnice $U_1(\mathbf{X}; \mu) = 0$ získali odhad shodující se s maximálně věrohodným odhadem, při jehož výpočtu jsme zanedbali závislost složek náhodného vektoru \mathbf{X} a se kterým pseudo-věrohodný odhad porovnáváme. Pro jednoduchost

jsme se omezili pouze na řešení rovnice $U_2(\mathbf{X}; \mu) = 0$. Ta vychází z dvourozměrné logaritmické věrohodnostní funkce $\ell_2(\mathbf{X}; \mu)$, která vychází ze sdružených hustot dvojic složek náhodného vektoru \mathbf{X} a díky tomu odráží závislost mezi složkami. Pro zjednodušení výpočtu využíváme pouze dvourozměrnou pseudo-věrohodnostní funkci $\ell_2(\mathbf{X}; \theta)$ také v ilustračním příkladu na straně 13. Dovolit si to můžeme díky tomu, že existuje více přístupů popisujících pseudo-věrohodnost. Další z nich využívají sdružená rozdělení většího množství marginálních rozdělení nebo rozdělení podmíněná, například v článku (Arnold a Strauss, 1991) nebo (Gourieroux a kol., 1984).

4. Odvození sdruženého rozdělení (str.17)

Chceme určit sdružené rozdělení, ze kterého pochází náhodný výběr dvourozměrných náhodných vektorů $(Y_1, X_1)^\top, \dots, (Y_n, X_n)^\top$. Pro náhodný vektor $(Y_i, X_i)^\top$, kde $i \in \{1, \dots, n\}$ platí, že náhodná veličina Y_i podmíněná náhodnou veličinou X_i má Poissonovo rozdělení s parametrem λX_i , kde $\lambda > 0$, a pro náhodnou veličinu X_i platí $X_i \sim R[0, 1]$.

Původní formulace: Pomocí vztahů

$$\begin{aligned} P[Y = k \wedge X = x] &= P[Y = k | X = x] \cdot P[X = x], \\ P[Y = k | X = x] &= \frac{(\lambda x)^k}{k!} \exp(-\lambda x), \\ P[X = x] &= \mathbf{I}_{[0, 1]}(x), \end{aligned}$$

kde $k \in \{0, 1, 2, \dots\}$ a $x > 0$, určíme rozdělení, ze kterého pochází náhodný výběr $(Y_1, X_1)^\top, \dots, (Y_n, X_n)^\top$.

Oprava: Původně jsme zamýšleli uvažovat náhodnou veličinu X s diskrétním rovnoměrným rozdělením a vztahy byly určeny pro tuto situaci. Změnou rozdělení náhodné veličiny X na $R[0, 1]$ vznikly nepřesnosti, neboť náhodná veličina $X \sim R[0, 1]$ má spojitě rozdělení a pravděpodobnost, že nabývá konkrétní hodnoty $x \in \mathbb{R}$ je nulová. První rovnost uvedená v práci platí triviálně, neboť obě strany jsou nulové. V druhé rovnosti problém není, avšak u třetí, jak správně upozornil oponent, došlo k záměně pravděpodobnosti a hustoty a správně je zápis $f_X(x) = \mathbf{I}_{[0, 1]}(x)$ pro $x \in \mathbb{R}$.

Při výpočtu sdružené hustoty $f_{Y, X}(y, x)$ náhodného vektoru $(Y, X)^\top$ vůči referenční míře $\mu \times \nu$, kde μ je čítecí míra na \mathbb{N}_0 a ν je Lebesgueova míra na \mathbb{R} , využijeme vztah

$$f_{Y|X}(y, x) = \frac{f_{Y, X}(y, x)}{f_X(x)} \cdot \mathbf{I}_{(0, +\infty)}(f_X(x)),$$

jehož úpravou získáme

$$f_{Y, X}(y, x) = f_{Y|X}(y, x) \cdot f_X(x) \cdot \mathbf{I}_{(0, +\infty)}(f_X(x)). \quad (1)$$

Hustota náhodné veličiny X má (vzhledem k Lebesgueově míře ν) tvar $f_X(x) = \mathbf{I}_{[0,1]}(x)$. Podmíněná hustota náhodné veličiny Y za X má (vzhledem k míře $\mu \times \nu$) tvar

$$f_{Y|X}(y, x) = \frac{(\lambda x)^y}{y!} \exp(-\lambda x), \quad y \in \mathbb{N}_0, x \in [0, 1], \lambda > 0.$$

Dosazením do vzorce (1) získáme, že sdružená hustota náhodného vektoru $(Y, X)^\top$ vzhledem k míře $\mu \times \nu$, je tvaru

$$f_{Y, X}(y, x; \lambda) = \begin{cases} \frac{(\lambda x)^y}{y!} \exp(-\lambda x), & \text{když } y \in \mathbb{N}_0, x \in [0, 1], \\ 0, & \text{jinak,} \end{cases}$$

kde $\lambda > 0$ je neznámý parametr.

5. Odhad založený na kvazi-věrohodnosti (str. 20)

Kvazi-skórová funkce náhodného vektoru $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$, kde $\forall i \in \{1, \dots, n\}$ platí $Y_i|X_i \sim \text{Poiss}(\lambda X_i)$ a $X_i \sim R[0, 1]$, má tvar

$$\sum_{i=1}^n \frac{Y_i - \lambda X_i}{\lambda X_i} = 0.$$

Jejím řešením je $\lambda = \frac{1}{n} \sum_{i=1}^n \frac{Y_i}{X_i}$ a odhad získaný pomocí kvazi-věrohodnosti je tedy tvaru

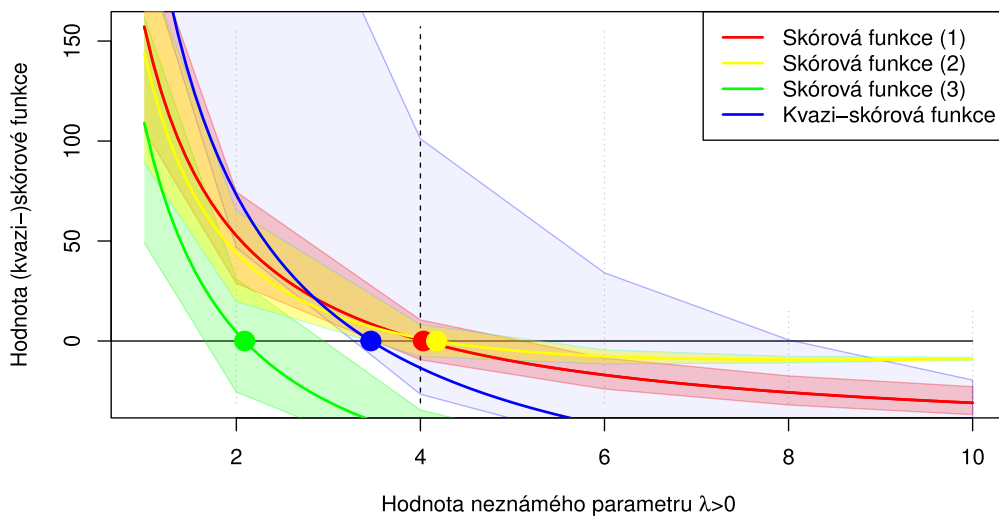
$$\hat{\lambda}_{\text{KL}} = \frac{1}{n} \sum_{i=1}^n \frac{Y_i}{X_i}.$$

Pomocí simulace, která odpovídá popisu na straně 20 v práci, porovnáme kvazi-věrohodný odhad s odhady založenými na maximální věrohodnosti. Výsledky jsou zaneseny do tabulky 1. Odhady jsou porovnávány pomocí nové sady dat oproti tabulce v práci na straně 20. Je tomu tak z důvodu chybějící podmínky pro jejich možnou reprodukci. Největší změny nastaly ve čtvrtém řádku, kde byly výsledky určeny pomocí opraveného tvaru kvazi-věrohodného odhadu.

Tabulka 1: Porovnání kvality odhadů neznámého parametru $\lambda > 0$ pomocí maximální věrohodnosti v **modelu 1.1** ($\hat{\lambda}_{\text{ML}(1)}$), **modelu 1.2** ($\hat{\lambda}_{\text{ML}(2)}$) a **modelu 1.3** ($\hat{\lambda}_{\text{ML}(3)}$) a pomocí kvazi-věrohodnosti ($\hat{\mu}_{\text{KL}}$). Porovnání pro různé rozsahy náhodných výběrů $n \in \mathbb{N}$ (s příslušnými směrodatnými odchylkami v závorce) pomocí mean square error kritéria na základě 1000 Monte-Carlo simulací.

	n = 10		n = 20		n = 50		n = 100	
$\hat{\lambda}_{\text{ML}(1)}$	0.803	(1.185)	0.401	(0.562)	0.171	(0.222)	0.075	(0.104)
$\hat{\lambda}_{\text{ML}(2)}$	1.167	(1.610)	0.678	(0.924)	0.262	(0.347)	0.121	(0.159)
$\hat{\lambda}_{\text{ML}(3)}$	4.299	(2.234)	4.221	(1.689)	4.072	(1.039)	4.038	(0.717)
$\hat{\lambda}_{\text{KL}}$	2.240	(10.321)	1.137	(5.921)	0.483	(1.921)	0.258	(0.781)

Nejnižší střední čtvercovou chybu má odhad $\hat{\lambda}_{\text{ML}(1)}$, který vychází z **modelu 1.3**. Výrazně nejhorší je pak odhad $\hat{\lambda}_{\text{ML}(3)}$, který vychází z **modelu 1.3**. U odhadu $\hat{\lambda}_{\text{KL}}$ založeného na kvazi-věrohodnosti lze pozorovat vysoké hodnoty směrodatných odchylek, obzvláště pro nižší rozsahy náhodného výběru. Pro všechny odhady se s rostoucím rozsahem náhodného výběru $n \in \mathbb{N}$ velikost střední čtvercové chyby zmenšuje a tedy přesnost odhadů zlepšuje. Nejvýraznější zlepšení v závislosti na rostoucím rozsahu náhodného výběru lze pozorovat u odhadu $\hat{\lambda}_{\text{KL}}$ založeného na kvazi-věrohodnosti.



Obrázek 1: Porovnání skórové funkce založené na sdruženém rozdělení náhodných vektorů $(Y_i, X_i)^\top$, pro $i \in \{1, \dots, 100\}$ (červená), skórové funkce založené na marginálním rozdělení náhodných veličin Y_i pro $i \in \{1, \dots, 100\}$ (žlutá), skórové funkce náhodného výběru z rozdělení $Poiss(\lambda)$ (zelená) a kvazi-skórové funkce (modrá) pro konkrétní náhodný výběr $(Y_1, X_1)^\top, \dots, (Y_{100}, X_{100})^\top$ s rozsahem $n = 100$. Průsečíky s osou x značí příslušné odhady neznámého parametru λ , jehož skutečná hodnota $\lambda_0 = 4$ je zvýrazněna černě čárkovaně. Spolehlivostní pásy značí plochu mezi minimy a maximy skórových funkcí a kvazi-skórové funkce v bodech 2, 4, 6, 8 a 10 lineárně interpolovanými přes 1000 Monte-Carlo simulací.

Pro rozsah náhodného výběru $n = 100$ výsledek ilustrujeme také na obrázku 1. (V práci je výsledek ilustrován na obrázku na straně 21 pro rozsah náhodného výběru $n = 10$. Spolehlivostní pásy pro upravený odhad založený na kvazi-věrohodnosti by pro takový rozsah byly moc široké. Lépe tak poslouží ilustrace na náhodných výběrech vyššího rozsahu.) I z něj je ihned patrná vysoká nepřesnost odhadu $\hat{\lambda}_{\text{ML}(3)}$. Lze si také všimnout, že spolehlivostní pás kvazi-skórové funkce je oproti těm, které přísluší skórovým funkcím, širší.

6. Interpretace numerických výsledků (str. 24)

V kapitole 2.3 porovnáváme pomocí simulace odhady neznámého parametru $\mu \in \mathbb{R}$ založené na maximální věrohodnosti a pseudo-věrohodnosti náhodného vektoru

\mathbf{X} s n -rozměrným normálním rozdělením $N_n(\mu, \Sigma)$, kde $\mu = (\mu, \dots, \mu)^\top \in \mathbb{R}^n$ a $\Sigma \in \mathbb{R}^{n \times n}$ je popsána na straně 21. Při výpočtu maximálně věrohodných odhadů $\hat{\mu}_{\text{ML}(1)}$ a $\hat{\mu}_{\text{ML}(2)}$ zanedbáváme vzájemnou závislost složek náhodného vektoru \mathbf{X} , což je důležité dodat k interpretaci výsledků, například v popisu tabulky na straně 24. Kdykoliv tedy v kapitole 2.3 mluvíme o maximálně věrohodném odhadu, máme na mysli maximálně věrohodný odhad založený na náhodném výběru z normálního rozdělení $N(\mu, 1)$.

7. Vlastnosti odvozených odhadů (doplnění)

V případě kvazi-věrohodných odhadů (kapitola 1.2) a pseudo-věrohodných odhadů (kapitola 1.3) platí, že za splnění určitých technických podmínek regularity jsou odhady neznámých parametrů konzistentní a asymptoticky normální. Více o chování a statistických vlastnostech odhadů založených na kvazi-věrohodnosti lze nalézt v článku (Douc a kol., 2017). Vlastnostem odhadů založených na pseudo-věrohodnosti se věnuje článek (Coeurjolly a Drouilhet, 1998).

Seznam použité literatury

- ARNOLD, B. C. a STRAUSS, D. (1991). Pseudolikelihood Estimation: Some Examples. *Sankhyā: The Indian Journal of Statistics, Series B (1960-2002)*, **53**(2), 233–243.
- COEURJOLLY, J. F. a DROUILHET, R. (1998). Gauss–Newton and full Newton methods in frequency–space seismic waveform inversion. *Geophysical Journal International*, **133**(2), 341–362.
- DOUC, R., FOKIANOS, K. a MOULINES, E. (2017). Asymptotic properties of quasi-maximum likelihood estimators in observation-driven time series models. *Electronic Journal of Statistics*, **11**(2), 2707–2740.
- GOURIEROUX, C., MONFORT, A. a TROGNON, A. (1984). Econometrica. *Sankhyā: The Indian Journal of Statistics, Series B (1960-2002)*, **52**(3), 681–700.
- PRATT, R. G., SHIN, C. a HICK, G. J. (2010). Asymptotic properties of the maximum pseudo-likelihood estimator for stationary Gibbs point processes including the Lennard-jones model. *Electronic Journal of Statistics*, **4**(none), 677–706.