MATEMATICKO-FYZIKÁLNÍ FAKULTA
Univerzita Karlova

BAKALÁŘSKÁ PRÁCE

Eva Březinová

Odhadování na principu věrohodnosti

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Matúš Maciak Ph.D.
Studijní program: Obecná matematika

Praha 2023

Prohlašuji, že jsem tuto bakalářskou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů. Tato práce nebyla využita k získání jiného nebo stejného titulu.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V .... dne ....

Podpis autora

Ráda bych poděkovala vedoucímu práce RNDr. Matúši Maciakovi Ph.D. za rady a čas věnovaný konzultacím. Dále bych chtěla poděkovat Rozálii Kluvancové a Marku Bedřichovi za příjemný společně strávený čas u psaní bakalářských prací.

Název práce: Odhadování na principu věrohodnosti

Autor: Eva Březinová

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Matúš Maciak Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: V této práci bude popsána metoda maximální věrohodnosti, která slouží k odhadování neznámých parametrů, na nichž závisí pravděpodobnostní rozdělení pozorovaných dat. Dále budou představeny metody z ní odvozené. Pozornost bude zaměřena převážně na kvazi-věrohodnost a pseudo-věrohodnost. Následovat bude také stručný popis profilové věrohodnosti, empirické věrohodnosti a podmíněné věrohodnosti. Součástí práce je simulační studie, ve které pomocí střední čtvercové chyby porovnáme kvalitu odhadů získaných na základě maximální věrohodnosti a kvazi-věrohodnosti, případně na základě maximální věrohodnosti a pseudo-věrohodnosti.

Klíčová slova: věrohodnost, maximální věrohodnost, kvazi-věrohodnost, pseudo-věrohodnost

Title: Likelihood based estimation

Author: Eva Březinová

Department: Department of Probability and Mathematical Statistics

Supervisor: RNDr. Matúš Maciak Ph.D., Department of Probability and Mathematical Statistics

Abstract: In this thesis we will describe the maximum likelihood method, method of estimating unknown parameters that determine the probability distribution of the observed data. We will also introduce other methods derived from the likelihood. We focus primarily on a quasi-likelihood and a pseudo-likelihood approach. Then we briefly describe profile likelihood, empirical likelihood, and conditional likelihood. The thesis includes a simulation study which compares the quality of the estimators based on the maximum likelihood and the quasi-likelihood or the maximum likelihood and the pseudo-likelihood using the mean squared error.

Keywords: likelihood, maximum likelihood, quasi-likelihood, pseudo-likelihood

Obsah

Úvod

Existuje celá řada metod matematické statistiky, které lze využít k odhadování neznámých parametrů, které popisují teoretický model, pomocí empirických dat.

Maximální věrohodnost vychází z apriorní znalosti hustoty pravděpodobnostního rozdělení pozorovaných dat, které závisí na neznámém parametru, který je odhadován. Je jí věnována první podkapitola teoretické části této práce. Kromě představení hlavní idey a zavedení definic je její využití ilustrováno na konkrétních příkladech. V neposlední řadě jsou v práci nastíněny její výhody a nevýhody. Nevýhody, například vysoká složitost výpočtu nebo nutnost znalosti celé informace o pravděpodobnostním rozdělení pozorovaných dat, nás vedou k jejím modifikacím. Jednou z nich je kvazi-věrohodnost, která je představena v druhé podkapitole a která se opírá pouze o znalost prvních dvou momentů pravděpodobnostního rozdělení pozorovaných dat. Ve třetí podkapitole je popsána pseudo-věrohodnost využívající marginální hustoty náhodných veličin a sdružené hustoty dvourozměrných náhodných vektorů. U obou výše zmíněných metod jsou kromě definic uvedeny i ilustrační příklady. Čtvrtá podkapitola je stručným pojednáním o dalších metodách odvozených z maximální věrohodnosti nebo ní inspirovaných. Objevuje se zde profilová věrohodnost, empirická věrohodnost a podmíněná věrohodnost.

Následuje empirická část obsahující dvě simulační studie, jejichž cílem je porovnání odhadů založených nejprve na maximální věrohodnosti a na kvazi-věrohodnosti (v případě, kdy známe první dva momenty) a posléze na maximální věrohodnosti a na pseudo-věrohodnosti (v případě, kdy detekujeme vzájemnou závislost naměřených dat). Kvalita odhadů neznámých parametrů je porovnána pomocí střední čtvercové chyby na základě 1000 Monte-Carlo simulací. Simulace jsou doplněny i o grafické znázornění.

1. Teoretická část

Jedním z hlavních cílů matematické statistiky je odhadování neznámých parametrů na základě naměřených empirických dat. Metod, které lze k hledání těchto odhadů použít, existuje celá řada, např. momentová metoda (popsaná v článku (Wooldridge, 2001)), dále metoda nejmenších čtverců (jejím vlastnostem se věnuje článek (Guo a kol., 2011)), mimo jiné i robustní odhady (jež Wilcox (2012) představuje ve své knize) a v neposlední řadě metodu maximální věrohodnosti, která se poprvé objevila v článku (Fisher, 1997) a jejíž vznik ve svém článku detailně popisuje Aldrich (1997).

V této kapitole bude představena metoda maximální věrohodnosti. Zazní její idea, budou zavedeny definice a na konkrétních příkladech budou ilustrovány výpočty maximálně věrohodných odhadů. Budou vyzdviženy výhody této metody, avšak vzápětí bude prostor věnován i nevýhodám. Ty povedou k jejím modifikacím. Hlavní důraz bude kladen na kvazi-věrohodnost a na pseudo-věrohodnost. Na závěr bude kapitola doplněna také o stručný popis profilové věrohodnosti, empirické věrohodnosti a podmíněné věrohodnosti.

1.1 Maximální věrohodnost

Cílem metody maximální věrohodnosti, jež vychází z apriorní znalosti empirických dat, je odhadnout neznámý parametr popisující teoretický model, na němž rozdělení pozorovaných dat závisí. Tato podkapitola se věnuje pouze situacím, kdy pozorovaná data tvoří náhodný výběr. Obsah podkapitoly se opírá o skripta, jejichž autorem je Nagy (2023).

Nechť $\theta \in \Theta \subseteq \mathbb{R}$ je jednorozměrný neznámý parametr, kde Θ je parametrický prostor. Nechť $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$ je pozorovaný náhodný výběr, jehož sdružená hustota $f_{\boldsymbol{X}}(\boldsymbol{x}; \theta)$ závisí na neznámém parametru $\theta \in \Theta \subseteq \mathbb{R}$, $\boldsymbol{x} \in \mathbb{R}^n$. Marginální hustotu stejně rozdělených náhodných veličin $X_1, \ldots, X_n$ značme $f(x; \theta)$, $\theta \in \Theta \subseteq \mathbb{R}$, $x \in \mathbb{R}$. Díky nezávislosti prvků náhodného výběru lze sdruženou hustotu náhodného vektoru $\boldsymbol{X}$ psát ve tvaru $f_{\boldsymbol{X}}(\boldsymbol{x}; \theta) = \prod_{i=1}^n f(x_i,\theta)$, kde $\theta \in \Theta \subseteq \mathbb{R}$, $\boldsymbol{x} = (x_1, \ldots, x_n)^\top \in \mathbb{R}^n$.

V konkrétní realizaci $\boldsymbol{X} \in \mathbb{R}^n$ je obsažena veškerá informace, kterou o odhadovaném parametru $\theta \in \Theta \subseteq \mathbb{R}$ máme. Na základě ní chceme určit, jak je náš odhad věrohodný. Co si pod pojmem věrohodnost představit? Ptáme se, „s jakou pravděpodobností“, bychom pro danou hodnotu θ, tedy pro případ, kdy náhodný výběr pochází z rozdělení tímto θ určeným, pozorovali tu hodnotu $\boldsymbol{X}$, kterou jsme pozorovali. Dále nás zajímá, pro jaké $\theta \in \Theta \subseteq \mathbb{R}$ je tato „pravděpodobnost“ největší. Takové θ nazveme maximálně věrohodný odhad a označíme $\widehat{\theta}_{\mathrm{ML}}$.

Poznámka. Všimněme si, že v předchozím odstavci dáváme slovo pravděpodobnost do uvozovek. Máme-li náhodný výběr s hustotou vůči čítací míře, je možné uvozovky vynechat a skutečně platí

$$f_{\boldsymbol{X}}(\boldsymbol{x}; \theta) = P_\theta(\boldsymbol{X} = \boldsymbol{x}).$$

Tvoří-li však $\boldsymbol{X}$ náhodný výběr z rozdělení s hustotou vůči nějaké σ-konečné míře

ν, zajímá nás výraz

$$f_{\boldsymbol{X}}(\boldsymbol{x};\,\theta) = \lim_{\varepsilon\to 0+} \frac{P_\theta(\boldsymbol{X} \in B(\boldsymbol{x},\varepsilon))}{\nu(B(\boldsymbol{x},\varepsilon))}, \text{ pro } \nu\text{-skoro všechna } \boldsymbol{x} \in \mathbb{R}^n,$$

kde $B(\boldsymbol{x},\varepsilon)$je uzavřená koule v R^n se středem $\boldsymbol{x}$ a poloměrem ε.

Definice 1 (věrohodnostní funkce)**.** *Buď $\boldsymbol{X} = (X_1,\ldots,X_n)^\top$ náhodný výběr, jehož distribuční funkce závisí na parametru $\theta \in \Theta \subseteq \mathbb{R}$. Předpokládejme, že marginální hustota $f(x;\,\theta)$, kde $x \in \mathbb{R}$, náhodné veličiny X_i, kde $i \in \{1,\ldots,n\}$, existuje vůči nějaké σ-konečné míře ν pro každé $\theta \in \Theta \subseteq \mathbb{R}$. Nechť $f_{\boldsymbol{X}}(\boldsymbol{x};\,\theta)$ je sdružená hustota náhodného vektoru $\boldsymbol{X}$, pro $\theta \in \Theta \subseteq \mathbb{R}$ a $\boldsymbol{x} \in \mathbb{R}^n$. Pak pro konkrétní realizaci $\boldsymbol{X} = (X_1,\ldots,X_n)^\top \in \mathbb{R}^n$ funkci $L : \Theta \to [0,\infty)$ danou předpisem*

$$L(\boldsymbol{X};\,\theta) = f_{\boldsymbol{X}}(\boldsymbol{X};\,\theta) = \prod_{i=1}^{n} f(X_i;\,\theta)$$

nazýváme věrohodnostní funkce parametru $\theta \in \Theta \subseteq \mathbb{R}$ v bodě $\boldsymbol{X} \in \mathbb{R}^n$.

Definice 2 (maximálně věrohodný odhad)**.** *Libovolnou hodnotu $\widehat{\theta}_{ML} \in \Theta \subseteq \mathbb{R}$, která maximalizuje věrohodnostní funkci vzhledem k argumentu $\theta \in \Theta \subseteq \mathbb{R}$, nazveme maximálně věrohodný odhad. Lze tedy psát*

$$\widehat{\theta}_{ML} = \underset{\theta\in\Theta}{\operatorname{argmax}} \quad L(\boldsymbol{X};\,\theta).$$

Maximálně věrohodný odhad, pro který budeme dále v textu často používat zkratku MLE, je měřitelnou funkcí proměnné $\boldsymbol{X}$ a nezávisí na $\theta \in \Theta \subseteq \mathbb{R}$. Můžeme se na něj dívat dvěma způsoby. Buď ho uvažujeme jako zobrazení $\widehat{\theta}_{\mathrm{ML}} : \mathbb{R}^n \to \Theta$, které každému $\boldsymbol{X} \in \mathbb{R}^n$ přiřadí prvek parametrického prostoru Θ tvaru $\widehat{\theta}_{\mathrm{ML}}(\boldsymbol{X})$, kde $\boldsymbol{X} \in \mathbb{R}^n$ je konkrétní realizace náhodného výběru. Maximálně věrohodný odhad je pak pevná hodnota z parametrického prostoru Θ. Alternativně lze na $\boldsymbol{X}$ nahlížet jako na náhodný vektor. V takovém případě je maximálně věrohodný odhad $\widehat{\theta}_{\mathrm{ML}} = \widehat{\theta}_{\mathrm{ML}}(\boldsymbol{X})$ náhodná veličina a jde o statistiku.

Maximalizace věrohodnostní funkce může být kvůli derivování součinu komplikovaná. Pomocí triku lze případnému problému předejít. Určíme nejprve logaritmus věrohodnostní funkce a až pro ten najdeme hodnotu parametru $\theta \in \Theta \subseteq \mathbb{R}$ maximalizující danou funkci. Logaritmus je na intervalu $(0,\infty)$ striktně rostoucí funkce, to zajistí, že i přesto, že se funkční hodnota změní, argument maxima zůstane stejný. Určení maxima bude snazší díky vztahu

$$\log f_{\boldsymbol{X}}(\boldsymbol{x};\,\theta) = \log \prod_{i=1}^{n} f(x_i,\theta) = \sum_{i=1}^{n} \log f(x_i,\theta), \quad \text{kde } \boldsymbol{x} = (x_1,\ldots,x_n)^\top,$$

a faktu, že derivování součtu je výrazně snazší než derivování součinu.

Definice 3 (logaritmická věrohodnostní funkce)**.** *Pro danou realizaci náhodného vektoru $\boldsymbol{X} \in \mathbb{R}^n$ funkci*

$$\ell(\boldsymbol{X};\,\theta) = \log\Big\{L(\boldsymbol{X};\,\theta)\Big\}$$

nazýváme logaritmická věrohodnostní funkce parametru $\theta \in \Theta \subseteq \mathbb{R}$.

Definice 4 (skórová funkce pro $\theta \subseteq \mathbb{R}$)**.** *Předpokládejme, že je logaritmická věrohodnostní funkce $\ell(\boldsymbol{X};\, \theta)$ diferencovatelná jako funkce proměnné $\theta \in \Theta \subseteq \mathbb{R}$, pak funkci*

$$S(\boldsymbol{X};\, \theta) = \frac{\partial \ell(\boldsymbol{X};\, \theta)}{\partial \theta}$$

nazýváme skórová funkce.

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

V situacích, kdy je odhadovaný parametr p-rozměrný, tj. $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^\top \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p$ pro $p \in \mathbb{N}$, má logaritmická věrohodnostní funkce tvar

$$\ell(\boldsymbol{X};\, \boldsymbol{\theta}) = \log\Big\{L(\boldsymbol{X};\, \boldsymbol{\theta})\Big\} = \log\Big\{\prod_{i=1}^{n} f(X_i;\, \boldsymbol{\theta})\Big\}.$$

Je-li navíc diferencovatelná jako funkce proměnné $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^\top \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p$, pak je *skórová funkce* určena jako vektor parciálních derivací a má tvar

$$S(\boldsymbol{X};\, \boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{x};\, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \left(\frac{\partial \ell(\boldsymbol{x}, \boldsymbol{\theta})}{\partial \theta_1}, \ldots, \frac{\partial \ell(\boldsymbol{x}, \theta)}{\partial \theta_p}\right)^\top$$

Soustava rovnic, která vznikne, položíme-li skórovou funkci rovnou nule se nazývá *soustava věrohodnostních rovnic.*

Poznámka. Aby mělo smysl soustavu věrohodnostních rovnic, případně věrohodnostní rovnici pro $p = 1$, zavést, předpokládáme, že logaritmická věrohodností funkce závisí na parametru $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p$, $p \in \mathbb{N}$, případně $\theta \in \Theta \subseteq \mathbb{R}$. V opačném případě by vyšla skórová funkce identicky nulová.

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

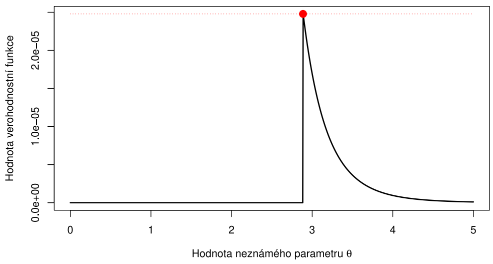
Ilustrace využití

V následující části práce ilustrujeme určení maximálně věrohodného odhadu, na konkrétních příkladech. V prvním z nich určíme MLE přímo z definice, tedy jako hodnotu z parametrického prostoru maximalizující věrohodnostní funkci. V druhém příkladu nalezneme MLE jako řešení soustavy věrohodnostních rovnic.

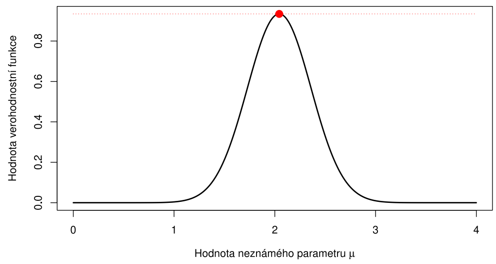
Maximalizace věrohodnostní funkce

Nechť $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$ je náhodný výběr z rovnoměrného rozdělení na intervalu $[0, \theta]$, kde $\theta > 0$ je neznámý parametr, který chceme odhadnout. Značíme $X_i \sim R[0, \theta], \quad \forall\, i \in \{1, \ldots, n\}$. Marginální hustoty náhodných veličin X_i, kde $i \in \{1, \ldots, n\}$, mají tvar

$$f(x;\, \theta) = \frac{1}{\theta}\mathbf{I}_{[0,\, \theta]}(x), \quad x \in \mathbb{R},$$



Obrázek 1.1: Graf věrohodnostní funkce konkrétní realizace náhodného výběru rozsahu 10 z rovnoměrného rozdělení $R[0,3]$.



Obrázek 1.2: Graf věrohodnostní funkce konkrétní realizace náhodného výběru rozsahu 10 z rozdělení $N(3,1)$.

kde definujeme indikátorovou funkci předpisem

$$\mathbf{I}_{[0,\,\theta]}(x) = \begin{cases} 1, & \text{pokud } x \in [0,\theta], \\ 0, & \text{jinak.} \end{cases}$$

Díky nezávislosti složek náhodného výběru vyjádříme sdruženou hustotu náhodného výběru $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$ jako součin marginálních hustot náhodných veličin X_i, kde $i \in \{1, \ldots, n\}$. Bude tvaru

$$f_{\boldsymbol{X}}(\boldsymbol{x};\, \theta) = \prod_{i=1}^{n} f(x_i;\, \theta) = \prod_{i=1}^{n} \frac{1}{\theta} \mathbf{I}_{[0,\,\theta]}(x_i) = \frac{1}{\theta^n} \mathbf{I}_{[\max_{i=1,\ldots,n} x_i \leq \theta]} \mathbf{I}_{[0 \leq \min_{i=1,\ldots,n} x_i]},$$

kde $\boldsymbol{x} = (x_1, \ldots, x_n)^\top \in \mathbb{R}^n$. Věrohodnostní funkce bude tvaru

$$L(\boldsymbol{X};\, \theta) = \frac{1}{\theta^n} \mathbf{I}_{[\max_{i=1,\ldots,n} X_i \leq \theta]} \mathbf{I}_{[0 \leq \min_{i=1,\ldots,n} X_i]},$$

kde $\boldsymbol{X} = (X_1, \ldots, X_n)^\top \in \mathbb{R}^n$ je konkrétní realizace náhodného výběru. Hodnotu věrohodnostní funkce maximalizujeme volbou $\theta = \max_{i=1,\ldots,n} X_i$, neboť pro ostatní hodnoty bude první z indikátorů nulový. Našli jsme maximálně věrohodný odhad $\widehat{\theta}_{\mathrm{ML}} = \max_{i=1,\ldots,n} X_i$.

Poznámka. V tomto případě musíme MLE neznámého parametru $\theta \in \Theta \subseteq \mathbb{R}$ hledat přímo podle definice 2 jako hodnotu maximalizující věrohodnostní funkci. K řešení nelze využít skórovou funkci, neboť věrohodnostní funkce náhodného výběru z rovnoměrného rozdělení není diferencovatelná podle proměnné θ, jak je patrné z grafu na obrázku 1.1.

Řešení soustavy věrohodnostních rovnic

Nechť $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$ je náhodný výběr z normálního rozdělení se střední hodnotou $\mu \in \mathbb{R}$ a rozptylem $\sigma^2 \in \mathbb{R}^+$, tedy $X_i \sim N(\mu, \sigma^2)$, $\forall\, i \in \{1, \ldots, n\}$. Neznámý parametr je dvourozměrný tvaru $\boldsymbol{\theta} = (\mu, \sigma^2)^\top \in \mathbb{R} \times \mathbb{R}^+$. Marginální hustoty náhodných veličin $X_1, \ldots, X_n$, kde $i \in \{1, \ldots, n\}$, jsou tvaru

$$f(x;\, \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\}, \quad \text{kde } x \in \mathbb{R}.$$

Věrohodnostní funkce má díky nezávislosti složek náhodného výběru tvar

$$L(\boldsymbol{X};\,\boldsymbol{\theta}) = \prod_{i=1}^{n} f(X_i;\,\boldsymbol{\theta}), \quad \text{kde } \boldsymbol{X} = (X_1, \ldots, X_n)^\top \in \mathbb{R}^n.$$

Jejím zlogaritmováním získáme logaritmickou věrohodnostní funkcí, která je tvaru

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

Nyní spočteme parciální derivace tohoto výrazu a určíme skórovou funkci. Nejprve derivujeme podle proměnné $\mu \in \mathbb{R}$. Získáme první složku skórové funkce, která je tvaru

$$\frac{\partial \ell(\boldsymbol{X};\,\boldsymbol{\theta})}{\partial \mu} = \sum_{i=1}^{n} X_i - n\mu.$$

Druhou složku skórové funkce získáme derivací logaritmické věrohodnostní funkce $\ell(\boldsymbol{X};\,\boldsymbol{\theta})$ podle proměnné $\sigma^2 \in \mathbb{R}^+$. Má tvar

$$\frac{\partial \ell(\boldsymbol{X};\,\boldsymbol{\theta})}{\partial \sigma^2} = \sum_{i=1}^{n}(X_i-\mu)^2 - n\sigma^2.$$

Obě složky skórové funkce položíme rovny 0 a výsledkem je soustava věrohodnostních rovnic

$$\begin{aligned} \sum_{i=1}^{n} X_i - n\mu &= 0, \\ \sum_{i=1}^{n}(X_i-\mu)^2 - n\sigma^2 &= 0, \end{aligned}$$

jejímž řešením je dvojice $\mu = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X_n}$, $\sigma^2 = \frac{\sum_{i=1}^{n}(X_i-\overline{X_n})^2}{n}$. Maximálně věrohodným odhadem střední hodnoty $\mu \in \mathbb{R}$ je výběrový průměr $\widehat{\mu}_{\mathrm{ML}} = \overline{X_n}$ a maximálně věrohodným odhadem rozptylu $\sigma^2 \in \mathbb{R}^+$ je $\widehat{\sigma^2}_{\mathrm{ML}} = \frac{\sum_{i=1}^{n}(X_i-\overline{X_n})^2}{n}$. Můžeme ho psát také jako $\widehat{\sigma^2}_{\mathrm{ML}} = \frac{n-1}{n}S_n^2$, kde $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X_n})^2$ značí výběrový rozptyl.

Dohromady dostáváme MLE parametru $\boldsymbol{\theta} \in \mathbb{R}\times\mathbb{R}^+$ tvaru

$$\widehat{\boldsymbol{\theta}}_{\mathrm{ML}} = \Big(\overline{X_n}, \frac{n-1}{n}S_n^2\Big)^\top.$$

Poznámka. V tomto příkladu již oproti předchozímu nebyl problém s diferencovatelností věrohodnostní funkce a mohli jsme tak MLE parametru $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}\times\mathbb{R}^+$ hledat jako řešení soustavy věrohodnostních rovnic.

Graf na obrázku 1.2 znázorňuje diferencovatelnost věrohodnostní funkce podle proměnné μ při fixním $\sigma^2 > 0$.

Výhody a nevýhody

Hlavní výhodou metody maximální věrohodnosti je to, že je založena na znalosti hustoty rozdělení, ze kterého pozorovaná data pocházejí. Díky tomu při jejím využití pracujeme s celou informací o tomto rozdělení. Maximálně věrohodný odhad má řadu žádoucích vlastností, které ve skriptech popisuje Nagy (2023, str. 43–48). Ze Zehnovy věty o invarianci je invariantní. Za splnění několika podmínek je také konzistentní, asymptoticky normální a eficientní. Obecně má maximálně věrohodný odhad hezké vlastnosti, pokud rozdělení pozorovaného náhodného výběru patří do exponenciální rodiny rozdělení. Patří do ní řada známých rozdělení (např. exponenciální, Poissonovo, normální či gamma).

Nastávají však i situace, kdy je pro nás využití celé informace o rozdělení náročné, a nutnost této znalosti se stává nevýhodou. Typicky jde o situace, kdy pozorovaná data netvoří náhodný výběr. To způsobí komplikace při určování sdružené hustoty pozorovaného náhodného vektoru. V případech, kdy složky náhodného vektoru netvoří náhodný výběr kvůli tomu, že nejsou stejně rozdělené, jsme stále schopni určit sdruženou hustotu jako součin marginálních hustot. Práce s získanou hustotou je však kvůli rozdílným činitelům výrazně složitější. Problém vyřešíme využitím *kvazi-věrohodnosti*, kterou blíže představíme v podkapitole 1.2. Je založena na vztahu mezi prvními dvěma momenty rozdělení, ze kterého pozorovaná data pocházejí. Tyto vztahy snadno dostaneme ze znalosti marginálních rozdělení. V případech, kdy složky náhodného vektoru nejsou nezávislé, nemůžeme sdruženou hustotu vyjádřit jako součin hustot marginálních a výpočet využívající hustotu vícerozměrného rozdělení pozorovaného náhodného vektoru je, kvůli jejímu často komplikovanému tvaru, příliš náročný. Maximální věrohodnost zaměníme za *pseudo-věrohodnost* využívající marginální hustoty náhodných veličin a sdružené hustoty dvourozměrných náhodných vektorů. Věnujeme jí podkapitolu 1.3.

1.2 Kvazi-věrohodnost

Jednou ze statistických metod (odhadujících neznámé parametry) odvozených z metody maximální věrohodnosti je *kvazi-věrohodnost*. Na rozdíl od metody maximální věrohodnosti nevychází ze znalosti celého pravděpodobnostního rozdělení pozorovaných dat, nýbrž pouze ze znalosti vztahu mezi střední hodnotou a rozptylem. Na základě toho odhaduje neznámý parametr $\theta \in \Theta \subseteq \mathbb{R}$, na kterém střední hodnota a rozptyl pozorovaných dat závisí. Tato podkapitola vychází ze článku (Nelder a Lee, 1992).

Poznámka. Omezíme na odhadování jednorozměrného parametru $\theta \in \Theta \subset \mathbb{R}$, kde Θ značí parametrický prostor.

Vztah mezi střední hodnotou a rozptylem se pro různá rozdělení výrazně liší. Na jedné straně je Poissonovo rozdělení, pro které si jsou střední hodnota a rozptyl rovny. U exponenciálního rozdělení je rozptyl roven druhé mocnině střední hodnoty. Úplným opakem je normální rozdělení, u kterého jsou střední hodnota a rozptyl dva různé parametry bez vzájemného vztahu.

V této kapitole budeme pracovat výhradně s náhodnými veličinami, potažmo náhodnými vektory, jejichž rozdělení patří do exponenciální rodiny rozdělení. Právě příslušnost rozdělení do této rodiny a vztah mezi jeho prvními dvěma momenty jsou veškeré informace, ze kterých při hledání odhadu neznámého parametru $\theta \in \Theta \subset \mathbb{R}$ pomocí kvazi-věrohodnosti vycházíme.

Exponenciální rodina rozdělení

Jelikož příslušnost rozdělení pozorovaného náhodného výběru k exponeciální rodině rozdělení je pro využití kvazi-věrohodnosti nutná, je na místě v první řadě exponenciální rodinu rozdělení definovat. Ve svých skriptech, o která se tato část opírá, tak činí Kulich (2023).

Definice 5 (kanonický tvar hustoty)**.** *Řekneme, že rozdělení náhodné veličiny X přísluší exponenciální rodině rozdělení pokud lze její hustotu (vůči nějaké σ-konečné míře) psát ve tvaru*

$$f(x;\, \theta, \phi) = \exp\left\{\frac{x\theta - b(\theta)}{\phi} + c(x, \phi)\right\}, \tag{1.1}$$

kde $\theta \in \Theta \subseteq \mathbb{R}$ je kanonický parametr, $\phi \in (0,\infty)$ je parametr rozptýlení a b, c jsou vhodné reálné funkce. Výraz (1.1) *nazýváme* kanonický tvar hustoty.

Poznámka. Do exponenciální rodiny rozdělení patří řada běžně používaných rozdělení jako normální, gamma, exponenciální, Poissonovo či alternativní.

Následující lemma popisuje vzájemný vztah střední hodnoty a rozptylu náhodné veličiny X a jejich závislost na neznámém parametru $\theta \in \Theta \in \mathbb{R}$.

Lemma 1. *Nechť rozdělení náhodné veličiny X pochází z exponenciální rodiny rozdělení. Potom existuje vytvořující funkce $P_X(t) = \mathsf{E}\, e^{tX}$ náhodné veličiny X. Je navíc konečná a platí $P_X(t) = \exp\left\{\frac{b(\theta)+t\phi}{\phi} - b(\theta)\right\}$.*

Je-li funkce $b(\theta)$ dvakrát spojitě diferencovatelná, pak je dvakrát spojitě diferencovatelná i funkce $P_X(t)$ v bodě $t = 0$ a platí

$$\mathsf{E}\, X = b'(\theta), \tag{1.2}$$

$$var\, X = \phi b''(\theta). \tag{1.3}$$

Důkaz. Kulich (2023, str. 14).

□

Poznámka. Jelikož platí, že $\operatorname{var} Y = \phi b''(\theta) > 0$ a $\phi > 0$, je b ryze konvexní funkce a funkce b' je ryze rostoucí. Funkce b' má dobře definovanou inverzní funkci a existuje tzv. rozptylová funkce V splňující

$$b''(\theta) = V(b'(\theta)), \quad \text{nebo}$$
$$b''((b')^{-1}(\mu)) = V(\mu).$$

Definice 6 (rozptylová funkce)**.** *Buď X náhodná veličina, jejíž rozdělení patří do exponenciální rodiny rozdělení. Funkci $V(\mu)$ splňující $var_\theta(X) = \phi V(\mu)$, kde ϕ je parametr rozptýlení, nazveme* rozptylová funkce.

Odhadování parametru

Nyní máme veškeré nástroje potřebné k zavedení *kvazi-věrohodnosti.* Hledání odhadu neznámého parametru pomocí kvazi-věrohodnosti (jako řešení kvazi-věrohodnostní rovnice) nachází analogii u maximální věrohodnosti (řešení věrohodnostní rovnice) s tím rozdílem, že při využití kvazi-věrohodnosti je postačující znalost prvních dvou momentů rozdělení pozorovaných dat. Dostáváme se k definici *kvazi-věrohodnosti*, která je v určitém smyslu aproximací logaritmické věrohodnostní funkce.

Definice 7 (kvazi-věrohodnost)**.** *Nechť $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$ je náhodný výběr z rozdělení, jež patří do rodiny exponenciálních rozdělení. Nechť $\mu \in \mathbb{R}$ značí střední hodnotu tohoto rozdělení, $\phi > 0$ je parametr rozptýlení a V je příslušná rozptylová funkce, tedy $var_\theta(X_i) = \phi V(\mu)$, kde $i \in \{1, \ldots, n\}$. Pak pro konkrétní realizaci $\boldsymbol{X} = (X_1, \ldots, X_n)^\top \in R^n$, definujme* kvazi-věrohodnost *vztahem*

$$Q(\boldsymbol{X};\, \boldsymbol{\mu}) = \sum_{i=1}^{n} \int_{x_i}^{\mu} \frac{X_i - t}{\phi V(\mu)}\, dt,$$

kde $\boldsymbol{\mu} = (\mu, \ldots, \mu)^\top \in \mathbb{R}^n$.

Poznámka. Často (například u GLM regresí) nastává situace, kdy jsou složky náhodného vektoru $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$ vzájemně nezávislé, ale jejich rozdělení mají různou střední hodnotu. Nechť náhodná veličina X_i má marginální rozdělení se střední hodnotou μ_i, parametrem rozptýlení $\phi > 0$ a rozptylovou funkcí V splňující $\operatorname{var}_\theta(X_i) = \phi V(\mu_i)$, pro $i \in \{1, \ldots, n\}$. V takovém případě definujeme kvazi-věrohodnost náhodného vektoru $\boldsymbol{X}$ jako součet kvazi-věrohodností jednotlivých složek, které značíme $Q_i(X_i;\, \mu_i), i \in \{1, \ldots, n\}$. Kvazi-věrohodnost náhodného vektoru $\boldsymbol{X}$ je tvaru

$$Q(\boldsymbol{X};\, \boldsymbol{\mu}) = \sum_{i=1}^{n} Q_i(X_i;\, \mu_i) = \sum_{i=1}^{n} \int_{x_i}^{\mu_i} \frac{X_i - t}{\phi V(\mu_i)}\, dt,$$

kde $\boldsymbol{X} = (X_1, \ldots, X_n)^\top \in \mathbb{R}^n$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$.

Definice 8 (kvazi-skórová funkce)**.** *Nechť* $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$ *je náhodný výběr z rozdělení patřícího do rodiny exponenciálních rozdělení. Nechť* $\mu \in \mathbb{R}$ *značí střední hodnotu tohoto rozdělení,* $\phi > 0$ *je parametr rozptýlení a* V *je příslušná rozptylová funkce, tedy* $var_\theta(X_i) = \phi V(\mu)$*, kde* $i \in \{1, \ldots, n\}$. *Pak pro konkrétní realizaci* $\boldsymbol{X} = (X_1, \ldots, X_n)^\top \in \mathbb{R}$ *a* $\boldsymbol{\mu} = (\mu, \ldots, \mu)^\top \in \mathbb{R}^n$ *funkci*

$$U(\boldsymbol{X};\, \boldsymbol{\mu}) = \sum_{i=1}^{n} \frac{X_i - \mu}{\phi V(\mu)}.$$

nazýváme kvazi-skórová funkce.

Poznámka. Analogicky jako u kvazi-věrohodnosti definujme kvazi-skórovou funkci náhodného vektoru $\boldsymbol{X}$, jehož složky mají rozdělení s odlišnými středními hodnotami, jako

$$U(\boldsymbol{X};\, \boldsymbol{\mu}) = \sum_{i=1}^{n} U_i(X_i;\, \mu_i) = \sum_{i=1}^{n} \frac{X_i - \mu_i}{\phi V(\mu_i)}.$$

Vztah mezi kvazi-věrohodností a kvazi-skórovou funkcí odpovídá vztahu mezi logaritmickou věrohodností funkcí a skórovou funkcí. Položíme-li kvazi-skórovou funkci rovnu nule, získáme *kvazi-věrohodnostní rovnici*, jež je analogií k věrohodnostní rovnici. Hledaný odhad neznámého parametru $\theta \in \Theta \subseteq \mathbb{R}$, který značíme $\widehat{\theta}_{\text{KL}}$, nalezneme jako její řešení, přičemž využijeme vztahy (1.2) a (1.3), které popisují vztah střední hodnoty $\mu \in \mathbb{R}$ (případně $\mu_i \in \mathbb{R}$) a neznámého parametru $\theta \in \Theta \in \mathbb{R}$.

Ilustrace využití

Nechť $X_1, \ldots, X_n$ je náhodný výběr z exponenciálního rozdělení. Značíme $X_i \sim Exp(\lambda), \forall\, i \in \{1, \ldots, n\}$, kde $\lambda \in \mathbb{R}$ je neznámý parametr. Víme, že platí $\mathsf{E}\, X_1 = \frac{1}{\lambda}$, $\operatorname{var} X_1 = \frac{1}{\lambda^2}$. Kvazi-skórovou funkci

$$U\left(\boldsymbol{X};\, \frac{1}{\lambda}\right) = \sum_{i=1}^{n} \frac{X_i - \frac{1}{\lambda}}{\frac{1}{\lambda^2}}$$

položíme rovnu nule a dostaneme kvazi-věrohodnostní rovnici, která je tvaru

$$\sum_{i=1}^{n} \frac{X_i - \frac{1}{\lambda}}{\frac{1}{\lambda^2}} = 0$$

a ze které vyjádříme odhad neznámého parametru $\lambda \in \mathbb{R}$ založený na kvazi-věrohodnosti tvaru $\widehat{\lambda}_{\text{KL}} = \frac{n}{\sum_{i=1}^{n} X_i} = \frac{1}{\overline{X}_n}$.

1.3 Pseudo-věrohodnost

Při hledání odhadu neznámého parametru nastávají situace, kdy je určení sdružené hustoty rozdělení pozorovaného náhodného vektoru příliš náročné nebo její tvar výrazně zkomplikuje následný výpočet. Často je tomu tak v případech, kdy složky vektoru nejsou nezávislé a nelze tak sdruženou hustotu určit jako součin hustot marginálních. Modifikací maximální věrohodnosti, která nám v takové situaci pomůže, je *pseudo-věrohodnost*, které věnujeme tuto kapitolu. Na rozdíl od maximální věrohodnosti nevychází ze sdružené hustoty celého pozorovaného náhodného vektoru, nýbrž z marginálních hustot jeho jednotlivých složek a sdružených hustot vektorů tvořených dvojicemi těchto složek. Při tvorbě této kapitoly jsme vycházeli ze článku (Cox a Reid, 2004).

Odhadování parametru

Nechť $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$, $n \in \mathbb{N}$, je náhodný vektor s hustotou $f_{\boldsymbol{X}}(\boldsymbol{x}; \theta)$, kde $\boldsymbol{x} \in \mathbb{R}^n$, $\theta \in \Theta \subset \mathbb{R}$ je neznámý parametr a Θ značí parametrický prostor. Pro jednoduchost uvažujme, že je jednorozměrný. Označme marginální hustotu náhodné veličiny X_i jako

$$f_i(x_i; \theta), \quad \text{kde}\, i \in \{1, \ldots, n\}$$

a 2-rozměrnou sdruženou hustotu náhodného vektoru $(X_i, X_j)^\top$ jako

$$f_{ij}(x_i, x_j; \theta), \quad \text{kde}\, i, j \in \{1, \ldots, n\}, i \neq j.$$

Definice 9 (logaritmická pseudo-věrohodnostní funkce)**.** *Mějme náhodný vektor $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$, $n \in \mathbb{N}$, kde X_i je náhodná veličina s husotou $f_i(x_i; \theta)$, $i \in \{1, \ldots, n\}$ a $\theta \in \Theta \subseteq \mathbb{R}$ je neznámý parametr. Dále nechť $a \in \mathbb{R}$ je vhodně volená konstanta. Pak pro konkrétní realizaci $\boldsymbol{X} = (X_1, \ldots, X_n)^\top \in \mathbb{R}^n$ nazveme funkce*

$$\begin{aligned} \ell_1(\boldsymbol{X}; \theta) &= \sum_{i=1}^{n} \log f_i(X_i; \theta), \qquad (1.4) \\ \ell_2(\boldsymbol{X}; \theta) &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \log f_{ij}(X_i, X_j; \theta) - an\ell_1(\boldsymbol{X}; \theta), \end{aligned}$$

popořadě jednorozměrná logaritmická pseudo-věrohodnostní funkce parametru $\theta \in \Theta \subseteq \mathbb{R}$ v bodě $\boldsymbol{X} \in \mathbb{R}^n$ a dvourozměrná logaritmická pseudo-věrohodnostní funkce parametru $\theta \in \Theta \subseteq \mathbb{R}$ v bodě $\boldsymbol{X} \in \mathbb{R}^n$.

Poznámka. Ve většině případů volíme $a > 0$. V případech, kdy $\ell_1(\boldsymbol{X}; \theta)$ nezávisí na $\theta \in \Theta \subseteq \mathbb{R}$, volíme $a = 0$. K volbě $a < 0$ dochází zřídka. Jde o situace, kdy je většina informace o $\theta \in \Theta \subseteq \mathbb{R}$ obsažena v $\ell_1(\boldsymbol{X}; \theta)$ a téměř žádná v $\ell_2(\boldsymbol{X}; \theta)$.

Poznámka. V případě, kdy jsou $X_1, \ldots, X_n$ vzájemně nezávislé, je výraz (1.4) roven logaritmické věrohodnostní funkci $\ell(\boldsymbol{X}; \theta)$ podle definice 3. Je tomu tak, neboť

$$\ell_1(\boldsymbol{X}; \theta) = \sum_{i=1}^{n} \log f_i(X_i; \theta) = \log \prod_{i=1}^{n} f_i(X_i; \theta) = \log \left(L(\boldsymbol{X}; \theta)\right) = \ell(\boldsymbol{X}; \theta),$$

kde nezávislost využíváme v předposlední rovnosti.

Dále definujeme pseudo-skórové funkce, pomocí kterých zapíšeme pseudo-věrohodnostní rovnice a které jsou analogií k skórové funkci.

Definice 10 (pseudo-skórová funkce)**.** *Buď $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$ náhodný vektor a $\theta \in \Theta \subset \mathbb{R}$ neznámý parametr. Jsou-li funkce $\ell_1(\boldsymbol{X};\, \theta)$ a $\ell_2(\boldsymbol{X};\, \theta)$ diferencovatelné jako funkce proměnné θ, pak pro konkrétní realizaci $\boldsymbol{X} = (X_1, \ldots, X_n)^\top \in \mathbb{R}^n$ nazveme funkce*

$$U_1(\boldsymbol{X};\, \theta) = \frac{\partial \ell_1(\boldsymbol{X};\, \theta)}{\partial \theta},$$
$$U_2(\boldsymbol{X};\, \theta) = \frac{\partial \ell_2(\boldsymbol{X};\, \theta)}{\partial \theta}.$$

popořadě jednorozměrná pseudo-skórová funkce a dvourozměrná pseudo-skórová funkce.

Pseudo-věrohodnostní rovnice získáme tak, že pseudo-skórové funkce položíme rovny nule. Hledaný odhad parametru $\theta \in \Theta \subseteq \mathbb{R}$ založený na pseudo-věrohodnosti, který značíme $\widehat{\theta}_{\mathrm{PL}}$, je jejich řešením.

Poznámka. Aby mělo smysl pseudo-věrohodnostní rovnice zavádět, předpokládejme, že logaritmické pseudo-věrohodností funkce závisí na parametru θ. V opačném případě by nám vyšly $U_1(\boldsymbol{X};\, \theta)$ a $U_2(\boldsymbol{X};\, \theta)$ identicky nulové.

Ilustrace využití

Nechť $\boldsymbol{X} = (X_1, X_2, X_3)^\top$ je náhodný vektor, který má 3-rozměrné normální rozdělení s vektorem středních hodnot $\boldsymbol{\mu} = (\mu, \mu, \mu)^\top \in \mathbb{R}^3$, kde $\mu \in \mathbb{R}$ je neznámý parametr, který budeme odhadovat, a s kovarianční maticí $\boldsymbol{\Sigma} \in \mathbb{R}^{3\times 3}$ tvaru

$$\boldsymbol{\Sigma} = \begin{pmatrix} 2.5 & 1.0 & -0.3 \\ 1.0 & 3.0 & 0.8 \\ -0.3 & 0.8 & 1.5 \end{pmatrix}.$$

Ze znalosti sdruženého rozdělení náhodného vektoru $\boldsymbol{X}$ dostáváme marginální rozdělení náhodných veličin X_1, X_2 a X_3. Platí $X_1 \sim N(\mu;\, 2.5)$, $X_2 \sim N(\mu;\, 3.0)$ a $X_3 \sim N(\mu;\, 1.5)$. Marginální hustoty jsou pro $i \in \{1, 2, 3\}$ tvaru

$$f_i(x;\, \mu) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma_i^2}\right\}, \quad \mu \in \mathbb{R}, x \in \mathbb{R},$$

kde $\sigma_1^2 = 2.5$, $\sigma_2^2 = 3.0$, $\sigma_1^3 = 1.5$ jsou příslušné hodnoty rozptylů náhodných veličin X_1, X_2 a X_3. Pomocí marginálních hustot určíme jednorozměrnou logaritmickou pseudo-věrohodnostní funkci, která je tvaru

$$\begin{aligned} \ell_1(\boldsymbol{X};\, \mu) &= \sum_{i=1}^{3} \log f_i(X_i;\, \mu) \\ &= -\tfrac{1}{2}\log\left(2\pi\sigma_1^2\right) - \tfrac{1}{5}(X_1-\mu)^2 - \tfrac{1}{2}\log\left(2\pi\sigma_2^2\right) - \tfrac{1}{6}(X_2-\mu)^2 \\ &\quad - \tfrac{1}{2}\log\left(2\pi\sigma_3^2\right) - \tfrac{1}{3}(X_3-\mu)^2, \end{aligned} \tag{1.5}$$

kde $\mu \in \mathbb{R}$, $\boldsymbol{X} = (X_1, X_2, X_3)^\top \in \mathbb{R}^3$.

Sdružená hustota náhodného vektoru $(X_i, X_j)^\top$, kde $i, j \in \{1, 2, 3\}$ a $i \neq j$, má tvar

$$f_{ij}(x_i, x_j;\, \mu) = \frac{1}{2\pi\sqrt{|\boldsymbol{\Sigma}_{i,j}|}} \exp\left\{-\frac{1}{2}\left[(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}_{i,j}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})^\top\right]\right\},$$

kde $\boldsymbol{\mu} = (\mu, \mu)^\top \in \mathbb{R}^2$, $\boldsymbol{x} = (x_i, x_j)^\top \in \mathbb{R}^2$. Výraz $|\boldsymbol{\Sigma}_{i,j}|$ značí determinant matice $\boldsymbol{\Sigma}_{i,j} \in \mathbb{R}^{2\times 2}$ a platí

$$|\boldsymbol{\Sigma}_{1,2}| = \begin{pmatrix} 2.5 & 1.0 \\ 1.0 & 3.0 \end{pmatrix} = 6.5; \quad |\boldsymbol{\Sigma}_{1,3}| = \begin{pmatrix} 2.5 & -0.3 \\ -0.3 & 1.5 \end{pmatrix} = 3.66;$$
$$|\boldsymbol{\Sigma}_{2,3}| = \begin{pmatrix} 3.0 & 0.8 \\ 0.8 & 1.5 \end{pmatrix} = 3.86.$$

Nechť $a = 1$, dvourozměrná logaritmická pseudo-věrohodnostní funkce má pro takovou volbu tvar

$$\begin{aligned}
\ell_2(\boldsymbol{X};\, \mu) &= \sum_{i=1}^{2}\sum_{j=i+1}^{3} \log f_{ij}(X_i, X_j;\, \mu) - 3\ell_1(\boldsymbol{X};\, \mu) \\
&= -\log\left(2\pi\sqrt{6.5}\right) - \tfrac{1}{13}\left[3(X_1-\mu)^2 - 2(X_1-\mu)(X_2-\mu) + 2.5(X_2-\mu)^2\right] \\
&\quad - \log\left(2\pi\sqrt{3.66}\right) - \tfrac{25}{183}\left[1.5(X_1-\mu)^2 + 0.6(X_1-\mu)(X_3-\mu) + 2.5(X_3-\mu)^2\right] \\
&\quad - \log\left(2\pi\sqrt{3.86}\right) - \tfrac{25}{193}\left[1.5(X_2-\mu)^2 - 1.6(X_2-\mu)(X_3-\mu) + 3(X_3-\mu)^2\right] \\
&\quad - \tfrac{1}{2}\log\left(2\pi\sigma_1^2\right) - \tfrac{1}{5}(X_1-\mu)^2 - \tfrac{1}{2}\log\left(2\pi\sigma_2^2\right) - \tfrac{1}{6}(X_2-\mu)^2 \\
&\quad - \tfrac{1}{2}\log\left(2\pi\sigma_3^2\right) - \tfrac{1}{3}(X_3-\mu)^2,
\end{aligned}$$

kde $\mu \in \mathbb{R}$, $\boldsymbol{X} = (X_1, X_2, X_3)^\top \in \mathbb{R}^3$ a tvar $\ell_1(\boldsymbol{X};\, \mu)$ je znám z rovnosti (1.5). Zderivováním podle proměnné μ vznikne dvourozměrná pseudo-skórová funkce. Položíme-li ji rovnou nule, vznikne pseudo-věrohodnostní rovnici, ze které vyjádříme proměnnou μ. Výsledkem je odhad neznámého parametru $\mu \in \mathbb{R}$ založený na pseudo-věrohodnosti tvaru

$$\widehat{\mu}_{\text{PL}} = \frac{\left(\frac{4}{13} + \frac{4}{13} + \frac{30}{61}\right) X_1 + \left(1 + \frac{3}{13} + \frac{35}{193}\right) X_2 + \left(2 + \frac{140}{183} + \frac{110}{193}\right) X_3}{\frac{21}{5} + \frac{7}{13} + \frac{230}{183} + \frac{145}{193}}.$$

1.4 Další metody

V této kapitole si stručně představíme metody, které jsou přímo odvozené z metody maximální věrohodnosti, jsou na ní založené nebo se metodou maximální věrohodnosti inspirují.

Profilová věrohodnost

Profilová věrohodnost nachází využití v situacích, kdy je neznámý parametr p-rozměrný tvaru $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^\top \in \Theta \subseteq \mathbb{R}^p$, $p \in \mathbb{N}$, ale odhadnout chceme např. pouze jeho část. Podrobně je představena v článku Murphy a Van Der Vaart (2000).

Metodolicky je založena na rozdělení vektoru $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$ na dva podvektory $\boldsymbol{\tau} = (\theta_1, \ldots, \theta_q)^\top$ a $\boldsymbol{\psi} = (\theta_{q+1}, \ldots, \theta_p)^\top$, kde $1 \leq q < p$, pro které platí $\boldsymbol{\theta}^\top = (\boldsymbol{\tau}^\top, \boldsymbol{\psi}^\top)$. Zájem je cílen na odhad vektoru $\boldsymbol{\tau} \in \mathbb{R}^p$. Vektor $\boldsymbol{\psi} \in \mathbb{R}^{p-q}$ nazveme *vektor rušivých parametrů*. Pro pevnou hodnotu $\boldsymbol{X} \in \mathbb{R}^n$ náhodného vektoru $\boldsymbol{X}$ definujeme *profilovou věrohodnostní funkci* parametru $\boldsymbol{\tau}$ předpisem

$$\ell^*(\boldsymbol{X};\, \boldsymbol{\tau}) = \sup_{\boldsymbol{\psi} \in \mathbb{R}^{p-q}} \ell(\boldsymbol{X};\, (\boldsymbol{\tau}^\top, \boldsymbol{\psi}^\top)^\top),$$

kde $\ell(\boldsymbol{X};\, (\boldsymbol{\tau}^\top, \boldsymbol{\psi}^\top)^\top) = \ell(\boldsymbol{X};\, \boldsymbol{\theta})$ je logaritmická věrohodnostní funkce parametru $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$ v bodě $\boldsymbol{X} \in \mathbb{R}^n$.

Poznámka. Na rozdíl od definice 3 zde pracujeme s vícerozměrným, konkrétně p-rozměrným parametrem $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p$. Tvar logaritmické věrohodostní funkce to neovlivní.

Platí, že maximalizace profilové věrohodnostní funkce v proměnné $\boldsymbol{\tau} \in \mathbb{R}^p$ je ekvivalentní maximalizaci logaritmické věrohodnostní funkce v obou proměnných $\boldsymbol{\tau} \in \mathbb{R}^p$ a $\boldsymbol{\psi} \in R^{p-q}$. Uplatnění profilová věrohodnost nachází také v Box-Coxové transformaci, která je popsána například v článku od Sakia (1992).

Empirická věrohodnost

Empirická věrohodnost byla poprvé představena Owenem ve článcích Owen (1988) a Owen (1990). Jde o neparametrickou statistickou metodu, která se využívá např. ke konstrukci intervalů spolehlivosti a k testování hypotéz. Lze ji také využít v regresních modelech, tomuto rozšíření se ve svém článku věnuje Owen (1991).

Mějme náhodný výběr $X_1, \ldots, X_n$ z rozdělení s distribuční funkcí $F_{\boldsymbol{X}}$. Empirická distribuční funkce $F_n(t) = \prod_{i=1}^n \mathbf{I}_{[X_i \leq t]}$ je neparametrickým maximálně věrohodným odhadem distribuční funkce $F_{\boldsymbol{X}}$. Je tomu tak díky tomu, že maximalizuje funkci

$$L(F) = \prod_{i=1}^{n} [F(X_i) - F(X_i-)]$$

přes všechny distribuční funkce, kde $F(X_i-) = \lim_{X \to X_i-} F(X)$.

S využitím této skutečnosti zavádíme *poměr empirické věrohodnosti*. Definujeme ho předpisem

$$R(F) = \frac{L(F)}{L(F_n)}.$$

Podmíněná věrohodnost

Podmíněná věrohodnost umožňuje modelovat pravděpodobnosti zachycení jedinců u daných událostí pomocí pozorovatelných charakteristik zachycených jedinců jako věk, pohlaví, váha atd. a událostí, při kterých můžeme jedince zachytit. Výsledné modely se využívají k odhadování velikosti uzavřené populace. Detailněji ji ve svém článku popisuje Huggins (1991).

2. Empirická část

Za pomoci simulační studie porovnáme odhady založené na maximální věrohodnosti s odhady založenými na kvazi-věrohodnosti (v případě znalosti prvních dvou momentů daného rozdělení) a s odhady založenými na pseudo-věrohodnosti (v případě závislosti složek náhodného vektoru). Připomeňme, že maximálně věrohodný odhad parametru $\theta \in \Theta \subseteq \mathbb{R}$ značíme jako $\widehat{\theta}_{\mathrm{ML}}$, odhady založené na kvazi-věrohodnosti jako $\widehat{\theta}_{\mathrm{KL}}$ a odhady založené na pseudo-věrohodnosti jako $\widehat{\theta}_{\mathrm{PL}}$.

2.1 Design simulace

. Pro každý rozsah náhodného výběru $n \in \{10, 20, 50, 100\}$ bylo vygenerovaných 1000 nezávislých náhodných výběrů. Pro každý z nich byl odhadnutý neznámý parametr pomocí daných metod založených na věrohodnosti. Odhady jsou porovnány pomocí střední čtvercové chyby (MSE), která je dána vzorcem $MSE_n(\widehat{\theta}) = (\theta_0 - \widehat{\theta})^2$, kde $\theta_0 \in \Theta \subseteq \mathbb{R}$ je skutečná hodnota neznámého parametru θ a $\widehat{\theta}$ je odhad neznámého parametru θ získaný příslušnou odhadovou metodou. Výsledky simulací jsou reportované v tabulkách 2.1 a 2.2.

2.2 Maximální věrohodnost vs. kvazi-věrohodnost

V prvním příkladě porovnáme odhady neznámého parametru založené na maximální věrohodnosti a kvazi-věrohodnosti pomocí simulační studie. Mějme náhodný výběr $(Y_1, X_1)^\top, \dots, (Y_n, X_n)^\top$. Předpokládejme, že pro $i \in \{1, \dots, n\}$ má Y_i podmíněno X_i Poissonovo rozdělení se střední hodnotou a rozptylem rovným λX_i, kde $\lambda > 0$ je neznámý parametr a X_i má pro $i \in \{1, \dots, n\}$ rovnoměrné rozdělení na intervalu $[0, 1]$. V následujících odstavcích sestrojíme odhad neznámého parametru $\lambda > 0$ nejprve třemi způsoby pomocí metody maximální věrohodnosti a poté pomocí kvazi-věrohodnosti.

Začneme výpočtem využívajícím maximální věrohodnost založenou na sdruženém rozdělení náhodných vektorů $(Y_i, X_i)^\top$, kde $i \in \{1, \dots, n\}$. Tento model značme **model 1.1** a získaný MLE značme $\widehat{\lambda}_{\mathrm{ML(1)}}$. Nechť má pro všechna $i \in \{1, \dots, n\}$ náhodná veličina Y_i podmíněná náhodnou veličinou X_i Poissonovo rozdělení s parametrem λX_i, tedy $Poiss(\lambda X_i)$. Zároveň nechť pro náhodnou veličinu X_i platí $X_i \sim R[0, 1]$, kde $i \in \{1, \dots, n\}$. Pomocí vztahů

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

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

$$f_{Y,\,X}(y, x;\, \lambda) = \frac{(\lambda x)^y}{y!} \exp\left(-\lambda x\right) \mathbf{I}_{[0,\,1]}(x), \quad \text{kde } \lambda > 0, \quad y \in \{0, 1, 2, \dots\}, x \in \mathbb{R}.$$

Věrohodnostní funkce je díky nezávislosti náhodného výběru součinem sdružených hustot náhodných vektorů $(Y_i, X_i)^\top$, kde $i \in \{1, \dots, n\}$. Je to funkce tvaru

$$L((\boldsymbol{Y}, \boldsymbol{X})^\top;\, \lambda) = \prod_{i=1}^{n} f_{Y,X}(Y_i, X_i;\, \lambda) = \prod_{i=1}^{n} \frac{(\lambda X_i)^{Y_i}}{Y_i!} \exp(-\lambda X_i) \mathbf{I}_{[0,1]}(X_i),$$

kde $\lambda > 0$, $\;(\boldsymbol{Y}, \boldsymbol{X})^\top = ((Y_1, X_1)^\top \dots, (Y_n, X_n)^\top)^\top$. Zlogaritmováním věrohodnostní funkce $L((\boldsymbol{Y}, \boldsymbol{X})^\top;\, \lambda)$ obdržíme logaritmickou věrohodnostní funkci

$$\ell((\boldsymbol{Y}, \boldsymbol{X})^\top;\, \lambda) = \sum_{i=1}^{n} \left[Y_i \log(\lambda X_i) - \lambda X_i + \log\left(\prod_{i=1}^{n} \frac{1}{Y_i!} \right) \right],$$

kde $X_i \in [0, 1]$ pro všechna $i \in \{1, \dots, n\}$. Její derivací podle proměnné λ je skórová funkce, kterou položíme rovnu nule a získáme věrohodnostní rovnici proměnné λ tvaru

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

Řešením je $\lambda = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} X_i}$, kde pro všechna $i \in \{1, \dots, n\}$ platí $Y_i \in \{0, 1, 2, \dots\}$ a $X_i \in \mathbb{R}$ a hledaný MLE je tak tvaru

$$\widehat{\lambda}_{\text{ML}(1)} = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} X_i}.$$

Druhý maximálně věrohodný odhad vychází ze znalosti marginálního rozdělení náhodných veličin Y_i, kde $i \in \{1, \dots, n\}$, tento model značme **model 1.2**. Marginální hustoty náhodných veličin Y_i lze získat ze sdružené hustoty náhodného vektoru $(Y_i, X_i)^\top$, kde $i \in \{1, \dots, n\}$ pomocí vztahu

$$\begin{aligned}
f_Y(y;\, \lambda) &= \int_{\mathbb{R}} \frac{(\lambda x)^y}{y!} \exp(-\lambda x) \mathbf{I}_{[0,1]}(x)\, dx = \int_0^1 \frac{(\lambda x)^y}{y!} \exp(-\lambda x)\, dx \\
&= \frac{1}{y!} \int_0^\lambda t^y \exp(-t) \frac{1}{\lambda}\, dt = \frac{1}{\lambda y!} \left\{ [-t^y \exp(-t)]_0^\lambda + \int_0^\lambda y t^{y-1} \exp(-t)\, dt \right\} \\
&= \frac{1}{\lambda y!} \left[(-\lambda^k \exp(-\lambda)) + y \int_0^\lambda t^{y-1} \exp(-t)\, dt \right] \\
&= \frac{1}{\lambda y!} (-\lambda^y \exp(-\lambda)) + \frac{1}{\lambda (y-1)!} (-\lambda^{y-1} \exp(-\lambda)) + \dots \\
&\quad + \frac{1}{\lambda} (-\lambda \exp(-\lambda)) + \frac{1}{\lambda} (1 - \exp(-\lambda)).
\end{aligned}$$

Z výpočtu plyne, že náhodný výběr $\boldsymbol{Y} = (Y_1, \dots, Y_n)^\top$, pochází z rozdělení, jehož hustota (a tudíž marginální hustota náhodných veličin Y_i, kde $i \in \{1, \dots, n\}$) je dána předpisem

$$f_Y(y;\, \lambda) = \frac{1}{\lambda} - \sum_{j=0}^{y} \frac{1}{\lambda j!} \lambda^j \exp(-\lambda), \quad \text{kde } \lambda > 0, y \in \{0, 1, 2, \dots\}, x \in \mathbb{R}.$$

Věrohodnostní funkce, založená na náhodném výběru $\boldsymbol{Y} = (Y_1, \dots, Y_n)^\top$, je díky nezávislosti součinem marginálních hustot náhodných veličin Y_i, kde $i \in$

$\{1, \ldots, n\}$ a má tvar

$$L(\boldsymbol{Y};\, \lambda) = \prod_{i=1}^{n} f_Y(Y_i;\, \lambda) = \prod_{i=1}^{n} \left[\frac{1}{\lambda} - \sum_{j=0}^{Y_i} \frac{1}{\lambda j!} \lambda^j \exp(-\lambda) \right],$$

kde $\lambda > 0, Y_i \in \{0, 1, 2, \ldots\}, X_i \in \mathbb{R}$. Logaritmická věrohodnostní funkce $\ell(\boldsymbol{Y};\, \lambda)$ založená na náhodném výběru $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$ je logaritmem věrohodnostní funkce $L(\boldsymbol{Y};\, \lambda)$. Její derivace podle proměnné λ dává skórovou funkci. Věrohodnostní rovnice pak vzniká položením skórové funkce rovné nule. V tomto případě má tvar

$$\sum_{i=1}^{n} \frac{-\frac{1}{\lambda^2} - \sum_{j=0}^{Y_i} \left[-\frac{\lambda^j \exp(-\lambda)}{\lambda^2 j!} + \frac{j \lambda^{j-1} \exp(-\lambda)}{\lambda j!} - \frac{\lambda^j \exp(-\lambda)}{\lambda j!} \right]}{\frac{1}{\lambda} - \sum_{j=0}^{Y_i} \frac{1}{\lambda j!} \lambda^j \exp(-\lambda)} = 0$$

a hledaný odhad $\widehat{\lambda}_{\text{ML}(2)}$ neznámého parametru λ bude jejím řešením. Explicitní vyjádření odhadu $\widehat{\lambda}_{\text{ML}(2)}$ zde není uvedeno, ale bude porovnám s ostatními odhady pomocí simulační studie.

U určení třetího maximálně věrohodného odhadu, značme $\widehat{\lambda}_{\text{ML}(3)}$, ignorujme skutečnost, že pravděpodobnostní rozdělení náhodných veličin Y_i závisí na náhodných veličinách X_i, kde $i \in \{1, \ldots, n\}$. Model značíme **model 1.3**. Nechť $Y_1, \ldots, Y_n$ tvoří náhodný výběr z Poissonova rozdělení, pro které platí $\mathsf{E}\, Y_1 = \operatorname{var} Y_1 = \lambda$, tedy $Y_1 \sim \mathit{Poiss}(\lambda)$, kde $\lambda > 0$ je neznámý parametr. Marginální hustota náhodné veličiny Y_i je pro všechna $i \in \{1, \ldots, n\}$ tvaru

$$f(y;\, \lambda) = \exp(-\lambda) \frac{\lambda^y}{y!}, \quad \text{kde } \lambda > 0, y \in \{0, 1, 2, \ldots\}.$$

Věrohodnostní funkce je sdruženou hustotou náhodného vektoru $\boldsymbol{Y}$. Díky nezávislosti složek jde o součin marginálních hustot náhodných veličin Y_i, kde $i \in \{1, \ldots, n\}$. Vzniklá funkce má tvar

$$L(\boldsymbol{Y};\, \lambda) = \exp(-n\lambda) \frac{\lambda^{\sum_{i=1}^n Y_i}}{\prod_{i=1}^n Y_i!}, \quad \text{kde } \lambda > 0, \boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top \in \{0, 1, 2, \ldots\}^n.$$

Logaritmická věrohodnostní funkce náhodného výběru $\boldsymbol{Y}$ je logaritmem věrohodnostní funkce $L(\boldsymbol{Y}, \lambda)$

$$\ell(\boldsymbol{Y};\, \lambda) = -n\lambda + \sum_{i=1}^{n} Y_i \log \lambda - \log \prod_{i=1}^{n} Y_i!,$$

kde $\lambda > 0, \boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top \in \{0, 1, 2, \ldots\}^n$. Derivací $\ell(\boldsymbol{Y};\, \lambda)$ podle proměnné λ je skórová funkce. Tu položíme rovnu nule a získáme věrohodnostní rovnici proměnné λ tvaru

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

Jejím řešením je $\lambda = \frac{1}{n} \sum_{i=1}^n Y_i$ a hledaný MLE má tak tvar

$$\widehat{\lambda}_{\text{ML}(3)} = \overline{Y_n}.$$

Tabulka 2.1: Porovnání kvality odhadů neznámého parametru $\lambda > 0$ pomocí maximální věrohodnosti v **modelu 1.1** ($\widehat{\lambda}_{\mathrm{ML}(1)}$), **modelu 1.2** ($\widehat{\lambda}_{\mathrm{ML}(2)}$) a **modelu 1.3** ($\widehat{\lambda}_{\mathrm{ML}(3)}$) a pomocí kvazi-věrohodnosti ($\widehat{\mu}_{\mathrm{KL}}$). 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	
$\widehat{\lambda}_{\mathrm{ML}(1)}$	0.828	*(1.146)*	0.433	*(0.671)*	0.165	*(0.234)*	0.084	*(0.121)*
$\widehat{\lambda}_{\mathrm{ML}(2)}$	1.299	*(1.825)*	0.661	*(1.024)*	0.254	*(0.352)*	0.132	*(0.192)*
$\widehat{\lambda}_{\mathrm{ML}(3)}$	4.288	*(2.300)*	4.106	*(1.644)*	4.100	*(1.054)*	4.041	*(0.739)*
$\widehat{\lambda}_{\mathrm{KL}}$	0.828	*(1.146)*	0.433	*(0.671)*	0.165	*(0.234)*	0.084	*(0.121)*

Nyní se dostáváme k určení odhadu pomocí kvazi-věrohodnosti, která je založena pouze na znalosti střední hodnoty a rozptylu Poissonova rozdělení náhodné veličiny Y_i podmíněné náhodnou veličinou X_i. Kvazi-skórovou funkci $U(\boldsymbol{Y}; \lambda \boldsymbol{X})$ náhodného vektoru $\boldsymbol{Y}$, kde $Y_i | X_i \sim Poiss(\lambda X_i)$ položíme rovnu nule a získáme kvazi-věrohodnostní rovnici

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

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

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

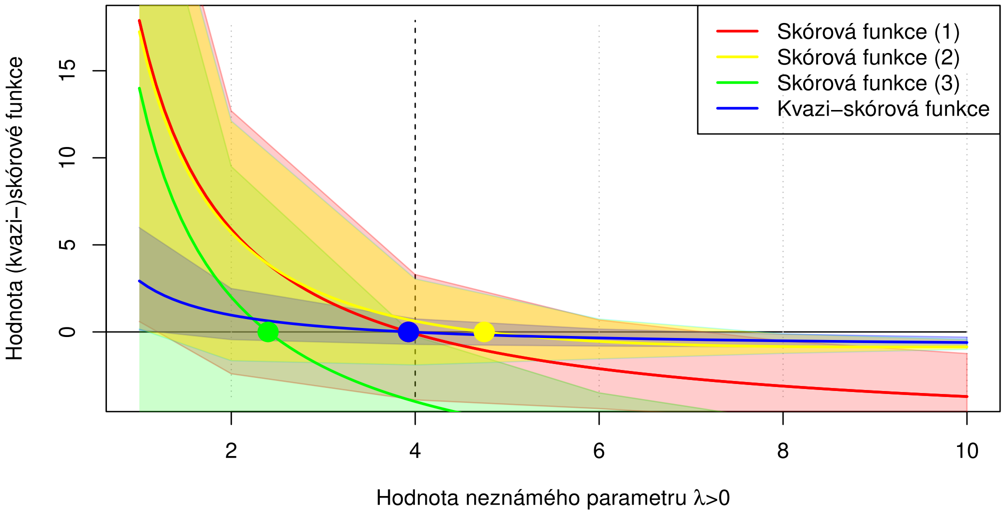
Simulace je provedena na náhodném výběru $(Y_1, X_1)^\top, \ldots, (Y_n, X_n)^\top$ rozsahu n, kde pro $i \in \{1, \ldots, n\}$ má náhodná veličina Y_i podmíněná náhodnou veličinou X_i rozdělením $Poiss(\lambda_0 X_i)$, kde $\lambda_0 = 4$ je skutečná hodnota parametru λ. Porovnání kvality odhadů $\widehat{\lambda}_{\mathrm{ML}(1)}$, $\widehat{\lambda}_{\mathrm{ML}(2)}$, $\widehat{\lambda}_{\mathrm{ML}(3)}$a $\widehat{\lambda}_{\mathrm{KL}}$ nabízí tabulka 2.1

Již ze vzorců odvozených výše je zjevné, že jsou získané odhady $\widehat{\lambda}_{\mathrm{ML}(1)}$ a $\widehat{\lambda}_{\mathrm{KL}}$ stejné a tudíž se neliší ani jejich střední čtvercové chyby. Výrazně nejhorší je odhad $\widehat{\lambda}_{\mathrm{ML}(3)}$, který vychází z **modelu 1.3**. Můžeme si všimnout, že u všech odhadů se s rostoucím rozsahem náhodného výběru n velikost střední čtvercové chyby zmenšuje a tedy přesnost odhadů zlepšuje.

Pro rozsah náhodného výběru $n = 10$ situaci ilustrujeme také na obrázku 2.1. I z něj je ihned patrná vysoká nepřesnost odhadu $\widehat{\lambda}_{\mathrm{ML}(3)}$. Lze si také všimnout, že spolehlivostní pás kvazi-skórové funkce je oproti těm, které přísluší skórovým funkcím, plošší.

2.3 Maximální věrohodnost vs. pseudo-věrohodnost

V následujícím příkladu porovnáváme odhady neznámého parametru založené na maximální věrohodnosti a pseudo-věrohodnosti pomocí simulační studie. Nechť náhodný vektor $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$ má n-rozměrné normální rozdělení $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, kde $\boldsymbol{\mu} = (\mu, \ldots, \mu)^\top \in R^n$, kde $\mu \in \mathbb{R}$ je neznámý parametr a $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ je



Obrázek 2.1: Porovnání skórové funkce založené na sdruženém rozdělení náhodných vektorů $(Y_i, X_i)^\top$, pro $i \in \{1, \dots, 10\}$ (červená), skórové funkce založené na marginálním rozdělení náhodných veličin Y_i pro $i \in \{1, \dots, 10\}$ (žlutá), skórové funkce náhodného výběru z rozdělení $Poiss(\lambda)$ (zelená) a kvazi-skórové funkce (modrá) pro konkrétní náhodný výběr $(\boldsymbol{Y}, \boldsymbol{X})^\top = ((Y_1, X_1)^\top, \dots, (Y_{10}, X_{10})^\top)^\top$ s rozsahem $n = 10$. 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.

kovarianční matice, která má následující tvar:

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho_1 & \rho_2 & 0 & \dots & 0 \\ \rho_1 & 1 & \rho_1 & \rho_2 & \dots & 0 \\ \rho_2 & \rho_1 & 1 & \rho_1 & \dots & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ 0 & \dots & \rho_2 & \rho_1 & 1 & \rho_1 \\ 0 & \dots & 0 & \rho_2 & \rho_1 & 1 \end{pmatrix}, \tag{2.1}$$

kde alespoň jedna z reálných hodnot ρ_1, ρ_2 je nenulová. Z kovarianční matice můžeme určit rozptyl náhodných veličin X_i a kovarianci mezi náhodnými veličinami X_i, X_j pro $i, j \in \{1, \dots, n\}$, kde $i \neq j$. Platí

$$\operatorname{var} X_i = 1, \text{ kde } i \in \{1, \dots, n\},$$

$$\operatorname{cov}(X_i, X_j) = \begin{cases} \rho_1, & \text{pro } i, j \in \{1, \dots, n\}, \text{je-li } j = i \pm 1, \\ \rho_2, & \text{pro } i, j \in \{1, \dots, n\}, \text{ je-li } j = i \pm 2, \\ 0, & \text{jinak.} \end{cases}$$

Ze znalosti sdruženého rozdělení náhodného vektoru $\boldsymbol{X}$ lze určit marginální rozdělení náhodných veličin X_i, kde $i \in \{1, \dots, n\}$. Platí $X_i \sim N(\mu, 1)$.

Začneme konstrukcí odhadu neznámého parametru $\mu \in \mathbb{R}$ pomocí metody maximální věrohodnosti, značíme $\widehat{\mu}_{\text{ML}}$. Sdružená hustota náhodného vektoru $\boldsymbol{X}$ je tvaru

$$f_{\boldsymbol{X}}(\boldsymbol{x}; \boldsymbol{\mu}) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right\}, \tag{2.2}$$

kde $|\boldsymbol{\Sigma}|$ je determinant matice $\boldsymbol{\Sigma}$ a $\boldsymbol{x} = (x_1, \dots, x_n)^\top \in \mathbb{R}^n$. Všimněme si, že abychom mohli výpočet provést, museli bychom spočíst inverzní matici řádu n a spočítat její determinant. Dalším komplikovaným krokem by bylo roznásobení výrazu uvnitř exponenciály. Výpočet si usnadníme zanedbáním vzájemné závislosti složek náhodného vektoru $\boldsymbol{X}$.

Poznámka. Závislost je patrná z nenulových kovariančních koeficientů.

V případě prvního odhadu budeme ignorovat závislost a budeme předpokládat,že prvky náhodného vektoru $\boldsymbol{X} = (X_1, \dots, X_n)^\top$ tvoří náhodný výběr z normálního rozdělení $N(\mu, 1)$. Marginální hustota náhodné veličiny X_i, kde $i \in \{1, \dots, n\}$, má tvar

$$f(x;\, \mu) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x-\mu)^2\right\}, \quad \text{kde } \mu \in \mathbb{R}, x \in \mathbb{R}.$$

Z předpokladu nezávislosti složek náhodného vektoru $\boldsymbol{X}$ je věrohodnostní funkce tvaru

$$L(\boldsymbol{X};\, \mu) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2}\sum_{i=1}^n (X_i-\mu)^2\right\}, \quad \mu \in \mathbb{R}, \boldsymbol{X} \in \mathbb{R}^n.$$

Logaritmická věrohodnostní funkce vznikne jejím zlogaritmováním a má tvar

$$\ell(\boldsymbol{X};\, \mu) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\sum_{i=1}^n (X_i-\mu)^2, \quad \mu \in \mathbb{R}, \boldsymbol{X} \in \mathbb{R}^n,$$

Derivací podle proměnné μ a položením rovné nule vznikne věrohodnostní rovnice

$$\sum_{i=1}^n (X_i - \mu) = 0,$$

ze které lze snadno vyjádřit $\mu = \frac{1}{n}\sum_{i=1}^n X_i$. Maximálně věrohodný odhad má tak tvar

$$\widehat{\mu}_{\text{ML}} = \overline{X_n}.$$

Přesuňme se k určení odhadu neznámého parametru $\mu \in \mathbb{R}$ pomocí pseudověrohodnosti, který využije i informaci o závislosti složek náhodného vektoru $\boldsymbol{X}$, získaný odhad značíme $\widehat{\mu}_{\text{PL}}$. V tomto případě není nutné zanedbávat závislost mezi jednotlivými složkami. Marginální hustota náhodné veličiny $X_i \sim N(\mu, 1)$ má tvar

$$f(x;\, \mu) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x-\mu)^2\right\}, \quad \mu \in \mathbb{R}, x \in \mathbb{R}.$$

Dále platí, že náhodný vektor $(X_i, X_j)^\top$ má dvourozměrné normální rozdělení $N_2((\mu, \mu)^\top, \boldsymbol{\Sigma}_{i,j})$, kde matici $\boldsymbol{\Sigma}_{i,j} \in \mathbb{R}^{2\times 2}$ lze získat z matice $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$ omezením se na řádky a sloupce s indexy i, j.

V případech, kdy $j = i \pm k$, kde $k \in \{1, 2\}$, získáme matici

$$\boldsymbol{\Sigma}_{i,j} = \begin{pmatrix} 1 & \rho_k \\ \rho_k & 1 \end{pmatrix}.$$

S využitím výrazu (2.2) pro vektor velikosti 2 místo n a záměnou matice $\mathbf{\Sigma} \in \mathbb{R}^{n\times n}$ za matici $\mathbf{\Sigma}_{i,j} \in \mathbb{R}^{2\times 2}$ lze nahlédnout, že sdružená hustota náhodného vektoru $(X_i, X_j)^\top$ tvaru

$$f_{ij}(x_i, x_j;\, \mu) =$$
$$= \frac{1}{2\pi\sqrt{|\mathbf{\Sigma}_{i,j}|}} \exp\left\{-\frac{1}{2|\mathbf{\Sigma}_{i,j}|}\left[(x_i-\mu)^2 - 2\rho_k(x_i-\mu)(x_j-\mu) + (x_j-\mu)^2\right]\right\},$$

kde $\mu \in \mathbb{R}$, $x_i, x_j \in \mathbb{R}$ a $|\mathbf{\Sigma}_{i,j}| = 1^2 - (\rho_k)^2$.

Analogicky se řeší případy, kdy $j \neq i \pm k$, kde $k \in \{1, 2\}$ a $j \neq i$. Matice $\mathbf{\Sigma}_{i,j}$ má tvar

$$\mathbf{\Sigma}_{i,j} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

a sdružená hustota náhodného vektoru $(X_i, X_j)^\top$ je v tomto případě rovna

$$f_{ij}(x_i, x_j;\, \mu) = \frac{1}{2\pi} \exp\left\{-\frac{1}{2}\left[(x_i-\mu)^2 + (x_j-\mu)^2\right]\right\},$$

kde $\mu \in \mathbb{R}$, $x_i, x_j \in \mathbb{R}$ a s využitím toho, že $|\mathbf{\Sigma}_{i,j}| = 1$. Jednorozměrná logaritmická pseudo-věrohodnostní funkce je tvaru

$$\ell_1(\boldsymbol{X};\, \mu) = \sum_{i=1}^{n} \log f(X_i;\, \mu) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\sum_{i=1}^{n}(X_i-\mu)^2,$$

kde $\mu \in \mathbb{R}$, $\boldsymbol{X} = (X_1, \dots, X_n)^\top \in \mathbb{R}^n$. Volbou $a = 1$ vznikne dvourozměrnou logaritmická pseudo-věrohodnostní funkce tvaru

$$\ell_2(\boldsymbol{X};\, \mu) = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \log f_{ij}(X_i, X_j;\, \mu) - na\ell_1(\boldsymbol{X};\, \mu).$$

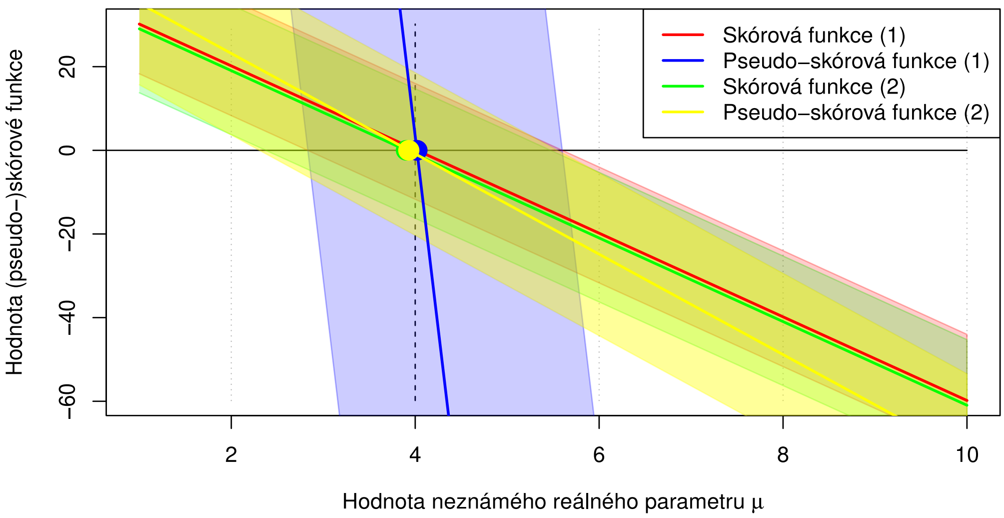
Následný výpočet je proveden analogicky jako v kapitole 1.3, kde ilustrujeme výpočet odhadu neznámého parametru založený na pseudověrohodnosti. Získaný odhad neznámého parametru μ založený na pseudo-věrohodnosti značíme $\widehat{\mu}_{\mathrm{PL}}$, nebude zde uveden jeho explicitní tvar, ale bude zahrnut v simulační studii.

Simulaci provedeme pro dva různé modely pro různé volby hodnot ρ_1 a ρ_2. V **modelu 2.1** nechť náhodný vektor $\boldsymbol{X} = (X_1, \dots, X_n)^\top$ pochází z n-rozměrného normálního rozdělení $N_n(\mu_0, \Sigma_1)$, kde $\boldsymbol{\mu}_0 = (\mu_0, \dots, \mu_0)^\top \in R^n$ a $\mu_0 = 4$. Nechť Σ_1 je matice vzniklá z matice 2.1 volbou $\rho_1 = 0.5$ a $\rho_2 = 0$. Porovnávané odhady v tomto případě značme $\widehat{\mu}_{\mathrm{ML}(1)}$ a $\widehat{\mu}_{\mathrm{PL}(1)}$. V **modelu 2.2** nechť generovaný náhodný vektor $\boldsymbol{X} = (X_1, \dots, X_n)^\top$ pochází z n-rozměrného normálního rozdělení $N_n(\mu_0, \Sigma_2)$, kde $\boldsymbol{\mu}_0 = (\mu_0, \dots, \mu_0)^\top \in R^n$ a $\mu_0 = 4$. Nechť Σ_2 je matice vzniklá z matice 2.1 volbou $\rho_1 = 0.5$ a $\rho_2 = 0.4$. Porovnávané odhady v tomto případě značme $\widehat{\mu}_{\mathrm{ML}(2)}$ a $\widehat{\mu}_{\mathrm{PL}(2)}$.

Porovnání kvality odhadů $\widehat{\mu}_{\mathrm{ML}(1)}$, $\widehat{\mu}_{\mathrm{PL}(1)}$, $\widehat{\mu}_{\mathrm{ML}(2)}$ a $\widehat{\mu}_{\mathrm{PL}(2)}$ nabízí tabulka 2.2. Nejmenší čtvercovou chybu má odhad $\widehat{\mu}_{\mathrm{PL}(1)}$. Střední čtvercové chyby jsou celkově v **modelu 2.1** nižší. Pro **model 2.2** má menší čtvercovou chybu odhad založený na maximální věrohodnosti, což je vzhledem k zanedbání závislosti překvapivé. Zároveň u všech odhadů platí, že s rostoucím $n \in \mathbb{N}$ se střední čtvercová chyba

Tabulka 2.2: Porovnání kvality odhadů neznámého parametru μ v **modelu 2.1** (pro náhodný výběr rozsahu n) založených na maximální věrohodnosti $\widehat{\mu}_{\mathrm{ML}(1)}$. a pseudo-věrohodnosti $\widehat{\mu}_{\mathrm{PL}(1)}$. a odhadů neznámého parametru μ založených na maximální věrohodnosti $\widehat{\mu}_{\mathrm{ML}(2)}$ a pseudo-věrohodnosti $\widehat{\mu}_{\mathrm{PL}(2)}$. v **modelu 2.2** (pro náhodný výběr rozsahu n) pomocí výběrového průměru (*směrodatné odchylky*) střední čtvercové chyby na základě 1000 Monte-Carlo simulací

	n = 10		n = 20		n = 50		n = 100	
$\widehat{\mu}_{\mathrm{ML}(1)}$	0.192	*(0.268)*	0.096	*(0.128)*	0.040	*(0.055)*	0.020	*(0.030)*
$\widehat{\mu}_{\mathrm{PL}(1)}$	0.191	*(0.268)*	0.095	*(0.128)*	0.040	*(0.055)*	0.020	*(0.030)*
$\widehat{\mu}_{\mathrm{ML}(2)}$	0.260	*(0.363)*	0.139	*(0.188)*	0.054	*(0.078)*	0.027	*(0.038)*
$\widehat{\mu}_{\mathrm{PL}(2)}$	0.276	*(0.384)*	0.145	*(0.196)*	0.055	*(0.078)*	0.027	*(0.039)*



Obrázek 2.2: Porovnání skórové funkce (červená) a pseudo-skórové funkce (modrá) založených na sdruženém rozdělení náhodného vektoru $\boldsymbol{X} = (X_1, \dots, X_{10})^\top$ rozsahu $n = 10$ z **modelu 2.1** pro konkrétní náhodný vektor $\boldsymbol{X} = (X_1, \dots, X_{10})^\top$ a skórové funkce (zelená) a pseudo-skórové funkce (žlutá) založených na sdruženém rozdělení náhodného vektoru $\boldsymbol{X} = (X_1, \dots, X_{10})^\top$ rozsahu $n = 10$ z **modelu 2.2**, ve kterém je na sobě závislých více složek, pro konkrétní náhodný výběr $\boldsymbol{X} = (X_1, \dots, X_{10})^\top$. Průsečíky s osou x značí příslušné odhady neznámého parametru $\mu \in \mathbb{R}$, jehož skutečná hodnota $\mu_0 = 4$ je zvýrazněna černě čárkovaně. Spolehlivostní pásy značí plochu mezi minimy a maximy skórových funkcí a pseudo-skórových funkcí v bodech $2, 4, 6, 8$ a 10 lineárně interpolovanými přes 1000 Monte-Carlo simulací.

zmenšuje a odhady jsou pro vyšší n kvalitnější. Pro rozsah náhodného výběru $n = 10$ situaci ilustrujeme také na grafu na obrázku 2.2. Lze nahlédnout, že pro konkrétní pozorované náhodné vektory se od sebe čtyři odhady neznámého parametru μ výrazně neliší. Rozdíl pozorujeme mezi skórovými a pseudo-skórovými funkcemi, kde Skórová funkce (1) klesá výrazně rychleji než ostatní funkce. Zároveň má nejširší spolehlivostní pás.

Závěr

V této práci byla popsána metoda maximální věrohodnosti, která slouží k odhadování neznámého teoretického parametru pomocí empirických dat. Její aplikace byla ilustrována na konkrétních příkladech. Zadefinována byla také kvazi-věrohodost, pomocí níž lze najít odhad neznámého parametru na základě znalosti prvních dvou momentů rozdělení empirických dat. Následovala kapitola věnující se pseudo-věrohodnosti, jež určuje odhad neznámého parametru pomocí marginálních hustot náhodných veličin a sdružených hustot dvourozměrných náhodných vektorů. Využití kvazi-věrohodnosti a pseudo-věrohodnosti bylo, stejně jako u maximální věrohodnosti, ilustrováno na konkrétních příkladech. Závěrem teoretické části byl stručný popis profilové, empirické a podmíněné věrohodnosti.

V empirické části práce byly pomocí simulačních studií porovnány odhady založené na maximální věrohodnosti a kvazi-věrohodnosti a následně na maximální věrohodnosti a pseudo-věrohodnosti. Kvalita odhadů byla porovnána pomocí střední čtvercové chyby příslušných odhadů přes 1000 Monte-Carlo simulací. Výsledky jsou zaznamenány v tabulkách 2.1 a 2.2 a ilustrovány grafy na obrázcích 2.1 a 2.2. V příkladu s Poissonovým rozdělením měl nejnižší střední čtvercovou chybu maximálně věrohodný odhad $\widehat{\lambda}_{\mathrm{ML}(1)}$ založený na sdruženém rozdělení náhodného vektoru $(Y_i, X_i)^\top$, kde $i \in \{1, \ldots, n\}$ (**model 1.1**). Stejné hodnoty střední čtvercové chyby byly obdrženy i pro odhad založený na kvazi-věrohodnosti $\widehat{\lambda}_{\mathrm{KL}}$. Zároveň s rostoucím rozsahem náhodného výběru docházelo ke zlepšování přesnosti všech odhadů. V příkladu s mnohorozměrným normálním rozdělením měl nejnižší střední čtvercovou chybu odhad $\widehat{\mu}_{\mathrm{PL}(1)}$ založený na pseudo-věrohodnosti z **modelu 2.1**. Zároveň s rostoucím rozsahem náhodného výběru docházelo ke zlepšování přesnosti všech odhadů.

Seznam použité literatury

ALDRICH, J. (1997). R. A. Fisher and the Making of Maximum Likelihood 1912 – 1922. *Statistical Science*, **12**(3), 162–176.

COX, D. R. a REID, N. (2004). A Note on Pseudolikelihood Constructed from Marginal Densities. *Oxford University Press on behalf of Biometrika Trust*, **91** (3), 729–737.

FISHER, R. A. (1997). On an Absolute Criterion for Fitting Frequency Curves. *Statistical Science*, **12**(1), 39–41.

GUO, D., WU, Y., SHITZ, S. S. a VERDÚ, S. (2011). Estimation in Gaussian Noise: Properties of the Minimum Mean-Square Error. *IEEE Transactions on Information Theory*, **57**(4), 2371–2385.

HUGGINS, R. M. (1991). Some Practical Aspects of a Conditional Likelihood Approach to Capture Experiments. *International Biometric Society*, **47**(2), 725–732.

KULICH, M. (2023). *NMST412/NMFM402 Generalized Linear Models*. URL `https://www.karlin.mff.cuni.cz/~kulich/vyuka/glm/doc/glm_notes_230216.pdf`. Course Notes,
Datum přístupu 19.7.2023.

MURPHY, S. A. a VAN DER VAART, A. W. (2000). On Profile Likelihood. *Journal of the American Statistical Association*, **95**(450), 449–465.

NAGY, S. (2023). *NMSA332: MATHEMATICAL STATISTICS II*. URL `http://www.karlin.mff.cuni.cz/~nagy/NMSA332/NMSA332.pdf`. Course Notes,
Datum přístupu 19.7.2023.

NELDER, J. A. a LEE, Y. (1992). Likelihood, Quasi-Likelihood and Pseudolikelihood: Some Comparisons. *Wiley for the Royal Statistical Society*, **54**(1), 273–284.

OWEN, A. B. (1988). Empirical Likelihood Ratio Confidence Intervals for a Single Functional. *Biometrika*, **75**(2), 237–249.

OWEN, A. B. (1990). Empirical Likelihood Ratio Confidence Regions. *Ann. Statist.*, **18**(1), 90–120.

OWEN, A. B. (1991). Empirical Likelihood for Linear Models. *Institute of Mathematical Statistics*, **19**(4), 1725–1747.

SAKIA, R. M. (1992). The Box-Cox Transformation Technique: A Review. *Journal of the Royal Statistical Society. Series D (The Statistician)*, **41**(2), 169–178.

WILCOX, R. R. (2012). *Introduction to Robust Estimation and Hypothesis Testing*. 3rd edition. Academic Press is an imprint of Elsevier, The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK. ISBN 978-0-12-386983-8.

WOOLDRIDGE, J. M. (2001). Applications of Generalized Method of Moments Estimation. *Journal of Economic Perspectives*, **15**(4), 87–100.