MATEMATICKO-FYZIKÁLNÍ FAKULTA
Univerzita Karlova

DIPLOMOVÁ PRÁCE

Eva Kielkowská

Metoda maximální věrohodnosti pro pozorování, která nejsou stejně rozdělená nebo nezávislá

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: Ing. Marek Omelka, Ph.D.
Studijní program: Matematika
Studijní obor: Pravděpodobnost, matematická statistika a ekonometrie

Praha 2017

Prohlašuji, že jsem tuto diplomovou práci vypracovala samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

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

V .... dne ....

Ráda bych poděkovala svému vedoucímu diplomové práce, dr. Omelkovi, nejen za odborné rady při zpracovávání tématu, ale také za vstřícnost a ochotu při konzultacích.

Název práce: Metoda maximální věrohodnosti pro pozorování, která nejsou stejně rozdělená nebo nezávislá

Autor: Eva Kielkowská

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: Ing. Marek Omelka, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: V práci se zabýváme metodou maximální věrohodnosti pro pozorování, která jsou nezávislá, ale nejsou stejně rozdělená. V první části jsou stanoveny podmínky pro konzistenci a asymptotickou normalitu maximálně věrohodných odhadů v tomto případě. Využívá se zde hlavně stejnoměrná integrovatelnost náhodných veličin. Ověření uvedených podmínek je ilustrováno na K-výběrovém problému. V druhé části se práce zaměřuje na situace, ve kterých odhady parametrů získáme minimalizací konvexních funkcí. Důkaz konzistence a asymptotické normality pro tyto odhady je založen na výsledcích pro konvexní náhodné funkce. Tento postup je možné použít pro metodu maximální věrohodnosti v modelech s logkonkávními hustotami. Příklad normálního lineárního modelu, logistické regrese a poissonovské regrese demonstruje použití výsledků představených v druhé části práce.

Klíčová slova: stejnoměrná integrovatelnost, konvexita, regresní modely

Title: Maximum likelihood theory for not i.i.d. observations

Author: Eva Kielkowská

Department: Department of Probability and Mathematical Statistics

Supervisor: Ing. Marek Omelka, Ph.D., Department of Probability and Mathematical Statistics

Abstract: Maximum likelihood approach for independent but not identically distributed observations is studied. In the first part of the thesis, conditions for consistency and asymptotic normality of the maximum likelihood estimates for this case are stated. Uniform integrability has a major role in proving the desired properties. K-sample problem serves as an example for using the described method. The second part is focused on estimates obtained by minimizing convex functions. Convexity is a key for showing the consistency and asymptotic normality of the estimates in this case. The results can be used for maximum likelihood when observations with logconcave densities are involved. Finally, normal linear model, logistic regression and Poisson regression examples are provided to present the application of the method.

Keywords: uniform integrability, convexity, regression models

Obsah

Úvod

Metoda maximální věrohodnosti, kterou se v této práci zabýváme, má široké uplatnění. Používá se v mnoha aplikacích pro odhadovaní parametrů a k následné inferenci. Typicky ji uvažujeme pro případ, kdy máme nezávislá a stejně rozdělená pozorování. V praxi se však můžeme setkat také s případy, kdy je některý z požadavků (stejná rozdělenost či nezávislost) porušen a tam už si s podmínkami, které potřebujeme pro konzistenci a asymptotickou normalitu v případě nezávislých a stejně rozdělených pozorování, nevystačíme.

Práce se zaměřuje na situace, kdy je porušen předpoklad stejného rozdělení pro všechna pozorování, avšak zůstává zachována jejich nezávislost. Hlavní motivací je využití metody maximální věrohodnosti pro regresní modely, ve kterých uvažujeme pevné regresory.

V první části jsou vysvětleny podmínky pro konzistenci a asymptotickou normalitu pro obecný případ s nestejně rozdělenými pozorováními. Jedná se o souhrn kritérií, která musí uvažovaný model splňovat, abychom měli zaručeny požadované výsledky. Obecný postup je následně ilustrován na K-výběrovém problému.

V druhé části se pak věnujeme speciálnímu případu, kdy funkce, na jejichž základě odhad získáváme, jsou v odhadovaném parametru konvexní. Zde využíváme fakt, že odhad metodou maximální věrohodnosti je vlastně speciálním případem M-odhadu. Tento přístup je následně ilustrován na lineárním regresním modelu, logistické a nakonec poissonovské regresi, ve kterých uvažujeme pevné regresory.

1. Metoda maximální věrohodnosti pro nezávislá, ale ne stejně rozdělená pozorování

Metodu maximální věrohodnosti běžně formulujeme pro nezávislá, stejně rozdělená pozorování. Díky tomuto a několika dalším předpokladům dostáváme konzistentní odhady parametrů a jejich asymptotickou normalitu. Nyní se budeme zabývat případem, kdy máme k dispozici nezávislá, ale ne stejně rozdělená pozorování. Podíváme se, jaké podmínky musí být splněny pro dosažení konzistentních a asymptoticky normálních odhadů v tomto případě. Kapitola vychází z článku Hoadley (1971), který obsahuje zde uvedené věty. Důkazy, které zde uvádíme, se ve stručnější podobě také nacházejí v tomto článku.

1.1 Zavedení pojmů a značení

Nejprve si zavedeme základními pojmy a značení, které budeme používat. Uvažujme posloupnost nezávislých náhodných veličin $Y_1, \dots, Y_n$, definovaných na pravděpodobnostním prostoru $(\Omega, \mathcal{F}, P_\theta)$, které nabývají hodnot na měřitelném prostoru $(\mathcal{Y}, \mathcal{A}, \mu)$ a nechť $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$. Předpokládejme dále, že Y_i má hustotu $f_i(y|\boldsymbol{\theta})$ vzhledem k μ, σ-konečné míře na $(\mathcal{Y}, \mathcal{A})$. Potom odpovídající věrohodnostní funkci definujeme jako $L_n(\boldsymbol{\theta}) = \prod_{i=1}^{n} f_i(Y_i|\boldsymbol{\theta})$.

Definice 1. *Odhad skutečné hodnoty parametru $\boldsymbol{\theta}_0$ metodou maximální věrohodnosti, který značíme jako $\hat{\boldsymbol{\theta}}_n$, definujeme jako prvek parametrického prostoru Θ splňující*

$$L_n(\hat{\boldsymbol{\theta}}_n) \geq L_n(\boldsymbol{\theta}) \quad \textit{pro všechna } \boldsymbol{\theta} \in \Theta. \tag{1.1}$$

Poznámka. Může se stát, že nerovnost (1.1) splňuje více bodů parametrického prostoru Θ, potom za odhad parametru $\boldsymbol{\theta}_0$ zvolíme libovolný z těchto bodů.

Pro přehlednost a stručnost dalšího zápisu zavedeme několik náhodných veličin.

$$R_i(\boldsymbol{\theta}) = \begin{cases} \log\left(\dfrac{f_i(Y_i|\boldsymbol{\theta})}{f_i(Y_i|\boldsymbol{\theta}_0)}\right) & \text{pro } f_i(Y_i|\boldsymbol{\theta}_0) > 0, \\ 0 & \text{jinak.} \end{cases} \tag{1.2}$$

Veličina $R_i(\boldsymbol{\theta})$ popisuje, jak se liší hustota pro daný parametr $\boldsymbol{\theta}$ od hustoty se skutečným parametrem $\boldsymbol{\theta}_0$. Pro splnění konzistence odhadu bude později třeba klást na tuto veličinu podmínky. Dále zavedeme náhodné veličiny

$$R_i(\boldsymbol{\theta},\rho) = \sup\{R_i(\boldsymbol{t}) : \|\boldsymbol{t} - \boldsymbol{\theta}\| \leq \rho\}, \tag{1.3}$$

$$V_i(r) = \sup\{R_i(\boldsymbol{\theta}) : \|\boldsymbol{\theta}\| > r\}. \tag{1.4}$$

kde $\|\cdot\|$ je Euklidovská norma na $\mathbb{R}^p$. Tyto veličiny popisují jaké nejvyšší hodnoty může nabývat R_i pro hodnoty parametru blízké danému $\boldsymbol{\theta}$, respektive vzdálené od $\boldsymbol{\theta}$ alespoň o r. Střední hodnoty uvedených veličin budeme značit malým písmenem, tedy

$$r_i(\boldsymbol{\theta},\rho) = \mathsf{E}\ R_i(\boldsymbol{\theta},\rho), \tag{1.5}$$

$$v_i(r) = \mathsf{E}\ V_i(r), \tag{1.6}$$

kde uvažujeme střední hodnotu vzhledem k hustotě ve skutečném parametru $\boldsymbol{\theta}_0$. Pokud budeme chtít zdůraznit, ve kterém parametru střední hodnotu uvažujeme, uvedeme jej v indexu, např. $\mathsf{E}_{\boldsymbol{\theta}}$. V souladu s obvyklým značením budeme označovat průměr hodnot $X_1, \ldots, X_n$ jako $\overline{X}_n$.

Pro náhodnou veličinu X a $B \geq 0$ definujeme

$$X^{(B)} = \begin{cases} X & \text{pro } X \geq -B, \\ -B & \text{jinak.} \end{cases}$$

Tedy $X^{(B)}$ je zdola „ořezaná“ verze původní náhodné veličiny.

Pro důkaz konzistence odhadu metodou maximální věrohodnosti budeme využívat stejnoměrnou integrovatelnost. Připomeneme tedy její definici a uvedeme několik výsledků, které budeme později potřebovat.

Definice 2. *(Billingsley, 2008, str. 216) Systém integrovatelných náhodných veličin $\{X_i : i \in I\}$ je stejnoměrně integrovatelný, jestliže*

$$\lim_{M\to\infty} \sup_{i\in I} \left\{ \int_{|X_i|>M} |X_i| dP \right\} = 0.$$

Lemma 1. *(Hoadley, 1971, Věta A.1) Postačující podmínka pro stejnoměrnou integrovatelnost $\{X_i : i \in I\}$ je existence kladných konstant K a δ takových, že $\mathsf{E}\,|X_i|^{1+\delta} \leq K$ pro všechna $i \in I$.*

Lemma 2. *(Hoadley, 1971, Věta A.3) Nechť U je podmnožina $\mathbb{R}^p$ a $X_i(u)$ pro pevné u je náhodná veličina. Jestliže $\{X_i(u) : i = 1,2,\ldots,u \in U\}$ je stejnoměrně integrovatelná a $\forall u \in U$ $\lim_{u\to u_0} X_i(u) = X_i$ stejnoměrně v i, P-skoro jistě, potom*

$$\lim_{u\to u_0} \mathsf{E}\,|X_i(u) - X_i| = 0 \quad \textit{stejnoměrně v } i,$$

tedy $\lim_{u\to u_0} \mathsf{E}\,X_i(u) = \mathsf{E}\,X_i$ stejnoměrně v i.

Lemma 3. *(Hoadley, 1971, Věta A.4) Nechť $\{X_i : i = 1,2,\ldots\}$ je posloupnost nezávislých náhodných veličin a dále nechť $\mu_i^{(B)} = \mathsf{E}\,X_i^{(B)}$. Jestliže $\exists\delta > 0$ a $0 < K < \infty$*

(i) $\mathsf{E}\,|X_i^{(0)}|^{1+\delta} \leq K$ *pro všechna* $i = 1,2,\ldots$

(ii) $\exists B \geq 0$ $\mu^{(B)} = \limsup_{n\to\infty} \bar{\mu}_n^{(B)} < 0,$

potom $\sum_{i=1}^n X_i \xrightarrow[n\to\infty]{P} -\infty$.

1.2 Konzistence

Nyní stanovíme podmínky pro konzistenci odhadu $\hat{\boldsymbol{\theta}}_n$ získaného metodou maximální věrohodnosti v případě nezávislých, ale ne stejně rozdělených pozorování $Y_1, \ldots, Y_n$.

C1. Θ je uzavřená podmnožina $\mathbb{R}^p$.

C2. $f_i(Y_i|\boldsymbol{\theta})$ je shora polospojitá funkce $\boldsymbol{\theta}$, stejnoměrně v i, P-skoro jistě.

C3. $\exists \delta > 0$ a $0 < K < \infty$ takové, že $\forall \boldsymbol{\theta} \in \Theta$ existuje $\rho^*(\boldsymbol{\theta}) > 0$ a $r > 0$, pro které

(i) $\mathsf{E}\,[R_i^{(0)}(\boldsymbol{\theta},\rho)]^{1+\delta} \leq K, \quad 0 \leq \rho \leq \rho^*(\boldsymbol{\theta}), \ \forall i = 1,2,\ldots$

(ii) $\mathsf{E}\,[V_i^{(0)}(r)]^{1+\delta} \leq K, \quad \forall i = 1,2,\ldots$

uvedené veličiny jsou definovány vztahy (1.3) a (1.4) a $R_k^{(0)}(\boldsymbol{\theta},\rho)$ označuje veličinu zdola omezenou 0.

C4. Existuje $B > 0$, pro které

(i) $\overline{r}^{(B)}(\boldsymbol{\theta}) = \limsup\limits_{n\to\infty} \ \overline{r}_n^{(B)}(\boldsymbol{\theta}) < 0, \quad \forall \boldsymbol{\theta} \neq \boldsymbol{\theta}_0;$

(ii) $\limsup\limits_{n\to\infty} \ \overline{v}_n^{(B)}(r) < 0$ pro r z podmínky C3.

kde hodnoty $\overline{r}_n^{(B)}(\boldsymbol{\theta})$ a $\overline{v}_n^{(B)}(r)$ jsou průměrem z $r_i^{(B)}(\boldsymbol{\theta})$, resp. $v_i^{(B)}(\boldsymbol{\theta})$ pro $i = 1,\ldots,n$, které jsme definovali vztahy (1.5), resp. (1.6).

C5. $R_i(\boldsymbol{\theta},\rho)$ a $V_i(r)$ jsou měřitelné funkce Y_i pro všechna $\boldsymbol{\theta} \in \Theta$ a $i = 1,2,\ldots$ a pro r z podmínky C3.

1.2.1 Diskuze podmínek pro konzistenci

Poznamenejme nyní, co nám říkají jednotlivé podmínky a které z nich jsou pro nás nejvíce omezující.

- První podmínkou (C1) je uzavřenost parametrického prostoru. Díky jeho uzavřenosti je možné pak některé vlastnosti dokázat pro konečně mnoho prvků tohoto parametrického prostoru, jejichž okolí tvoří pokrytí celého parametrického prostoru, příp. jeho části (viz důkaz věty 4).

  Tato podmínka není pro běžné formulace rozdělení splněna (například alternativní rozdělení s parametrem $p \in (0,1)$ nebo normální rozdělení s parametry $\mu \in (-\infty,\infty)$ a $\sigma^2 \in (0,\infty)$). Pro splnění první podmínky je ale možné odrazit se od krajů o $\varepsilon > 0$ a tento oříznutý parametrický prostor brát v úvahu i s jeho hranicemi. Většinou neuvažujeme parametr v krajních bodech otevřeného parametrického prostoru (např. alternativní rozdělení s pravděpodobností úspěchu 0 nebo 1), protože rozdělení s tímto parametrem často nedává dobrý smysl. Přesto je možné získat konzistenci odhadu i pro parametr ležící na hranici parametrického prostoru.

- Pro splnění druhé podmínky (C2) je potřeba ověřit, že pro všechna $\boldsymbol{\theta}^* \in \Theta$ platí

$$\forall \varepsilon > 0 \; \exists \delta > 0 \; \sup_{i \in \mathbb{N}} (f_i(Y_i|\boldsymbol{\theta}) - f_i(Y_i|\boldsymbol{\theta}^*)) < \varepsilon \quad \forall \boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| < \delta.$$

- Třetí podmínka (C3) zaručuje stejnoměrnou integrovatelnost pro veličiny $R_i^{(0)}(\boldsymbol{\theta},\rho)$ a $V_i^{(0)}(r)$ a umožní nám použití lemmatu 3 v důkazu konzistence. Veličiny zde uvažujeme zdola „ořezané“ nulou, a to z toho důvodu, že situace, ve kterých veličina $R_i(\boldsymbol{\theta})$, která je základem těchto veličin, nabývá velkých záporných hodnot je v případě, že $f_i(Y_i|\boldsymbol{\theta})$ je blízké nule (pro $f_i(Y_i|\boldsymbol{\theta}) = 0$ bychom dostali $R_i(\boldsymbol{\theta}) = -\infty$). Tento případ by však neměl narušit konzistenci, neboť věrohodnost je pro takovýto parametr $\boldsymbol{\theta}$ malá a je tedy nepravděpodobné, že by se jednalo o vhodný odhad $\hat{\boldsymbol{\theta}}_n$.

- Jestliže pro všechna $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ platí $f_i(Y_i|\boldsymbol{\theta}) \neq f_i(Y_i|\boldsymbol{\theta}_0)$ μ-skoro jistě (tomuto říkáme identifikovatelnost), pak lze ukázat, že $r_i(\boldsymbol{\theta}) < 0$ (Pawitan, 2001, Information inequality, str. 239). Podmínkou C4(i) požadujeme, aby platilo $\limsup_{n\to\infty} \overline{r}_n^{(B)}(\boldsymbol{\theta})$, tedy uvedená podmínka identifikovatelnosti nemusí platit pro každou hustotu zvlášť.

- V poslední podmínce (C5) požadujeme měřitelnost funkcí $R_i(\boldsymbol{\theta},\rho)$ a $V_i(r)$. Ačkoliv je tato podmínka nezbytná, abychom mohli s těmito funkcemi pracovat, v praxi se nepotkáváme s takovými funkcemi, které by nebyly měřitelné. Je na místě poznamenat, že ačkoliv hustoty pro běžná rozdělení měřitelné jsou (tedy také logaritmus jejich podílu jak jej uvažujeme v (1.3)), ale veličiny, které zde uvažujeme, získáme jako supremum přes nespočetnou množinu, kde už měřitelnost nemusí být zachována.

Věta 4. *Hoadley (1971) Pokud jsou splněny podmínky C1-C5, potom*

$$\hat{\boldsymbol{\theta}}_n \xrightarrow{P} \boldsymbol{\theta}_0 \quad \text{pro } n \to \infty.$$

Důkaz.

Vezměme libovolné $\eta > 0$ a definujme $\Theta(\eta) = \Theta \setminus U_\eta(\boldsymbol{\theta}_0)$, kde $U_\eta(\boldsymbol{\theta}_0)$ je otevřené η-okolí bodu $\boldsymbol{\theta}_0$. Stačí ukázat, že

$$\mathsf{P}[\hat{\boldsymbol{\theta}}_n \in \Theta(\eta)] \xrightarrow[n\to\infty]{} 0. \tag{1.7}$$

Z definice 1 víme, že pro odhad $\hat{\boldsymbol{\theta}}_n$ platí

$$L_n(\hat{\boldsymbol{\theta}}_n) \geq L_n(\boldsymbol{\theta}) \quad \text{pro všechna } \boldsymbol{\theta} \in \Theta.$$

Dosadíme-li za věrohodnostní funkci, dostáváme

$$\prod_{i=1}^n f_i(Y_i|\hat{\boldsymbol{\theta}}_n) \geq \prod_{i=1}^n f_i(Y_i|\boldsymbol{\theta}),$$

$$\log \prod_{i=1}^n \frac{f_i(Y_i|\hat{\boldsymbol{\theta}}_n)}{f_i(Y_i|\boldsymbol{\theta})} \geq 0 \tag{1.8}$$

pro všechna $\boldsymbol{\theta} \in \Theta$, tedy i pro $\boldsymbol{\theta}_0$. Uvažujme nyní $\hat{\boldsymbol{\theta}}_n \in \Theta(\eta)$, pak můžeme snadno ukázat

$$\sup \left\{ \log \prod_{i=1}^{n} \left(\frac{f_i(Y_i|\boldsymbol{\theta})}{f_i(Y_i|\boldsymbol{\theta}_0)} \right) : \boldsymbol{\theta} \in \Theta(\eta) \right\} \geq 0. \tag{1.9}$$

Díky (1.8) totiž víme, že alespoň pro $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_n$ je výraz v závorce větší nebo roven nule. Podle (1.2) nyní přepíšeme (1.9) jako

$$\sup \left\{ \sum_{i=1}^{n} R_i(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta(\eta) \right\} \geq 0. \tag{1.10}$$

Označíme-li levou stranu nerovnosti (1.10) jako R_n^*, pak jsme právě ukázali, že $\{\hat{\boldsymbol{\theta}}_n \in \Theta(\eta)\} \subset \{R_n^* \geq 0\}$. Pro důkaz (1.7) tedy stačí ukázat, že $R_n^* \xrightarrow[n\to\infty]{P} -\infty$.

Vezměme r z podmínky C3 a definujme

$$\begin{aligned}
\omega &= \Theta(\eta) \cap \{\|\boldsymbol{\theta}\| \leq r\}, \\
R_{n,1}^* &= \sup \left\{ \sum_{i=1}^{n} R_i(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \omega \right\}, \\
R_{n,2}^* &= \sup \left\{ \sum_{i=1}^{n} R_i(\boldsymbol{\theta}) : \|\boldsymbol{\theta}\| > r \right\}.
\end{aligned}$$

Nyní stačí ukázat, že $R_{n,1}^* \xrightarrow[n\to\infty]{P} -\infty$ a $R_{n,2}^* \xrightarrow[n\to\infty]{P} -\infty$, což provedeme ve dvou následujících částech.

1. Dokažme nejdříve $\mathbf{R}_{\mathbf{n},\mathbf{1}}^* \xrightarrow[n\to\infty]{P} -\infty$.

Množina ω, kterou v $R_{n,1}^*$ uvažujeme, je kompaktní množina. Vzhledem k tomu, že množina Θ je uzavřená, je uzavřená také množina $\Theta(\eta)$. Průnikem s množinou $\{\|\boldsymbol{\theta}\| \leq r\}$, která je uzavřená a zároveň omezená, tak dostáváme kompaktní množinu ω. Množina $\{U_{\rho(\boldsymbol{\theta})}(\boldsymbol{\theta}), \boldsymbol{\theta} \in \omega\}$, kde $U_{\rho(\boldsymbol{\theta})}(\boldsymbol{\theta})$ je otevřené okolí bodu $\boldsymbol{\theta}$ o poloměru $\rho(\boldsymbol{\theta})$, tvoří otevřené pokrytí ω. Existují tedy $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_g \in \omega$ takové, že

$$\omega \subset \bigcup_{j=1}^{g} U_{\rho(\boldsymbol{\theta}_j)}(\boldsymbol{\theta}_j).$$

Pro $R_{n,1}^*$ a $B > 0$ dále platí

$$R_{n,1}^* = \sup \left\{ \sum_{i=1}^{n} R_i(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \omega \right\} \leq \max \left\{ \sum_{i=1}^{n} R_i^{(B)}(\boldsymbol{\theta}_j, \rho(\boldsymbol{\theta}_j)) : 1 \leq j \leq g \right\}. \tag{1.11}$$

Abychom ukázali, že $R_{n,1}^* \xrightarrow[n\to\infty]{P} -\infty$, stačí ukázat, že

$$\sum_{i=}^{n} R_i^{(B)}(\boldsymbol{\theta}_j, \rho(\boldsymbol{\theta}_j)) \xrightarrow[n\to\infty]{P} -\infty$$

pro všechna $j = 1, \ldots, g$ pomocí lemmatu 3.

Nyní ověříme předpoklady lemmatu 3 pro $\mathbf{R_i^{(B)}}(\boldsymbol{\theta}_\mathbf{j},\rho(\boldsymbol{\theta}_\mathbf{j}))$.

Předpoklad (i) tohoto lemmatu je zaručen přímo podmínkou C3(i), je ovšem potřeba ověřit druhý předpoklad

$$\limsup_{n\to\infty} \bar{r}_n^{(B)}(\boldsymbol{\theta}_j,\rho(\boldsymbol{\theta}_j)) < 0 \quad \text{pro } j = 1,\ldots,g.$$

Obdobnou vlastnost předpokládáme pro $\bar{r}_n^{(B)}(\boldsymbol{\theta})$, ukážeme tedy nejprve, že

$$r_i^{(B)}(\boldsymbol{\theta},\rho) \to r_i^{(B)}(\boldsymbol{\theta}) \quad \text{pro } \rho \searrow 0$$

stejnoměrně v i, a to s pomocí lemmatu 2.

Ověříme předpoklady lemmatu 2.

Podmínka C3(i) podle lemmatu 1 implikuje, že $\{R_i^{(0)}(\boldsymbol{\theta},\rho) : i = 1,2,\ldots; 0 \leq \rho \leq \rho^*\}$ je stejnoměrně integrovatelná, a tedy i pro $B > 0$ je $\{R_i^{(B)}(\boldsymbol{\theta},\rho) : i = 1,2,\ldots; 0 \leq \rho \leq \rho^*\}$ stejnoměrně integrovatelná, což je první předpoklad lemmatu 2. Dále ukážeme, že

$$R_i^{(B)}(\boldsymbol{\theta},\rho) \searrow R_i^{(B)}(\boldsymbol{\theta}) \quad \text{pro } \rho \searrow 0$$

stejnoměrně v i, P-skoro jistě. Chceme tedy ukázat, že

$$\sup_{i\in\mathbb{N}}(R_i^{(B)}(\boldsymbol{\theta},\rho) - R_i^{(B)}(\boldsymbol{\theta})) \leq \varepsilon \quad \text{pro } \rho \searrow 0,$$

což plyne z podmínky C2, která říká, že pro všechna $\boldsymbol{\theta}^* \in \Theta$ platí

$$\forall\varepsilon > 0\ \exists\delta > 0\ \sup_{i\in\mathbb{N}}(f_i(Y_i|\boldsymbol{\theta}) - f_i(Y_i|\boldsymbol{\theta}^*)) < \varepsilon \quad \forall\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| < \delta$$

a **druhý předpoklad lemmatu 2 je splněn**.

Použijeme nyní lemma 2 na $\{R_i^{(B)}(\boldsymbol{\theta},\rho) : i = 1,2,\ldots; 0 \leq \rho \leq \rho^*\}$ a dostáváme

$$r_i^{(B)}(\boldsymbol{\theta},\rho) \to r_i^{(B)}(\boldsymbol{\theta}) \text{ pro } \rho \searrow 0, \text{ stejnoměrně v } i$$

S touto znalostí pokračujme v ověření předpokladů lemmatu 3. Z definice stejnoměrné konvergence máme, že pro každé $\boldsymbol{\theta} \in \omega$ existuje $\rho(\boldsymbol{\theta}) \leq \rho^*$, pro které

$$r_i^{(B)}(\boldsymbol{\theta},\rho) - r_i^{(B)}(\boldsymbol{\theta}) < \varepsilon \quad \text{pro všechna } i \in \mathbb{N}. \tag{1.12}$$

Zvolíme-li v (1.12) $\varepsilon = -\dfrac{\bar{r}^{(B)}(\theta)}{2}$ (toto je dle podmínky C4 kladné), dostáváme

$$r_i^{(B)}(\boldsymbol{\theta},\rho) - r_i^{(B)}(\boldsymbol{\theta}) < -\frac{\bar{r}^{(B)}(\boldsymbol{\theta})}{2} \quad \text{pro všechna } i \in \mathbb{N}. \tag{1.13}$$

Obě strany nerovnosti (1.13) nyní sečteme přes i, kde $i \in \{1,\ldots,n\}$ a vydělíme n.

$$\bar{r}_n^{(B)}(\boldsymbol{\theta},\rho) - \bar{r}_n^{(B)}(\boldsymbol{\theta}) < -\frac{\bar{r}^{(B)}(\boldsymbol{\theta})}{2}. \tag{1.14}$$

Na obě strany (1.14) nyní použijeme lim sup a dostáváme

$$\begin{aligned}
\limsup_{n\to\infty} \overline{r}_n^{(B)}(\boldsymbol{\theta},\rho) - \limsup_{n\to\infty} \overline{r}_n^{(B)}(\boldsymbol{\theta}) &< \limsup_{n\to\infty} -\frac{\overline{r}^{(B)}(\boldsymbol{\theta})}{2} \\
\limsup_{n\to\infty} \overline{r}_n^{(B)}(\boldsymbol{\theta},\rho) - \overline{r}^{(B)}(\boldsymbol{\theta}) &< -\frac{\overline{r}^{(B)}(\boldsymbol{\theta})}{2} \\
\limsup_{n\to\infty} \overline{r}_n^{(B)}(\boldsymbol{\theta},\rho) &< \frac{\overline{r}^{(B)}(\boldsymbol{\theta})}{2} \quad \text{pro všechna } \boldsymbol{\theta} \in \omega.
\end{aligned}$$

Podle podmínky C4(i) dostáváme, že

$$\limsup_{n\to\infty} \overline{r}_n^{(B)}(\boldsymbol{\theta},\rho) < 0 \quad \text{pro všechna } \boldsymbol{\theta} \in \omega. \tag{1.15}$$

Platnost vztahu (1.15) pro všechna $\boldsymbol{\theta} \in \omega$ zajišťuje jeho platnost i pro $\boldsymbol{\theta}_j$, $j = 1, \ldots, g$. Ukázali jsme tak, že je splněn i druhý předpoklad lemmatu 3 pro množinu $\{R_i^{(B)}(\boldsymbol{\theta}_j,\rho(\boldsymbol{\theta}_j)) : i = 1,2,\ldots\}$, z čehož plyne

$$\sum_{i=1}^{n} R_i^{(B)}(\boldsymbol{\theta}_j,\rho(\boldsymbol{\theta}_j)) \xrightarrow[n\to\infty]{P} -\infty.$$

Ze vztahu (1.11) tedy plyne, že také $R_{n,1}^* \xrightarrow[n\to\infty]{P} -\infty$.

2. Zbývá ukázat, že také $\mathbf{R}_{\mathbf{n,2}}^* \xrightarrow[n\to\infty]{P} -\infty$.

Použijeme opět lemma 3, nyní pro $V_i(r)$. Splnění předpokladů zajišťují podmínky C3(ii) a C4(ii), a tedy přímo dostáváme, že $\sum_{i=1}^{n} V_i(r) \xrightarrow[n\to\infty]{P} -\infty$. Odtud již snadno vidíme, že

$$\begin{aligned}
R_{n,2}^* &= \sup\left\{\sum_{i=1}^{n} R_k(\boldsymbol{\theta}) : \|\boldsymbol{\theta}\| > r\right\} \\
&\le \sum_{i=1}^{n} \sup\left\{R_i(\boldsymbol{\theta}) : \|\boldsymbol{\theta}\| > r\right\} \\
&= \sum_{i=1}^{n} V_i(r) \xrightarrow[n\to\infty]{P} -\infty.
\end{aligned}$$

Tímto je důkaz dokončen.

□

1.3 Asymptotická normalita

Pro statistickou inferenci nestačí znát pouze vhodný odhad zkoumaného parametru, který je konzistentní, ale také rozdělení tohoto odhadu, ať už přesné nebo asymptotické. V této kapitole stanovíme podmínky pro asymptotickou normalitu maximálně věrohodného odhadu při nezávislých, ale ne stejně rozdělených pozorováních.

Pro stručnější zápis podmínek a pohodlnější zápis v důkazu si opět zavedeme pomocné veličiny. Označíme si postupně logaritmus hustoty $f_i(y|\boldsymbol{\theta})$ a také jeho první a druhou derivaci podle $\boldsymbol{\theta}$, respektive vektor prvních parciálních derivací a matici druhých parciálních derivací v případě, že $\boldsymbol{\theta} \in \mathbb{R}^p$, kde $p > 1$. Tedy

$$\boldsymbol{\Phi}_i(y,\boldsymbol{\theta}) = \log f_i(y|\boldsymbol{\theta}),$$

p-rozměrný vektor prvních parciálních derivací označíme jako $\boldsymbol{\Phi}'_i(y,\boldsymbol{\theta})$, jehož prvky jsou $\boldsymbol{\Phi}'_{i,j}(y,\boldsymbol{\theta}) = \frac{\partial}{\partial\theta_j}\boldsymbol{\Phi}_i(y,\boldsymbol{\theta})$. Matice druhých parciálních derivací má rozměr $p \times p$ a tvoří ji prvky $\boldsymbol{\Phi}'_{i,jl}(y,\boldsymbol{\theta}) = \frac{\partial^2}{\partial\theta_j\partial\theta_l}\boldsymbol{\Phi}_i(y,\boldsymbol{\theta})$. Pro důkaz asymptotické normality dále využijeme stejnoměrnou verzi slabého zákona velkých čísel, kterou zde uvedeme jako pomocné lemma pro pozdější důkaz.

Lemma 5 (Stejnoměrný slabý zákon velkých čísel)**.** *(Hoadley, 1971, Věta A.5.) Nechť $\{Y_i : i = 1,2,\ldots\}$ jsou nezávislé náhodné veličiny se stavovým prostorem $(\mathcal{Y},\mathcal{A})$. Nechť $H_i : \mathcal{Y} \times S \to \mathbb{R}$, kde $S \subset \mathbb{R}^p$ je kompaktní množina. Označme $h_i(\mathbf{s}) = \mathsf{E}\, H_i(Y_i,\mathbf{s})$. Jestliže*

(i) $H_i(Y_i,\mathbf{s})$ je $\mathcal{A}$-měřitelná pro všechna $\mathbf{s} \in S$,

(ii) $H_i(Y_i,\mathbf{s})$ je spojitá na S, stejnoměrně v i, P-skoro jistě,

(iii) existuje měřitelná funkce $B_i : \mathcal{Y} \to \mathbb{R}$ taková, že $|H_i(y,\mathbf{s})| < B_i(y)$ pro všechna $\mathbf{s} \in S$ a $\exists\delta > 0$ a $0 < K < \infty$ takové, že $\mathsf{E}\,|B_i(Y_i)|^{1+\delta} \leq K$,

potom

(i) $h_i(\mathbf{s})$ je spojitá na S, stejnoměrně v i,

(ii) $\sup_{\mathbf{s}\in S}\{|\frac{1}{n}\sum_{i=1}^n H_i(Y_i,\mathbf{s}) - \overline{h}_n(\mathbf{s})|\} \xrightarrow[n\to\infty]{P} 0$.

1.3.1 Podmínky pro asymptotickou normalitu

Následuje seznam podmínek, který zaručí asymptotickou normalitu odhadu metodou maximální věrohodnosti v případě, kdy pozorování jsou nezávislá, ale ne stejně rozdělená.

N1. $\boldsymbol{\theta}_0$ leží uvnitř parametrického prostoru Θ (dále označme jako $\mathrm{int}(\Theta)$).

N2. $\hat{\boldsymbol{\theta}}_n \xrightarrow[n\to\infty]{P} \boldsymbol{\theta}_0$.

N3. $\boldsymbol{\Phi}'_i(Y_i,\boldsymbol{\theta})$ a $\boldsymbol{\Phi}''_i(Y_i,\boldsymbol{\theta})$ existují P-skoro jistě.

N4. $\boldsymbol{\Phi}''_i(Y_i,\boldsymbol{\theta})$ je spojitá funkce v $\boldsymbol{\theta}$, stejnoměrně v i, P-skoro jistě a je měřitelnou funkcí Y_i.

N5. $\mathsf{E}_{\boldsymbol{\theta}}[\boldsymbol{\Phi}'_i(Y_i,\boldsymbol{\theta})] = \mathbf{0}_p \quad$ pro $\forall i \in \mathbb{N}$ a $\forall\boldsymbol{\theta} \in \Theta$.

N6. $I_i(\boldsymbol{\theta}) = \mathsf{E}_{\boldsymbol{\theta}}[\boldsymbol{\Phi}'_i(Y_i,\boldsymbol{\theta})(\boldsymbol{\Phi}'_i(Y_i,\boldsymbol{\theta}))^\top] = -\,\mathsf{E}_{\boldsymbol{\theta}}[\boldsymbol{\Phi}''_i(Y_i,\boldsymbol{\theta})] \quad$ pro $i \in \mathbb{N}$ a $\forall\boldsymbol{\theta} \in \Theta$.

N7. $\overline{I}_n(\boldsymbol{\theta}) \xrightarrow[n\to\infty]{} \overline{I}(\boldsymbol{\theta})\ \ \forall\boldsymbol{\theta} \in \Theta$, kde $\overline{I}(\boldsymbol{\theta})$ je pozitivně definitní.

Připomeňme, že $\overline{I}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^n I_i(\boldsymbol{\theta})$.

N8. Pro nějaké $\delta > 0$, $\sum_{i=1}^{n} \mathsf{E}\, |\boldsymbol{\lambda}^T \boldsymbol{\Phi}_i'(Y_i,\boldsymbol{\theta}_0)|^{2+\delta}/n^{(2+\delta)/2} \xrightarrow[n\to\infty]{} 0$ pro všechna $\boldsymbol{\lambda} \in \mathbb{R}^p$.

N9. Existuje $\varepsilon > 0$, $\delta > 0$, $0 < K < \infty$ a náhodné veličiny $B_{i,jl}(Y_i)$ takové, že $\forall i, j, l$

(i) $\sup \{|\boldsymbol{\Phi}''_{i,jl}(Y_i,\boldsymbol{t})| : \| \boldsymbol{t} - \boldsymbol{\theta}_0 \| \leq \varepsilon\} \leq B_{i,jl}(Y_i)$,

(ii) $\mathsf{E}\, |B_{i,jl}(Y_i)|^{1+\delta} \leq K$.

1.3.2 Diskuze podmínek pro asymptotickou normalitu

- První podmínka (N1) požaduje, aby skutečný parametr ležel uvnitř parametrického prostoru. Pro běžné formulace pravděpodobnostních rozdělení je tato podmínka splněna, jak už bylo zmíněno v poznámkách k podmínkám pro konzistenci odhadu.

- Dále v podmínce N2 předpokládáme konzistenci odhadu (potažmo tedy všechny podmínky pro konzistenci), kterou potřebujeme, abychom ukázali, že asymptotický rozptyl $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ konverguje k inverzi očekávané informační matice ve skutečném parametru. Ačkoliv tedy tato podmínka znamená splnění podmínek C1-C5, nedává dobrý smysl snažit se ukázat asymptotickou normalitu pro nekonzistentní odhady.

- Ve třetí podmínce (N3) požadujeme existenci prvních a druhých parciálních derivací funkce $\boldsymbol{\Phi}_i(Y_i,\boldsymbol{\theta})$ podle parametru $\boldsymbol{\theta}$. Jedná se o velmi důležitou podmínku, neboť na derivacích je kromě jiných podmínek založena očekávaná informační matice. Pro běžně používaná rozdělení bývá tato podmínka splněna.

- Další podmínka (N4) požaduje spojitost $\boldsymbol{\Phi}_i''(Y_i,\boldsymbol{\theta})$ stejnoměrně v i a její měřitelnost v Y_i. Poznamenejme, že pro vícerozměrný parametr $\boldsymbol{\theta}$ představuje $\boldsymbol{\Phi}_i''(Y_i,\boldsymbol{\theta})$ matici $p \times p$, a tedy požadujeme výše uvedené vlastnosti pro její prvky.

- Podmínky N5 a N6 odpovídají podmínkám, které potřebujeme i v případě nezávislých a stejně rozdělených pozorování.

- Nazveme-li matici $I_i(\boldsymbol{\theta})$ očekávanou informační maticí pro i-té pozorování, klademe v další podmínce (N7) požadavek na průměr z těchto očekávaných informačních matic přes všechna pozorování. Tento průměr má konvergovat k pozitivně definitní matici $\overline{I}(\boldsymbol{\theta})$, kterou díky této vlastnosti budeme moci invertovat. Podmínku N7 společně s podmínkou N8 potřebujeme, abychom mohli použít Ljapunovovu centrální limitní větu na vektory $\boldsymbol{\Phi}_i'(Y_i,\boldsymbol{\theta}_0)$.

- Poslední podmínka (N9) zajišťuje, že je veličina $\boldsymbol{\Phi}''_{i,jl}(Y_i,\boldsymbol{t})$ v okolí $\boldsymbol{\theta}_0$ omezena stejnoměrně integrovatelnou náhodnou veličinou. Tato podmínka je potřebná při dokazování podoby asymptotického rozptylu $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$.

1.4 Srovnání s případem stejně rozdělených pozorování

Srovnejme nyní podmínky, které potřebujeme pro asymptotickou normalitu v případě, kdy nemáme stejně rozdělená pozorování, s případem, kdy se jedná o stejně rozdělené veličiny. Uveďme nyní podmínky regularity pro případ, kdy $Y_1, \ldots, Y_n$ jsou nezávislé a stejně rozdělené náhodné veličiny s hustotou $f(y|\boldsymbol{\theta})$, kde $\boldsymbol{\theta} \in \mathbb{R}^p$ a skutečná hodnota parametru $\boldsymbol{\theta}$ je $\boldsymbol{\theta}_0$. Označme obecně Y jako veličinu s hustotou $f(y|\boldsymbol{\theta})$.

Poznámka. Formulace podmínek regularity se v různých literaturách mírně odlišuje, zde použijeme jejich podobu podle Lehmann a Casella (2003, str.462).

- R1. Existuje otevřené okolí skutečného parametru $\boldsymbol{\theta}_0$, které leží v Θ takové, že pro skoro všechna y můžeme hustotu $f(y|\boldsymbol{\theta})$ třikrát derivovat pro všechna $\boldsymbol{\theta} \in U_\delta(\boldsymbol{\theta}_0)$.

- R2. Pro všechna $\boldsymbol{\theta} \in \Theta$ platí

  (i) $\mathsf{E}\left[\frac{\partial \log f(Y|\boldsymbol{\theta})}{\partial \theta_j}\right] = 0 \quad \forall j = 1, \ldots, p,$

  (ii) $\mathsf{E}\left[-\frac{\partial^2 \log f(Y|\boldsymbol{\theta})}{\partial \theta_j \partial \theta_l}\right] = \mathsf{E}\left[\frac{\partial \log f(Y|\boldsymbol{\theta})}{\partial \theta_j}\frac{\partial \log f(Y|\boldsymbol{\theta})}{\partial \theta_l}\right] = I_{jl}(\theta) \ \ \forall j,l = 1, \ldots, p.$

- R3. Fisherova informační matice $I(\boldsymbol{\theta})$ je pozitivně definitní pro všechna $\boldsymbol{\theta} \in U_\delta(\boldsymbol{\theta}_0)$ a její prvky $I_{jl}(\boldsymbol{\theta})$ jsou konečné.

- R4. Pro všechna $\boldsymbol{\theta} \in U_\delta(\boldsymbol{\theta}_0)$ a pro všechna $j,l,m \in \{1, \ldots, p\}$ existují funkce M_{jlm} takové, že platí

$$\left|\frac{\partial^3 \log f(y|\boldsymbol{\theta})}{\partial \theta_j \partial \theta_l \partial \theta_m}\right| \leq M_{jlm}(y),$$
$$\mathsf{E}\left[M_{jlm}(Y)\right] < \infty.$$

Souhrnně, před rozebráním jednotlivých bodů, můžeme uvést, že v případě nezávislých a stejně rozdělených pozorování máme jedinou hustotu, která je shodná pro všechna pozorování.

- Podmínka na to, aby skutečný parametr ležel uvnitř parametrického prostoru je zachována v obou případech.

- Požadavek na existenci derivací, respektive parciálních derivací, je také shodný v obou případech (odpovídá podmínce R1). Spojitost druhých parciálních derivací (N4) je také obsažena v podmínce R2.

- Podmínky N5 a N6 odpovídají podmínce regularity R2.

- Vezmeme-li nyní podmínku N7, která hovoří o konvergenci matice $\overline{I}_n(\boldsymbol{\theta})$, musíme si nejdříve uvědomit, že v případě stejně rozdělených a nezávislých pozorování jsou očekávané informační matice pro všechna pozorování stejné, a tedy průměrem z nich dostáváme opět očekávanou informační matici pro jedno pozorování, a tedy podmínka N7 odpovídá podmínce R3 z podmínek regularity.

- Podmínky N8 a N9 pro případ, kdy máme stejně rozdělená a nezávislá pozorování, nepotřebujeme, neboť jejich potřeba vychází z nutnosti použít centrální limitní větu pro náhodné vektory, které nejsou stejně rozdělené. V případě se stejně rozdělenými pozorováními si poradíme i bez těchto podmínek.

Věta 6. *Hoadley (1971) Jestliže jsou podmínky N1 až N9 splněny, potom*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}_p(\mathbf{0}_{\boldsymbol{p}}, \overline{I}^{-1}(\boldsymbol{\theta}_0)).$$

Důkaz. Předpokládejme, že parametrický prostor Θ je otevřená množina a dále, že odhad $\hat{\boldsymbol{\theta}}_n$ je konzistentním odhadem skutečného parametru $\boldsymbol{\theta}_0$ (platnost podmínek N1 a N2). Tedy existuje $\eta > 0$ takové, že $U_\eta(\boldsymbol{\theta}_0) \subset \Theta$ a $\mathsf{P}[\hat{\boldsymbol{\theta}}_n \in U_\eta(\boldsymbol{\theta}_0)] = 1 - \varepsilon_n$, pro $\varepsilon_n \to 0$. Skórová funkce v bodě $\hat{\boldsymbol{\theta}}_n$ je s pravděpodobností $1 - \varepsilon_n$ rovna nule, tedy

$$\frac{\partial}{\partial \boldsymbol{\theta}} \log L_n(\boldsymbol{\theta})\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n} = \mathbf{0}_{\boldsymbol{p}},$$

což můžeme přepsat pomocí výše zavedeného značení jako

$$\sum_{i=1}^{n} \boldsymbol{\Phi}_i'(Y_i, \hat{\boldsymbol{\theta}}_n) = \mathbf{0}_{\boldsymbol{p}}. \tag{1.16}$$

Definujme vektorovou funkci $\Psi_i(\gamma) = \boldsymbol{\Phi}_i'(y, \boldsymbol{\theta} + \gamma(\boldsymbol{t} - \boldsymbol{\theta}))$. Využijeme nyní základní větu integrálního počtu (Billingsley, 2008, str.224) pro nově definovanou funkci $\Psi_i(\gamma)$ v tomto tvaru

$$\Psi_i(1) - \Psi_i(0) = \int_0^1 \Psi'(\xi) d\xi \tag{1.17}$$

a aplikujeme ji na jednotlivé prvky vektorové funkce $\Psi_i(\gamma)$. Pokud dosadíme do (1.17) podle definice funkce Ψ, pak dostáváme

$$\begin{aligned}\boldsymbol{\Phi}_i'(y, \boldsymbol{t}) - \boldsymbol{\Phi}_i'(y, \boldsymbol{\theta}) &= \int_0^1 \frac{d}{d\gamma} \boldsymbol{\Phi}_i'(y, \boldsymbol{\theta} + \gamma(\boldsymbol{t} - \boldsymbol{\theta}))\bigg|_{\gamma=\xi} d\xi \\ &= \left(\int_0^1 \boldsymbol{\Phi}_i''(y, \boldsymbol{\theta} + \xi(\boldsymbol{t} - \boldsymbol{\theta})) d\xi\right)(\boldsymbol{t} - \boldsymbol{\theta}).\end{aligned} \tag{1.18}$$

Položíme-li nyní $y = Y_i$, $\boldsymbol{t} = \hat{\boldsymbol{\theta}}_n$, $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, sečteme rovnici (1.18) přes i, kde $i = 1, \ldots, n$ máme

$$\sum_{i=1}^{n} \boldsymbol{\Phi}_i'(Y_i, \hat{\boldsymbol{\theta}}_n) - \sum_{i=1}^{n} \boldsymbol{\Phi}_i'(Y_i, \boldsymbol{\theta}_0) = \sum_{i=1}^{n} \left(\int_0^1 \boldsymbol{\Phi}_i''(Y_i, \boldsymbol{\theta}_0 + \xi(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)) d\xi\right)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0). \tag{1.19}$$

Použitím (1.16) se zjednoduší levá strana rovnice (1.19), rovnici poté ještě vynásobíme $-n^{-\frac{1}{2}}$ a dostaneme:

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} \boldsymbol{\Phi}_i'(Y_i, \boldsymbol{\theta}_0) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \left(-\int_0^1 \boldsymbol{\Phi}_i''(Y_i, \boldsymbol{\theta}_0 + \xi(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)) d\xi\right)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0). \tag{1.20}$$

Označme nyní

$$\Gamma_n = \frac{1}{n} \sum_{i=1}^{n} \left(-\int_0^1 \boldsymbol{\Phi}_i''(Y_i, \boldsymbol{\theta}_0 + \xi(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)) d\xi\right) \tag{1.21}$$

a poté můžeme (1.20) přepsat do tvaru

$$n^{-1/2}\sum_{i=1}^{n}\mathbf{\Phi}_i'(Y_i,\boldsymbol{\theta}_0)=\Gamma_n[n^{1/2}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)]. \tag{1.22}$$

Na pravé straně (1.22) máme nyní, kromě matice I_n, výraz jehož asymptotickou normalitu chceme ukázat. Využijeme k tomu vyjádření na levé straně (1.22), pro které budeme moci použít vhodnou centrální limitní větu a nakonec ukážeme, že $\Gamma_n \xrightarrow{P} \overline{I}(\boldsymbol{\theta}_0)$ a tím bude důkaz dokončen.

Náhodné p-rozměrné vektory $\mathbf{\Phi}_i'(Y_i,\boldsymbol{\theta}_0)$ jsou díky nezávislosti Y_i také nezávislé, mají nulovou střední hodnotu (podmínka N5) a rozptylovou matici $I_i(\boldsymbol{\theta})$ (podmínka N6). Na vektory $\mathbf{\Phi}_i'(Y_i,\boldsymbol{\theta}_0)$ nyní použijeme mnohorozměrnou Ljapunovovu centrální limitní větu (uvedena v příloze jako věta A.1). Předpoklady této věty jsou díky dalším podmínkám pro asymptotickou normalitu splněny. Podmínka N7 zajišťuje, že průměr z rozptylových matic jde k nějaké konečné, pozitivně definitní matici $\overline{I}(\boldsymbol{\theta})$ (předpoklad (i) věty A.1), a podmínka N8 je přímo Ljapunovovou podmínkou (předpoklad (ii) věty A.1), jejíž platnost jsme chtěli ověřit. Předpoklady mnohorozměrné Ljapunovovy centrální limitní věty pro vektory $\mathbf{\Phi}_i'(Y_i,\boldsymbol{\theta}_0)$ jsou splněny, z čehož plyne, že

$$n^{-1/2}\sum_{i=1}^{n}\mathbf{\Phi}_i'(Y_i,\boldsymbol{\theta}_0)\xrightarrow[n\to\infty]{\mathcal{D}}\mathcal{N}_p(\mathbf{0}_p,\overline{I}(\boldsymbol{\theta}_0)).$$

Nyní zbývá ukázat, že $\Gamma_n \xrightarrow[n\to\infty]{P} \overline{I}(\boldsymbol{\theta}_0)$.

Dle podmínky N7 víme, že $\overline{I}_n(\boldsymbol{\theta}) \xrightarrow[n\to\infty]{} \overline{I}(\boldsymbol{\theta})$. Ukážeme tedy nejdříve, že Γ_n po prvcích konverguje stejnoměrně v pravděpodobnosti k matici $\overline{I}_n(\boldsymbol{\theta})$. Matice Γ_n je dána předpisem (1.21), kde prvky matice $-\mathbf{\Phi}_i''(Y_i,\mathbf{s})$ tvoří náhodné veličiny $-\mathbf{\Phi}_{i,jl}''(Y_i,\mathbf{s})$. Střední hodnoty těchto veličin tvoří prvky matice $\overline{I}_n(\boldsymbol{s})$. Pro důkaz tedy použijeme stejnoměrný slabý zákon velkých čísel (uvedený dříve jako lemma 5). Pro ověření předpokladů lemmatu pro veličiny $-\mathbf{\Phi}_{i,jl}''(Y_i,\mathbf{s})$ a pro $\mathbf{s}\in S$, kde $S=\{\mathbf{s};\parallel \mathbf{s}-\boldsymbol{\theta}_0\parallel\leq\varepsilon\}$, použijeme podmínky N6, N4 a N9. Podmínka N6 popisuje, jak vypadají prvky matice $\overline{I}_i(\boldsymbol{\theta})$. Podmínka N4 přímo odpovídá předpokladům (*i*) a (*ii*) lemmatu 5 a předpoklad (*iii*) je splněn díky podmínce N9. Dostáváme tedy

$$\sup_{\mathbf{s}\in S}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}-\mathbf{\Phi}_{i,jl}''(Y_i,\mathbf{s})-\overline{I}_{n,jl}(\mathbf{s})\right|\right\}\xrightarrow[n\to\infty]{P}0. \tag{1.23}$$

Jestliže v (1.23) vezmeme $\mathbf{s}=\boldsymbol{\theta}_0+\xi(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)$, kde $0\leq\xi\leq1$, a použijeme podmínku N7 ($\overline{I}_n(\boldsymbol{\theta})\xrightarrow[n\to\infty]{}\overline{I}(\boldsymbol{\theta})$), dostaneme

$$\frac{1}{n}\sum_{i=1}^{n}-\mathbf{\Phi}_{i,jl}''(Y_i,\boldsymbol{\theta}_0+\xi(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0))\xrightarrow[n\to\infty]{P}\overline{I}_{jl}(\boldsymbol{\theta}_0+\xi(\mathbf{t}-\boldsymbol{\theta}_0)) \tag{1.24}$$

stejnoměrně v ξ a pro $\parallel\mathbf{t}-\boldsymbol{\theta}_0\parallel\leq\varepsilon$, neboť $\parallel\mathbf{s}-\boldsymbol{\theta}_0\parallel\leq\parallel\mathbf{t}-\boldsymbol{\theta}_0\parallel$ a podle lemmatu 5 konvergence platí na ε-okolí skutečného parametru.

Nyní už stačí ukázat, že výraz na pravé straně (1.24) konverguje k požadovanému $\overline{I}_{jl}(\boldsymbol{\theta}_0)$. Abychom to ukázali, využijeme druhé části lemmatu 5 (ii), která nám dává spojitost $\overline{I}_{i,jl}$ na S, stejnoměrně v i, tedy platí

$$\lim_{\mathbf{t}\to\boldsymbol{\theta}_0} \overline{I}_{i,jl}(\boldsymbol{\theta}_0 + \xi(\mathbf{t} - \boldsymbol{\theta}_0)) = \overline{I}_{i,jl}(\boldsymbol{\theta}_0) \quad \text{stejnoměrně v } i.$$

Díky podmínce N7 ($\overline{I}_n(\boldsymbol{\theta}) \xrightarrow[n\to\infty]{} \overline{I}(\boldsymbol{\theta})$) pak dostáváme spojitost i pro $\overline{I}_{jl}$, tedy pravá strana (1.24) pro $t \to \boldsymbol{\theta}_0$ jde skutečně k $\overline{I}_{jl}(\boldsymbol{\theta}_0)$. Jelikož v našem výrazu (1.21) je $\mathbf{t} = \hat{\boldsymbol{\theta}}_n$, platí díky konzistenci odhadu (podmínka N2), že $\Gamma_n \xrightarrow[n\to\infty]{P} \overline{I}(\boldsymbol{\theta}_0)$ a důkaz je tímto dokončen.

□

1.5 Příklad: K-výběrový problém

Je mnoho situací, kdy se setkáváme s nezávislými, ale ne stejně rozdělenými pozorováními. Jedním z možných příkladů je K-výběrový problém. Představme si K-výběrový problém, ve kterém se jedná o výběry z alternativního rozdělení s různou pravděpodobností úspěchu pro každý výběr, označme ji p_k pro $k = 1, \dots, K$, kde $p_k \in (0,1)$. Naším cílem je odhad parametru $\boldsymbol{p} = (p_1, \dots, p_K)^\top$ a inference o něm. Označme skutečný parametr jako $\boldsymbol{p}_0 = (p_{10}, \dots, p_{K0})^\top$. Předpokládejme, že pozorované veličiny jsou následujícího tvaru

$$\begin{aligned} Y_{11}, \dots, Y_{1n_1} &\sim Alt(p_1), \\ Y_{21}, \dots, Y_{2n_2} &\sim Alt(p_2), \\ &\vdots \\ Y_{K1}, \dots, Y_{Kn_K} &\sim Alt(p_K), \end{aligned} \tag{1.25}$$

kde $n_1 + n_2 + \ldots + n_K = n$ a výběry jsou mezi sebou nezávislé. Jedná se tedy o případ, kdy jsou pozorování nezávislá, ale nepocházejí ze stejného rozdělení (případ, který jsme uvedli v části 1.1). Nyní ověříme, že jsou pro tento příklad splněny podmínky pro konzistenci a asymptotickou normalitu odhadu parametru $\boldsymbol{p}$.

Poznamenejme, že pro tento případ je možné konzistenci odhadu parametru $\boldsymbol{p}$ a jeho asymptotickou normalitu ukázat i jednodušeji. Konzistenci odhadu $\hat{\boldsymbol{p}}_n$ získáme po složkách, neboť odhad $\hat{p}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} Y_{k,i}$ je nestranný a var $\hat{p}_k = \frac{p_k(1-p_k)}{n_1}$, tedy pro $n_1 \to \infty$ konverguje k 0, z čehož plyne, že odhad $\hat{p}_k$ je konzistentní. Po složkách dostaneme konzistenci $\hat{\boldsymbol{p}}_n$ pro $\min(n_1, \dots, n_K) \to \infty$. Asymptotickou normalitu $\hat{\boldsymbol{p}}_n$ získáme z centrální limitní věty a díky nezávislosti jednotlivých výběrů. Aplikace obecného přístupu na tento příklad tedy pouze ilustruje použití vět 4 a 6.

Pro ověřování podmínek využijeme zapsání jednotlivých výběrů do sdruženého výběru $\{Y_1, \dots, Y_n\}$. Označíme jednotlivé výběry postupně jako V_1 až V_K a poté můžeme psát $Y_i \in V_k$, což znamená, že pozorování Y_i pochází z k-tého výběru, tedy z alternativního rozdělení s pravděpodobností úspěchu p_k. Začneme ověřováním podmínek pro konzistenci.

Ověření podmínek pro konzistenci

C1.

První podmínka požaduje, aby byl parametrický prostor, zde $\Theta = (0,1)^K$, uzavřený. Omezíme se tedy na prostor $\langle \varepsilon, 1-\varepsilon \rangle^K$ pro malé $\varepsilon > 0$, abychom se vyhnuli technickým obtížím na hranici parametrického prostoru a zároveň splnili první podmínku. Rozšíření pro $p_k = 0$ a $p_k = 1$ je možné, ale zbytečně technicky náročné a z praktického hlediska nedává příliš smysluplnou interpretaci (nulová pravděpodobnost úspěchu nebo pravděpodobnost úspěchu rovna jedné).

C2.

V druhé podmínce je potřeba ověřit, že $f_i(Y_i|\boldsymbol{p})$ je shora polospojitá funkce $\boldsymbol{p}$, stejnoměrně v i, P-skoro jistě. Uvědomme si, jak vypadají funkce f_i v tomto případě. Jelikož všechna pozorování pocházejí z alternativního rozdělení, můžeme jednoduše psát $f_i(Y_i|\boldsymbol{p}) = p_k^{Y_i}(1-p_k)^{1-Y_i}$ pro $Y_i \in V_k$, $i = 1, \ldots, n$. Máme tedy celkem K možností, jak může hustota vypadat. Chceme ukázat, že $\forall \boldsymbol{p}^* \in \Theta$

$$\forall \varepsilon > 0 \; \exists \delta > 0 : \sup_{i \in \{1,\ldots,n\}} \left(f_i(Y_i|\boldsymbol{p}) - f_i(Y_i|\boldsymbol{p}^*) \right) < \varepsilon \quad \text{pro } \boldsymbol{p} \in U_\delta(\boldsymbol{p}^*).$$

Vzhledem k tomu, že veličina Y_i může nabývat pouze hodnot 0 a 1, můžeme psát

$$\begin{aligned} Y_i = 1: &\quad f_i(Y_i|\boldsymbol{p}) - f_i(Y_i|\boldsymbol{p}^*) = \boldsymbol{p} - \boldsymbol{p}^*, \\ Y_i = 0: &\quad f_i(Y_i|\boldsymbol{p}) - f_i(Y_i|\boldsymbol{p}^*) = 1 - \boldsymbol{p} - (1 - \boldsymbol{p}^*) = \boldsymbol{p}^* - \boldsymbol{p}, \end{aligned}$$

tedy pro $\boldsymbol{p} \in U_\delta(\boldsymbol{p}^*)$, stačí jen volit $\delta < \varepsilon$ a tato podmínka je splněna.

C3.

(i) Ve třetí podmínce chceme nejprve ověřit, zda existuje $\rho^*(\boldsymbol{p}) > 0$ a kladné K a δ takové, že

$$\mathsf{E}\,[R_i^{(0)}(\boldsymbol{p},\rho)]^{1+\delta} \leq K, \quad 0 \leq \rho \leq \rho^*.$$

Podíváme se, jak vypadá v tomto případě veličina $R_k(\boldsymbol{p})$. Podle vztahu (1.2) dostáváme

$$R_i(\boldsymbol{p}) = \log\left(\frac{p_k^{Y_i}(1-p_k)^{1-Y_i}}{p_{k0}^{Y_i}(1-p_{k0})^{1-Y_i}} \right) \quad \text{pro } Y_i \in V_k,$$

kde Y_i nabývá pouze hodnot 0 a 1. Pro konkrétní i máme

$$\left.\begin{aligned} Y_i = 1: &\quad R_i(\boldsymbol{p}) = \log\left(\frac{p_k}{p_{k0}}\right) \\ Y_i = 0: &\quad R_i(\boldsymbol{p}) = \log\left(\frac{1-p_k}{1-p_{k0}}\right) \end{aligned}\right\} < \log\left(\frac{p_k}{p_{k0}}\right) \qquad \text{pro} \quad p_{k0} < p_k,$$

a pro $p_{k0} \geq p_k$ dostaneme

$$R_i(\boldsymbol{p}) \leq \log\left(\frac{1-p_k}{1-p_{k0}}\right).$$

tedy pro jakékoliv p_k můžeme psát

$$R_i(\boldsymbol{p}) \leq \max\left(\log\left(\frac{p_k}{p_{k0}}\right), \log\left(\frac{1-p_k}{1-p_{k0}}\right)\right),$$

tedy máme výraz, který nezávisí na Y_i. Veličina $R_i(\boldsymbol{p},\rho)$, která je supremem z $R_i(\mathbf{t})$ pro $\|\boldsymbol{t}-\boldsymbol{\theta}\| \leq \rho$, také nezávisí na Y_i a je menší nebo rovna uvedenému výrazu. Při výpočtu střední hodnoty tedy tyto výrazy můžeme nahradit konstantou a poté najdeme vhodné K a δ, pro které je splněna podmínka C3(i).

(ii) V druhé části třetí podmínky, chceme ověřit, zda i pro velká $\boldsymbol{p}$ je rozdíl v logaritmech hustot pro dané $\boldsymbol{p}$ a pro jeho skutečnou hodnotu $\boldsymbol{p}_0$ ve střední hodnotě shora omezen. Vzhledem k tomu, že náš parametrický prostor tvoří $\langle\varepsilon,1-\varepsilon\rangle^K$, víme, že $\|\boldsymbol{p}\|$ bude maximální pro $p_i = 1-\varepsilon$, kdy bude $\|\boldsymbol{p}\| = \sqrt{K}(1-\varepsilon)$. Zvolíme-li jako r číslo o trochu větší než $\sqrt{K}(1-\varepsilon)$, pak víme, že nenajdeme $\boldsymbol{p}$, které by splňovalo $\|\boldsymbol{p}\| > r$, a tedy supremum z $R_k(\boldsymbol{p})$ přes tuto množinu bude $-\infty$. Vzhledem k tomu, že uvažujeme omezenou veličinu $V_i^{(0)}(r)$, víme, že pro $K = 0$ je podmínka C3(ii) splněna.

C4.

(i) Ukážeme dále, že existuje $B > 0$, pro které

$$\limsup_{n\to\infty} \overline{r}_n^{(B)}(\boldsymbol{p}) < 0 \quad \text{pro } \boldsymbol{p} \neq \boldsymbol{p}_0.$$

Připomeňme, že $\overline{r}_n^{(B)}(p)$ je průměrem z $\mathsf{E}\ R_i(\boldsymbol{p})$ pro $i = 1,\ldots,n$. Můžeme tedy využít toho, že $\mathsf{E}\log f_i(Y_i|\boldsymbol{p}) < \mathsf{E}\log f_i(Y_i|\boldsymbol{p}_0)$ pro $\boldsymbol{p} \neq \boldsymbol{p}_0$ (Pawitan, 2001, Věta 9.5, Information inequality). Tedy $\mathsf{E}\ R_i(\boldsymbol{p}) < 0$ pro všechna i, tedy také průměr z těchto čísel musí být záporný.

(ii) Použijeme-li výsledky získané při ověřování podmínky C3(ii), kde jsme zjistili, že je možné najít r takové, že supremum z $R_k(\boldsymbol{p})$ přes $\|\boldsymbol{p}\| > r$ je $-\infty$, dostáváme přímo, že také podmínka C4(ii) je splněna. Hodnoty $v_k^{(B)}(r)$ jsou rovny $-B$, kde $B \geq 0$, tedy záporné a tedy také jejich limita bude záporná, což jsme chtěli ověřit.

C5.

Poslední podmínku klademe na veličiny $R_i(\boldsymbol{p},\rho)$ a $V_i(r)$, které mají být měřitelné funkce Y_i. Díky tomu, že veličina Y_i nabývá pouze dvou hodnot je také podmínka na měřitelnost splněna.

Ověřili jsme všechny podmínky pro konzistenci a tedy $\hat{\boldsymbol{p}}_n \xrightarrow[n\to\infty]{P} \boldsymbol{p}_0$. Přejdeme nyní k podmínkám pro asymptotickou normalitu. Podmínka na konzistenci odhadu (N2) je již ověřena, podívejme se nyní na další podmínky.

Ověření podmínek pro asymptotickou normalitu

N1.

V první podmínce požadujeme, aby skutečný parametr ležel uvnitř parametrického prostoru Θ. Zde tedy můžeme předpokládat $\boldsymbol{p}_0 \in (0{,}1)^K$, čímž je podmínka splněna.

Připomeňme jen, že pro konzistenci jsme uvažovali parametrický prostor $\Theta = \langle \varepsilon, 1-\varepsilon \rangle^K$. Pro $\varepsilon \to 0$ dostáváme $\langle 0{,}1 \rangle^K$, a tedy $\boldsymbol{p}_0 \in \operatorname{int}(\langle 0{,}1 \rangle^K)$ odpovídá $\boldsymbol{p}_0 \in (0{,}1)^K$.

N3.

Ověřme dále existenci prvních a druhých parciálních derivací logaritmů jednotlivých hustot ($\boldsymbol{\Phi}_i'(Y_i,\boldsymbol{p})$ a $\boldsymbol{\Phi}_i''(Y_i,\boldsymbol{p})$). Vezměme tedy pevné k a uvažujme hustotu $f_i(Y_i|\boldsymbol{p}) = p_k^{Y_i}(1-p_k)^{1-Y_i}$ pro $Y_i \in V_k$. Po zlogaritmování dostáváme funkci $Y_i \log p_k + (1-Y_i)\log(1-p_k)$, jejíž první a druhé parciální derivace existují. Vektor prvních parciálních derivací má pouze jeden nenulový prvek a vypadá následovně

$$\boldsymbol{\Phi}_i'(Y_i,\boldsymbol{p}) = \left(0, \dots, 0, \frac{Y_i}{p_k} - \frac{1-Y_i}{1-p_k}, 0, \dots, 0\right)^\top. \tag{1.26}$$

Matice druhých parciálních derivací obsahuje pro dané k pouze jeden nenulový prvek na místě (k,k), a to $-\frac{Y_i}{(p_k)^2} - \frac{1-Y_i}{(1-p_k)^2}$.

N4.

U matice druhých parciálních derivací chceme ověřit spojitost v parametru $\boldsymbol{p}$ (resp. spojitost prvků této matice). Funkce $-\frac{Y_i}{(p_k)^2} - \frac{1-Y_i}{(1-p_k)^2}$ je v p_k spojitá.

N5.

K ověření této podmínky použijeme vztah (1.26). Obdobně jako u předchozí podmínky postupujeme po složkách, stačí tedy pouze ověřit, že

$$\mathsf{E}_{\boldsymbol{p}}\left[\frac{Y_i}{p_k} - \frac{1-Y_i}{1-p_k}\right] = 0,$$

což můžeme provést jednoduchým výpočtem, neboť víme, že $Y_i \in V_k$ pochází z alternativního rozdělení s pravděpodobností úspěchu p_k:

$$\mathsf{E}_{\boldsymbol{p}}\left[\frac{Y_i}{p_k} - \frac{1-Y_i}{1-p_k}\right] = p_k\left(\frac{1}{p_k}\right) + (1-p_k)\left(-\frac{1}{1-p_k}\right) = 0.$$

Podmínka N5 je tedy splněna.

N6.

Další podmínku také ověříme přímým výpočtem. Nejprve podle (1.26) vypočteme $\mathsf{E}_{\boldsymbol{p}}[\boldsymbol{\Phi}_i'(Y_i,\boldsymbol{p})(\boldsymbol{\Phi}_i'(Y_i,\boldsymbol{p}))^\top]$, což je opět matice s pouze jedním nenulovým prvkem na místě (k,k), a to

$$\mathsf{E}_{\boldsymbol{p}}\left[\left(\frac{Y_i}{p_k} - \frac{1-Y_i}{1-p_k}\right)^2\right] = p_k\left(\frac{1}{p_k}\right)^2 + (1-p_k)\left(-\frac{1}{1-p_k}\right)^2 = \frac{1}{p_k} + \frac{1}{1-p_k}.$$

Vypočteme-li nyní střední hodnotu z jediného nenulového prvku matice $\boldsymbol{\Phi}_i''(Y_i,\boldsymbol{p})$ dojdeme ke stejnému výsledku, neboť

$$
\begin{aligned}
-\mathsf{E}_{\boldsymbol{p}}\left[-\frac{Y_i}{{p_k}^2}-\frac{1-Y_i}{(1-p_k)^2}\right] &= -\left[p_k\left(\frac{-1}{(p_k)^2}\right)+(1-p_k)\left(\frac{-1}{(1-p_k)^2}\right)\right]\\
&= \frac{1}{p_k}+\frac{1}{1-p_k},
\end{aligned}
$$

čímž jsme ověřili další z podmínek.

N7.

Z předchozího bodu víme, jak vypadá očekávaná informační matice pro jedno pozorování. Vezmeme-li nyní průměr ze všech těchto matic, dostaneme diagonální matici, která vypadá následovně

$$
\begin{aligned}
\overline{I}_n(\boldsymbol{p}) &= \frac{1}{n}\sum_{k=1}^{n} I_k(\boldsymbol{p})\\
&= \frac{1}{n}\begin{pmatrix}
n_1\left(\frac{1}{p_1}+\frac{1}{1-p_1}\right) & 0 & \dots & \dots & 0\\
0 & \ddots & 0 & & \\
\vdots & 0 & n_k\left(\frac{1}{p_k}+\frac{1}{1-p_k}\right) & 0 & \\
\vdots & \dots & 0 & \ddots & \\
0 & \dots & \dots & 0 & n_K\left(\frac{1}{p_K}+\frac{1}{1-p_K}\right)
\end{pmatrix}.
\end{aligned}
$$

Uvažujeme, že parametr $\boldsymbol{p} \in (0,1)^K$. Pro ukázání konvergence výše uvedené matice k pozitivně definitní matici tak stačí, aby platilo $\frac{n_i}{n} \to c_i$ pro všechna i, kde $0 < c_i < \infty$.

N8.

Pro ověření další podmínky potřebujeme nejprve výraz $\boldsymbol{\lambda}^\top \boldsymbol{\Phi}_i'(Y_i,\boldsymbol{p}_0)$. Z podmínky N3 máme podobu vektoru prvních parciálních derivací, a tak pro $Y_i \in V_k$ dostáváme výraz $\lambda_k \cdot \left(\frac{Y_i}{p_{k0}} - \frac{1-Y_i}{1-p_{k0}}\right)$. Zvolíme nyní $\delta = 1$ a pokračujeme výpočtem střední hodnoty

$$
\begin{aligned}
\mathsf{E}\left[\left|\lambda_k\cdot\left(\frac{Y_i}{p_{k0}}-\frac{1-Y_i}{1-p_{k0}}\right)\right|^3\right] &= \left|\frac{\lambda_k}{p_{k0}}\right|^3 p_{k0} + \left|-\frac{\lambda_k}{1-p_{k0}}\right|^3 (1-p_{k0})\\
&\le d_k,
\end{aligned}
$$

kde $d_k < \infty$ je konstanta pro k-tý výběr. Nyní sečteme střední hodnoty přes všechna pozorování. Podle formulace K-výběrového problému (1.25) se n_1-krát objeví $Y_i \in V_1$, n_2-krát objeví $Y_i \in V_2$, …, n_K-krát objeví $Y_i \in V_K$, celkem pro sumu dostáváme omezení shora výrazem $n_1 d_1 + n_2 d_2 + \ldots + n_K d_K$. Uvážíme-li dále $\max\limits_{k=1,\dots,K} d_k$, kterým nahradíme jednotlivá d_k, dostáváme omezení shora výrazem $n \cdot \max\limits_{k=1,\dots,K} d_k$. Po vydělení $n^{\frac{3}{2}}$ dostáváme výraz $\frac{1}{\sqrt{n}} \max\limits_{k=1,\dots,K} d_k$, který pro $n \to \infty$ konverguje k 0. Tímto je podmínka N8 splněna.

N9.

Splnění poslední podmínky závisí na prvcích matice druhých parciálních derivací $\boldsymbol{\Phi}_i''(Y_i,\boldsymbol{t})$. V tomto případě má tato matice pouze jediný nenulový prvek, a to výraz $-\frac{Y_i}{(t_k)^2} - \frac{1-Y_i}{(1-t_k)^2}$ na místě (k,k), uvažujeme-li $Y_i \in V_k$ a t_k je k-tý prvek vektoru $\boldsymbol{t}$. Chceme tedy ověřit, že existuje $B_{i,kk}(Y_i)$, která je stejnoměrně integrovatelná a splňuje

$$\sup \left\{ \left| -\frac{Y_i}{(t_k)^2} - \frac{1-Y_i}{(1-t_k)^2} \right| : \| \boldsymbol{t} - \boldsymbol{p}_0 \| \leq \varepsilon \right\} \leq B_{i,kk}(Y_i).$$

Výraz, jehož supremum chceme ověřit, závisí pouze na jednom prvku z vektoru parametrů, hledáme tedy supremum přes taková t_k, že $|t_k - p_{k0}| \leq \varepsilon$. Vzhledem k tomu, že je t_k odraženo jak od nuly, tak od jedničky a veličina Y_i nabývá pouze hodnot 0 a 1, můžeme za $B_{i,kk}(Y_i)$ zvolit vhodnou konstantu $c_k < \infty$ a podmínka N9 je splněna.

Ověřili jsme tedy všechny podmínky potřebné pro asymptotickou normalitu $\hat{\boldsymbol{p}}_n$. Pokud tedy platí, že $\frac{n_i}{n} \to c_i$ pro všechna i, kde $0 < c_i < \infty$ (požadavek, který potřebujeme ke splnění podmínky N7), dostáváme podle věty 6

$$\sqrt{n}(\hat{\boldsymbol{p}}_n - \boldsymbol{p}_0) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}_K(\boldsymbol{0_K}, \overline{I}^{-1}(\boldsymbol{p}_0)),$$

kde $\overline{I}^{-1}(\boldsymbol{p}_0))$ je inverzní matice k matici, ke které konverguje $\overline{I}_n(\boldsymbol{p})$ z podmínky N7 pro $n \to \infty$.

2. Asymptotické vlastnosti odhadů získaných minimalizací konvexních funkcí

V předchozí kapitole jsme uvedli podmínky, za kterých máme zaručenu konzistenci a asymptotickou normalitu maximálně věrohodných odhadů v případě nezávislých, ale ne stejně rozdělených pozorování. Nyní využijeme asymptotických vlastností odhadů, které jsou definovány pomocí minimalizace konvexních funkcí. Tyto výsledky poté použijeme pro metodu maximální věrohodnosti, čímž dosáhneme slabších a jednodušeji ověřitelných předpokladů.

Tato kapitola se opírá o stejnojmenný článek (Hjort a Pollard, 2011), ve kterém najdeme zde uváděné věty a jejich důkazy, případně náznaky důkazů. Také příklady logistické a poissonovské regrese v další sekci pocházejí z tohoto článku a jsou zde podrobněji rozpracovány.

2.1 Úvod

Mějme posloupnost nezávislých náhodných veličin $Y_1, \dots, Y_n$ a uvažujme parametr $\boldsymbol{\theta} \in \mathbb{R}^p$, jehož odhad získáme minimalizací výrazu $\sum_{i=1}^n g_i(Y_i, \boldsymbol{\theta})$, kde $g_i(y_i, \boldsymbol{\theta})$ jsou funkce konvexní v $\boldsymbol{\theta}$. Právě konvexita nám umožní použít následující větu, která zajistí konzistenci i asymptotickou normalitu pro náš odhad současně.

Věta 7. *(Hjort a Pollard, 2011, Basic Corollary, str.2) Nechť $A_n(\mathbf{s})$ pro $s \in S$ je posloupnost konvexních náhodných funkcí, kde $S \subset \mathbb{R}^p$ je otevřená a konvexní množina. Předpokládejme dále, že $A_n(\mathbf{s})$ lze reprezentovat jako $\frac{1}{2}\mathbf{s}^\top V \mathbf{s} + \boldsymbol{U_n}^\top \boldsymbol{s} + C_n + r_n(\mathbf{s})$, kde V je symetrická pozitivně definitní matice, $\boldsymbol{U_n}$ je posloupnost vektorů omezená v pravděpodobnosti, C_n je libovolná a $r_n(\mathbf{s}) \xrightarrow[n\to\infty]{P} 0$ pro každé $\mathbf{s}$. Nechť $\boldsymbol{\alpha}_n = \operatorname*{argmin}_{s \in S} A_n(\boldsymbol{s})$, potom $\boldsymbol{\alpha}_n = \boldsymbol{\beta}_n + o_p(1)$, kde $\boldsymbol{\beta}_n = \operatorname*{argmin}_{s\in S} \frac{1}{2}\mathbf{s}^\top V \mathbf{s} + \boldsymbol{U_n}^\top \boldsymbol{s} + C_n = -V^{-1}\boldsymbol{U_n}$. Pokud také $\boldsymbol{U_n} \xrightarrow[n\to\infty]{\mathcal{D}} \boldsymbol{U}$, potom $\boldsymbol{\alpha}_n \xrightarrow[n\to\infty]{\mathcal{D}} -V^{-1}\boldsymbol{U}$.*

Výše uvedená věta umožňuje převést hledání argumentu minima z nějaké konvexní funkce na hledání argumentu minima kvadratické aproximace této funkce. Uvažujeme takovou aproximaci, že rozdíl původní funkce a její aproximace je zanedbatelný v pravděpodobnosti. Zvolíme-li tedy jako aproximaci původní funkce funkci, jejíž argument minima známe, je naše úloha vyřešena. Při splnění několika předpokladů totiž věta říká, že pak rozdíl argumentů minima těchto funkcí je zanedbatelný v pravděpodobnosti.

Důkaz věty je uveden v Hjort a Pollard (2011), uveďme zde alespoň základní myšlenky, na kterých je postaven.

Důkaz vychází ze dvou poznatků pro konvexní funkce. Prvním z nich je přechod od bodové konvergence konvexních náhodných funkcí ke stejnoměrné konvergenci na kompaktních množinách (obě ve smyslu konvergence v pravděpodobnosti). Víme-li tedy, že funkce $A_n(\boldsymbol{s})$ z věty 7 konverguje v pravděpodobnosti k funkci $\frac{1}{2}\mathbf{s}^\top V \mathbf{s} + \boldsymbol{U_n}^\top \boldsymbol{s} + C_n$ pro všechna $\boldsymbol{s} \in \mathbb{R}^p$, neboť $r_n(\mathbf{s}) \xrightarrow[n\to\infty]{P} 0$, víme tedy, že na kompaktních podmnožinách $\mathbb{R}^p$ se jedná o stejnoměrnou konvergenci (Hjort a Pollard, 2011, Lemma 1, str.2). Druhý poznatek nám dává představu o tom, jak mohou být argumenty minima dvou funkcí, kde jedna je aproximací druhé, od sebe nejvýše vzdáleny (Hjort a Pollard, 2011, Lemma 2, str.2). Dostáváme hranici pro pravděpodobnost, se kterou vzdálenost těchto argumentů minim překročí předem stanovený rozdíl.

Spojením těchto dvou poznatků dostáváme požadovanou větu. Víme, že rozdíl aproximované a aproximující funkce konverguje stejnoměrně k nule, a tedy dostáváme, že rozdíl mezi argumenty minim je v pravděpodobnosti zanedbatelný.

Vzhledem k tomu, že nepřistupujeme k hledání argumentu minima pomocí derivování minimalizované funkce a následně hledání kořene tohoto výrazu, nepotřebujeme klást na hustoty pozorování některé z předpokladů, které jsme potřebovali v předchozí kapitole, v obecném případě.

Je vhodné na tomto místě poznamenat, že získáváme-li odhad parametru $\boldsymbol{\theta}$ na základě minimalizace $\sum_{i=1}^n g_i(Y_i,\boldsymbol{\theta})$ jedná se o M-odhad, který lze použít pro široké spektrum ztrátových funkcí g. V této práci se ovšem zaměříme na aplikaci tohoto způsobu odhadování na metodu maximální věrohodnosti.

2.2 Konzistence a asymptotická normalita

Označme opět skutečný parametr jako $\boldsymbol{\theta}_0$ a jeho odhad $\hat{\boldsymbol{\theta}}_n$. Předpokládejme, že pro funkci $g_i(y_i,\boldsymbol{\theta})$, která je konvexní v $\boldsymbol{\theta}$, existuje rozvoj okolo skutečného parametru $\boldsymbol{\theta}_0$ takový, že pro $\boldsymbol{t} \in \mathbb{R}^p$

$$g_i(y_i,\boldsymbol{\theta}_0 + \mathbf{t}) - g_i(y_i,\boldsymbol{\theta}_0) = \boldsymbol{D}_i(y_i,\boldsymbol{\theta}_0)^\top \mathbf{t} + R_i(y_i,\mathbf{t}), \tag{2.1}$$

kde $\mathsf{E}\ \boldsymbol{D}_i(y_i,\boldsymbol{\theta}_0) = \mathbf{0}_p$ a dále

$$\mathsf{E}\ R_i(Y_i,\mathbf{t}) = \tfrac{1}{2}\mathbf{t}^\top A_i \mathbf{t} + v_{i,0}(\mathbf{t}), \tag{2.2}$$

$$\mathrm{var} R_i(Y_i,\mathbf{t}) = v_i(\mathbf{t}). \tag{2.3}$$

Označme B_i varianční matici pro $\boldsymbol{D}_i(y_i,\boldsymbol{\theta}_0)$ a dále $J_n = \sum_{i=1}^n A_i$ a $K_n = \sum_{i=1}^n B_i$.

Pro důkaz asymptotické normality budeme využívat Feller-Lindebergovu centrální limitní větu v její úpravě pro lineární kombinace náhodných vektorů (uvedena v příloze jako věta A.2). Budeme tedy uvažovat tzv. Lindebergovu podmínku v následující podobě

$$\forall \boldsymbol{\lambda} \in \mathbb{R}^p\ \forall \varepsilon > 0 : \lim_{n\to\infty} \sum_{i=1}^n \mathsf{E}\left[\frac{1}{n}(\boldsymbol{\lambda}^\top \boldsymbol{D}_i(Y_i,\boldsymbol{\theta}_0))^2 \mathbb{I}\left\{\left|\frac{1}{\sqrt{n}}\boldsymbol{\lambda}^\top \boldsymbol{D}_i(Y_i,\boldsymbol{\theta}_0)\right| \geq \varepsilon\right\}\right] = 0. \tag{2.4}$$

Věta 8. *(Hjort a Pollard, 2011, Věta 2.2) Nechť platí $\sum_{i=1}^n v_{i,0}(\frac{\mathbf{s}}{\sqrt{n}}) \xrightarrow[n\to\infty]{} 0$ a $\sum_{i=1}^n v_i(\frac{\mathbf{s}}{\sqrt{n}}) \xrightarrow[n\to\infty]{} 0$ pro každé $\boldsymbol{s} \in \mathbb{R}^p$ a nechť dále $\frac{J_n}{n} \xrightarrow[n\to\infty]{} J$ a $\frac{K_n}{n} \xrightarrow[n\to\infty]{} K$, kde J je pozitivně definitní. Potom*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = -J^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{D}_i(Y_i,\boldsymbol{\theta}_0) + o_p(1).$$

a odhad $\hat{\boldsymbol{\theta}}_n$ je $\sqrt{n}$-konzistentní odhad skutečného parametru $\boldsymbol{\theta}_0$. Jestliže jsou splněny Lindebergovy podmínky pro posloupnost $\frac{1}{\sqrt{n}}\boldsymbol{D}_i(Y_i,\boldsymbol{\theta}_0)$, viz (2.4), potom také

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}_p(\mathbf{0}_p, J^{-1}KJ^{-1}).$$

Poznámka. Výrazy používané ve větě 8 jsme definovali dříve, viz (2.2) a (2.3).

Důkaz. Uvažujme funkci

$$G_n(\mathbf{s}) = \sum_{i=1}^{n} \left(g_i \left(Y_i, \boldsymbol{\theta}_0 + \frac{\mathbf{s}}{\sqrt{n}} \right) - g_i(Y_i, \boldsymbol{\theta}_0) \right) \quad \text{pro } \mathbf{s} \in \mathbb{R}^p.$$

Můžeme snadno nahlédnout, že se jedná o konvexní funkci v $\mathbf{s}$, neboť výraz $\sum_{i=1}^{n} g_i \left(Y_i, \boldsymbol{\theta}_0 + \frac{\mathbf{s}}{\sqrt{n}} \right)$ je součet konvexních funkcí a výraz, který od tohoto součtu odečítáme ($\sum_{i=1}^{n} g_i(Y_i, \boldsymbol{\theta}_0)$) nezávisí na proměnné $\mathbf{s}$. Funkce $G_n(\mathbf{s})$ nabývá minima v bodě $\mathbf{s} = \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$, jelikož $\hat{\boldsymbol{\theta}}_n$ získáme jako minimum ze $\sum_{i=1}^{n} g_i(Y_i, \boldsymbol{\theta})$. Využijeme nyní vztahu 2.1 pro všechna pozorování a dostáváme

$$\begin{aligned} G_n(\mathbf{s}) &= \sum_{i=1}^{n} \boldsymbol{D}_i(Y_i, \boldsymbol{\theta}_0)^\top \left(\frac{\mathbf{s}}{\sqrt{n}} \right) + R_i \left(Y_i, \frac{\mathbf{s}}{\sqrt{n}} \right) \\ &= \sum_{i=1}^{n} \left(\boldsymbol{D}_i(Y_i, \boldsymbol{\theta}_0)^\top \left(\frac{\mathbf{s}}{\sqrt{n}} \right) + R_i \left(Y_i, \frac{\mathbf{s}}{\sqrt{n}} \right) - \mathsf{E}\, R_i \left(Y_i, \frac{\mathbf{s}}{\sqrt{n}} \right) \right) \\ &\qquad + \sum_{i=1}^{n} \mathsf{E}\, R_i \left(Y_i, \frac{\mathbf{s}}{\sqrt{n}} \right). \end{aligned} \tag{2.5}$$

Označme nyní $\boldsymbol{U_n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{D}_i(Y_i, \boldsymbol{\theta}_0)$ a dále

$$r_n(\boldsymbol{s}) = \sum_{i=1}^{n} \left(R_i \left(Y_i, \frac{\mathbf{s}}{\sqrt{n}} \right) - \mathsf{E}\, R_i \left(Y_i, \frac{\mathbf{s}}{\sqrt{n}} \right) \right).$$

Podle vztahu (2.2) můžeme dále rozepsat

$$\begin{aligned} \sum_{i=1}^{n} \mathsf{E}\, R_i \left(Y_i, \frac{\mathbf{s}}{\sqrt{n}} \right) &= \sum_{i=1}^{n} \left(\frac{1}{2} \left(\frac{\mathbf{s}}{\sqrt{n}} \right)^\top A_i \left(\frac{\mathbf{s}}{\sqrt{n}} \right) + v_{i,0} \left(\frac{\mathbf{s}}{\sqrt{n}} \right) \right) \\ &= \frac{1}{2} \mathbf{s}^\top \frac{J_n}{n} \mathbf{s} + \sum_{i=1}^{n} v_{i,0} \left(\frac{\mathbf{s}}{\sqrt{n}} \right), \end{aligned}$$

neboť $\sum_{i=1}^{n} A_i = J_n$. Využijeme-li výše uvedených úprav, můžeme funkci $G_n(\boldsymbol{s})$ z (2.5) nyní zapsat jako

$$G_n(\mathbf{s}) = \boldsymbol{U_n}^\top \mathbf{s} + \frac{1}{2} \mathbf{s}^\top \frac{J_n}{n} \mathbf{s} + \sum_{i=1}^{n} v_{i,0} \left(\frac{\mathbf{s}}{\sqrt{n}} \right) + r_n(\mathbf{s}),$$

což je reprezentace požadovaná ve větě 7. Ověříme tedy předpoklady této věty.

Jak už jsme ukázali dříve, funkce $G_n(\mathbf{s})$ je konvexní, svého minima nabývá pro $\mathbf{s} = \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ a můžeme ji aproximovat funkcí $\frac{1}{2}\mathbf{s}^\top J \mathbf{s} + \boldsymbol{U_n}^\top \mathbf{s}$, neboť platí, že $\sum_{i=1}^{n} v_{i,0} \left(\frac{\mathbf{s}}{\sqrt{n}} \right) \xrightarrow[n\to\infty]{} 0$, $\frac{J_n}{n} \xrightarrow[n\to\infty]{} J$ a vzápětí také ukážeme, že $r_n(\mathbf{s}) \xrightarrow[n\to\infty]{P} 0$. Matice J je podle předpokladu pozitivně definitní a také symetrická, protože ji získáme jako střední hodnotu ze součtu matic druhých parciálních derivací funkcí g_i. Tyto funkce jsou v $\boldsymbol{\theta}$ konvexní, a tedy i spojité, což zajišťuje, že matice druhých parciálních derivací je symetrická.

Dále potřebujeme ukázat, že posloupnost vektorů $\boldsymbol{U_n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{D}_i(Y_i,\boldsymbol{\theta}_0)$ je omezená v pravděpodobnosti. To plyne z toho, že víme $\mathsf{E}\, \boldsymbol{D}_i(Y_i,\boldsymbol{\theta}_0) = 0$ a také $\operatorname{var} \boldsymbol{D}_i(Y_i,\boldsymbol{\theta}_0) = B_i$ a $\sum_{i=1}^{n} B_i = K_n$ a $\frac{K_n}{n} \xrightarrow[n\to\infty]{} K$, tedy

$$\operatorname{var} \boldsymbol{U_n} = \tfrac{1}{n} \sum_{i=1}^{n} \operatorname{var} \boldsymbol{D}_i(Y_i,\boldsymbol{\theta}_0) = \tfrac{1}{n} \sum_{i=1}^{n} B_i = \tfrac{K_n}{n} \xrightarrow[n\to\infty]{} K.$$

Posledním předpokladem, který musíme ověřit je $r_n(\mathbf{s}) \xrightarrow[n\to\infty]{P} 0$ pro všechna $\mathbf{s} \in \mathbb{R}^p$. To ukážeme snadno, neboť posloupnost $r_n(\mathbf{s})$ má nulovou střední hodnotu a pro její rozptyl platí

$$\operatorname{var}\Big(r_n(\mathbf{s})\Big) = \operatorname{var} \sum_{i=1}^{n} R_i\left(Y_i,\frac{\mathbf{s}}{\sqrt{n}}\right) = \sum_{i=1}^{n} \operatorname{var}\left(R_i\left(Y_i,\frac{\mathbf{s}}{\sqrt{n}}\right)\right) = \sum_{i=1}^{n} v_i\left(\frac{\mathbf{s}}{\sqrt{n}}\right),$$

a dále $\sum_{i=1}^{n} v_i\left(\frac{\mathbf{s}}{\sqrt{n}}\right) \xrightarrow[n\to\infty]{} 0$ podle předpokladu věty. Z věty 7 tedy plyne, že

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = \boldsymbol{\beta}_n + o_p(1), \tag{2.6}$$

kde $\boldsymbol{\beta}_n$ je argument minimalizující funkci, kterou funkci $G_n(\mathbf{s})$ aproximujeme, což je funkce $\frac{1}{2}\mathbf{s}^\top J\mathbf{s}+\boldsymbol{U_n}^\top\mathbf{s}$. Tato funkce nabývá minimální hodnoty pro $\boldsymbol{s} = -J^{-1}\boldsymbol{U_n}$. Po dosazení tak z (2.6) dostáváme

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = -J^{-1}\tfrac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{D}_i(Y_i,\boldsymbol{\theta}_0) + o_p(1), \tag{2.7}$$

čímž je důkaz první části věty 8 dokončen.

Ukázali jsme zároveň, že odhad $\hat{\boldsymbol{\theta}}_n$ je $\sqrt{n}$-konzistentní. Posloupnost náhodných vektorů $\boldsymbol{U_n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{D}_i(Y_i,\boldsymbol{\theta}_0)$ je omezená v pravděpodobnosti a po vynásobení pevnou maticí J^{-1} a přičtení členu, který je řádu $o_p(1)$ (pravá strana (2.7)), tak stále máme posloupnost omezenou v pravděpodobnosti, což musí platit i pro posloupnost na levé straně (2.7). Máme tedy zaručenu konzistenci $\hat{\boldsymbol{\theta}}_n$.

Dále chceme ukázat asymptotickou normalitu. Předpokládáme, že posloupnost $\frac{1}{\sqrt{n}}\boldsymbol{D}_i(Y_i,\boldsymbol{\theta}_0)$ splňuje Lindebergovu podmínku. Z toho podle Feller-Lindebergovy věty (věta A.2) plyne, že $\boldsymbol{U_n} = \sum_{i=1}^{n} \frac{1}{\sqrt{n}}\boldsymbol{D}_i(Y_i,\boldsymbol{\theta}_0)$ konverguje k $\mathcal{N}_p(\mathbf{0}_p,K)$, neboť $\lim_{n\to\infty} \sum_{i=1}^{n} \operatorname{var}(\frac{1}{\sqrt{n}}\boldsymbol{D}_i(Y_i,\boldsymbol{\theta}_0)) = K$. Odtud tedy snadno dostáváme

$$-J^{-1}\boldsymbol{U_n} \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}_p(\mathbf{0}_p,J^{-1}KJ^{-1}),$$

čímž je důkaz dokončen.

□

2.3 Metoda maximální věrohodnosti pro logkonkávní hustoty

Chceme-li větu 8 použít v případě odhadu metodou maximální věrohodnosti, stačí si pouze uvědomit, že maximálně věrohodný odhad $\hat{\boldsymbol{\theta}}_n$ je takový, který splňuje následující nerovnost

$$L_n(\hat{\boldsymbol{\theta}}_n) \geq L_n(\boldsymbol{\theta}) \quad \text{pro všechna } \boldsymbol{\theta} \in \Theta,$$

tedy můžeme psát

$$\hat{\boldsymbol{\theta}}_n = \operatorname*{argmax}_{\boldsymbol{\theta}\in\Theta} L_n(\boldsymbol{\theta}) = \operatorname*{argmax}_{\boldsymbol{\theta}\in\Theta} \prod_{i=1}^{n} f_i(Y_i|\boldsymbol{\theta}).$$

Přejdeme-li k maximalizaci logaritmické věrohodnosti a následně k minimalizační úloze, dostáváme

$$\hat{\boldsymbol{\theta}}_n = \operatorname*{argmax}_{\boldsymbol{\theta}\in\Theta} \sum_{i=1}^{n} \log f_i(Y_i|\boldsymbol{\theta}) = \operatorname*{argmin}_{\boldsymbol{\theta}\in\Theta} \sum_{i=1}^{n} -\log f_i(Y_i|\boldsymbol{\theta}).$$

Označíme-li $g_i(y_i,\boldsymbol{\theta}) = -\log f_i(y_i|\boldsymbol{\theta})$, kde $\log f_i(y_i|\boldsymbol{\theta})$ je konkávní funkce v $\boldsymbol{\theta}$, můžeme na funkce $g_i(y_i,\boldsymbol{\theta})$ aplikovat větu 8, ze které dostáváme

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = J^{-1} \sum_{i=1}^{n} \boldsymbol{D}_i(Y_i,\boldsymbol{\theta}_0) + o_P(1).$$

Jestliže je navíc splněna Lindebergova podmínka pro posloupnost $\frac{1}{\sqrt{n}}\boldsymbol{D}_i(Y_i,\boldsymbol{\theta}_0)$ a máme navíc, že $K = J$ dostáváme

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}_p(\mathbf{0}_p,J^{-1}).$$

Při využití Taylorova rozvoje pro získání rozvoje (2.1), potřebujeme pro rovnost matic K a J rovnost $K_n = J_n$, tedy

$$\sum_{i=1}^{n} \operatorname{var}\left(-\frac{\partial \log f_i(Y_i|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right) = \sum_{i=1}^{n} \mathsf{E}\, \frac{\partial^2 \log f_i(Y_i|\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}. \tag{2.8}$$

Vzhledem k tomu, že výraz, jehož rozptyl na levé straně (2.8) počítáme, má nulovou střední hodnotu, odpovídá rovnost (2.8) podmínce N6 z obecného případu (část 1.3) a také podmínce R2(ii), kterou potřebujeme v případě stejně rozdělených pozorování (část 1.4).

2.4 Uvedení na příkladech

V této části ukážeme teoretické výsledky pro odhady získané minimalizací konvexních funkcí na konkrétních modelech. Zaměříme se na tři regresní modely, a to na lineární, logistický a poissonovský regresní model a odhad regresních koeficientů v těchto modelech. U každého z příkladů nejprve popíšeme model, se kterým budeme pracovat a poté si připravíme jeho reprezentaci (jako v sekci 2.2), kterou budeme potřebovat k následnému ověření předpokladů věty 8, kterou v těchto případech použijeme k důkazu konzistence a asymptotické normality odhadů regresních koeficientů.

2.4.1 Normální lineární model

Uvažujme nyní normální lineární model s pevnými regresory (tzn. nenáhodnými), tedy

$$Y_i \sim \mathcal{N}(\boldsymbol{x}_i^\top \boldsymbol{\beta},\sigma^2) \quad \text{pro } i = 1,\dots,n, \tag{2.9}$$

kde Y_i jsou nezávislé, $\boldsymbol{x}_i \in \mathbb{R}^p$ je vektor regresorů, $\boldsymbol{\beta} \in \mathbb{R}^p$ značí vektor regresních koeficientů, $0 < \sigma^2 < \infty$ a označme jako $\boldsymbol{\beta}_0$ skutečný vektor regresních koeficientů. Předpokládáme tedy, že $Y_1, \ldots, Y_n$ pocházejí z rozdělení s hustotou

$$f_i(y_i|\boldsymbol{\beta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta})^2}{2\sigma^2}\right).$$

Naším cílem je ukázat vlastnosti odhadu parametru $\boldsymbol{\beta}$. Na tomto místě je vhodné poznamenat, že pro normální lineární model máme explicitní vzorec pro $\hat{\boldsymbol{\beta}}$ a víme, že platí

$$\hat{\boldsymbol{\beta}} \sim \mathcal{N}_p(\mathbf{0}_p, \sigma^2 (X^\top X)^{-1}),$$

kde X je matice modelu (2.9) o rozměrech $n \times p$, která má v řádcích vektory regresorů pro jednotlivá pozorování a má hodnost p. Vzhledem k tomu, že předpokládáme normalitu Y_i, máme přesné rozdělení pro $\hat{\boldsymbol{\beta}}$. Použití věty 8 k ukázání konzistence a asymptotické normality v tomto případě tedy není vhodným řešením, ale ukazuje postup při aplikaci věty 8.

Budeme se soustředit na vlastnosti odhadu regresních koeficientů $\boldsymbol{\beta}$, poznamenejme však také pár slov k rozptylu. V tomto příkladě uvažujeme shodný rozptyl σ^2 pro všechna pozorování. Náš výsledek na tom, zda rozptyl σ^2 známe, či neznáme, nezávisí. V případě, že je rozptyl známý představuje vektor regresních koeficientů $\boldsymbol{\beta}$ všechny neznámé parametry v modelu. Je-li rozptyl neznámý a je tedy nutné jej odhadnout, používáme k tomu statistiku založenou na reziduálním součtu čtverců lineárního modelu, která je s odhadem regresních koeficientů nezávislá.

Uvažujme nyní funkci $g_i(y_i,\boldsymbol{\beta}) = -\log f_i(y_i|\boldsymbol{\beta})$, tedy

$$g_i(y_i,\boldsymbol{\beta}) = \log\sqrt{2\pi} + \tfrac{1}{2}\log(\sigma^2) + \frac{(y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta})^2}{2\sigma^2}.$$

Pro získání reprezentace požadované pro použití věty 8 rozvineme funkci $g_i(y_i,\boldsymbol{\beta})$ pomocí Taylorova rozvoje druhého řádu okolo bodu $\boldsymbol{\beta}_0$. Vidíme, že se jedná o kvadratickou funkci, získáme tedy její přesné vyjádření. Je také zřejmé, že se jedná o funkci konvexní v $\boldsymbol{\beta}$, což se potvrdí i následujícím výpočtem.

V reprezentaci (2.1) dostáváme v tomto případě

$$\boldsymbol{D}_i(Y_i,\boldsymbol{\beta}_0) = \frac{(Y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}_0)}{\sigma^2}(-\boldsymbol{x}_i), \tag{2.10}$$

$$R_i(Y_i,\mathbf{t}) = \frac{1}{2}\boldsymbol{t}^\top \tilde{A}_i(Y_i)\boldsymbol{t}, \tag{2.11}$$

kde $\tilde{A}_i(Y_i) = \frac{\partial^2 g_i}{\partial\beta\partial\beta^\top}$ a vypadá následovně

$$\tilde{A}_i(Y_i) = \frac{\boldsymbol{x}_i \boldsymbol{x}_i^\top}{\sigma^2}.$$

Vidíme, že matice $\tilde{A}_i(Y_i)$, která je maticí druhých parciálních derivací funkce $g_i(y_i,\boldsymbol{\beta})$, je pozitivně semidefinitní, což potvrzuje, že funkce $g_i(y_i,\boldsymbol{\beta})$ je v $\boldsymbol{\beta}$ konvexní.

Ověření předpokladů věty 8.

Pro připomenutí, požadujeme, aby $\frac{J_n}{n}$ konvergovala k pozitivně definitní matici, $\frac{K_n}{n}$ byla konečná a také, aby $\sum_{i=1}^n v_{i,0}(\mathbf{s}/\sqrt{n}) \xrightarrow[n\to\infty]{} 0$ a $\sum_{i=1}^n v_i(\mathbf{s}/\sqrt{n}) \xrightarrow[n\to\infty]{} 0$ pro všechna $\boldsymbol{s} \in \mathbb{R}^p$.

Vezměme nejdříve první podmínku. Matice J_n je definována jako součet matic A_i, kde matice A_i je střední hodnotou z matice $\tilde{A}_i$. V tomto případě ale matice $\tilde{A}_i$ neobsahuje žádný náhodný člen, a tedy $A_i = \tilde{A}_i$ a můžeme rovnou přejít k výpočtu

$$\frac{J_n}{n} = \frac{1}{n}\sum_{i=1}^n \frac{\boldsymbol{x}_i\boldsymbol{x}_i^\top}{\sigma^2}.$$

Stačí předpokládat, že $\frac{1}{n}\sum_{i=1}^n \frac{\boldsymbol{x}_i\boldsymbol{x}_i^\top}{\sigma^2}$ konverguje k pozitivně definitní matici.

Podívejme se dále, jak vypadá matice $\frac{K_n}{n}$. K_n získáme jako součet variančních matic pro $\boldsymbol{D}_i(Y_i,\boldsymbol{\beta}_0)$ pro $i = 1,\ldots,n$. Jelikož náhodný vektor $\boldsymbol{D}_i(Y_i)$ má nulovou střední hodnotu, můžeme psát

$$B_i = \mathsf{E}\ \boldsymbol{D}_i(Y_i,\boldsymbol{\beta}_0)\boldsymbol{D}_i(Y_i,\boldsymbol{\beta}_0)^\top = \frac{\boldsymbol{x}_i\boldsymbol{x}_i^\top}{\sigma^2}.$$

Sečtením matic B_i přes $i = 1,\ldots,n$ dostáváme matici $K_n = \frac{X^\top X}{\sigma^2}$, která se v tomto případě rovná matici J_n. Předpokládejme, že

$$\frac{1}{n}\sum_{i=1}^n \frac{\boldsymbol{x}_i\boldsymbol{x}_i^\top}{\sigma^2} \xrightarrow[n\to\infty]{} K. \tag{2.12}$$

Podle (2.11) vidíme, že $v_{i,0}(\mathbf{t})$ je přímo rovno nule, neboť jsme provedli kvadratickou „aproximaci“ kvadratické funkce, tedy nemáme žádný zbytkový člen. Další předpoklad je tedy splněn.

Nakonec zbývá ověřit, že $\sum_{i=1}^n v_i(\mathbf{s}/\sqrt{n}) \xrightarrow[n\to\infty]{} 0$, kde $v_i(\mathbf{s}/\sqrt{n})$ je definováno vztahem (2.3) jako rozptyl $R_i(Y_i,\mathbf{t})$. Jelikož tento již neobsahuje žádnou náhodnou veličinu, je rozptyl roven nule, a tedy i tato podmínka je splněna.

Předpokládáme-li, že platí (2.12), jsou splněny všechny předpoklady věty 8 a dostáváme, že

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = \left(\frac{1}{n}\sum_{i=1}^n \frac{\boldsymbol{x}_i\boldsymbol{x}_i^\top}{\sigma^2}\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\frac{(Y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta}_0)}{\sigma^2}\boldsymbol{x}_i\right) + o_p(1).$$

Asymptotická normalita

Pro zajištění asymptotické normality odhadu $\boldsymbol{\beta}$ potřebujeme ověřit Lindebergovu podmínku (viz (2.4)) pro posloupnost $\frac{1}{\sqrt{n}}\frac{(Y_i-\boldsymbol{x}_i^\top\boldsymbol{\beta}_0)}{\sigma^2}(-\boldsymbol{x}_i)$. V tomto případě je ale vhodnější přejít k ověření Ljapunovovy podmínky (Billingsley, 2008, str.362), ze které platnost Lindebergovy podmínky vyplývá (Billingsley, 2008, str.362).

Ljapunovova podmínka má tento tvar:

$$\forall \boldsymbol{\lambda} \in \mathbb{R}^p\ \exists \delta > 0 : \lim_{n\to\infty}\sum_{i=1}^n \mathsf{E}\left[\left|\frac{1}{\sqrt{n}}\frac{(Y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta}_0)}{\sigma^2}\boldsymbol{\lambda}^\top\boldsymbol{x}_i\right|^{2+\delta}\right] = 0.$$

Označme nyní

$$L_n(\delta) := \sum_{i=1}^n \mathsf{E}\,[|\frac{1}{\sqrt{n}}\frac{(Y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta}_0)}{\sigma^2}\boldsymbol{\lambda}^\top\boldsymbol{x}_i|^{2+\delta}].$$

Chceme ukázat, že $L_n(\delta)$ můžeme shora omezit výrazem, který konverguje k 0, pak bude Ljapunovova podmínka splněna.

$$\begin{aligned}
L_n(\delta) &= \sum_{i=1}^{n} \mathsf{E}\left[\left|\frac{1}{\sqrt{n}}\frac{(Y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}_0)}{\sigma^2}\right|^{2+\delta}\right] (\boldsymbol{\lambda}^\top \boldsymbol{x}_i)^2 \left|\boldsymbol{\lambda}^\top \boldsymbol{x}_i\right|^{\delta} \\
&\leq \frac{1}{n^{\delta/2}} \|\boldsymbol{\lambda}\|^{\delta} \max_{i=1,\dots,n} \|\boldsymbol{x}_i\|^{\delta} \boldsymbol{\lambda}^\top \frac{1}{n} \sum_{i=1}^{n} \mathsf{E}\left[\left|\frac{(Y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}_0)}{\sigma^2}\right|^{2+\delta}\right] \boldsymbol{x}_i \boldsymbol{x}_i^\top \boldsymbol{\lambda},
\end{aligned}$$

kde jsme využili Cauchyho-Schwartzovy nerovnosti pro $|\boldsymbol{\lambda}^\top \boldsymbol{x}_i|^{\delta}$ a následně $\|\boldsymbol{x}_i\|$ omezili shora výrazem $\max\limits_{i=1,\dots,n} \|\boldsymbol{x}_i\|$, abychom dostali výraz, který nezávisí na indexu i. Konvergenci tohoto výrazu k 0 pak můžeme zajistit tím, že předpokládáme

$$\frac{1}{n} \sum_{i=1}^{n} \mathsf{E}\left[\left|\frac{(Y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}_0)}{\sigma^2}\right|^{2+\delta}\right] \boldsymbol{x}_i \boldsymbol{x}_i^\top \xrightarrow[n\to\infty]{} C, \text{ kde } C \text{ je konečná,} \tag{2.13}$$

$$\frac{1}{\sqrt{n}} \max_{i=1,\dots,n} \|\boldsymbol{x}_i\| \xrightarrow[n\to\infty]{} 0. \tag{2.14}$$

Závěr

Při splnění podmínek (2.13) a (2.14) dostáváme konzistenci a asymptotickou normalitu pro odhad vektoru regresních koeficientů $\boldsymbol{\beta}$.

Srovnání s modelem s náhodnými regresory

Uvažujeme-li lineární model s náhodnými regresory, pak předpokládáme, že pozorujeme nezávislé a stejně rozdělené náhodné vektory $(Y_i, \boldsymbol{X}_i)$ pro $i = 1, \dots, n$, pro jejichž podmíněnou střední hodnotu platí $\mathsf{E}[Y_i|\boldsymbol{X}_i] = \boldsymbol{\beta}^\top \boldsymbol{X}_i$, kde $\boldsymbol{\beta} \in \mathbb{R}^p$ je vektor regresních koeficientů a $\mathrm{var}[Y_i|\boldsymbol{X}_i] = \sigma^2$ (označme tyto tři podmínky jako P1).

Předpokládáme-li normalitu odezev Y_i, tedy $Y_i|\boldsymbol{X}_i \sim \mathcal{N}(0,\sigma^2)$, máme zaručeno také podmíněné normální rozdělení pro $\hat{\boldsymbol{\beta}}_n$, obdobně jako v případě s pevnými regresory. Jestliže nepředpokládáme normální rozdělení pro odezvu, můžeme stanovit podmínky, které zaručí silnou konzistenci a asymptotickou normalitu $\hat{\boldsymbol{\beta}}_n$ (viz Komárek (2016)). Jsou to následující podmínky (označme je P2)

(i) $\mathsf{E}\,|X_{1,j}X_{1,l}| < \infty \quad j,l = 1, \dots, p,$

(ii) $\mathsf{E}\,(\boldsymbol{X}_1 \boldsymbol{X}_1^\top) = W,$ kde W je pozitivně definitní matice,

kde $\boldsymbol{X}_1$ je náhodný vektor regresorů. Z podmínek P1 a P2 poté s použitím silného zákona velkých čísel plyne, že $\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i^\top \xrightarrow[n\to\infty]{s.j.} W$ a také platí

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_n &\xrightarrow[n\to\infty]{s.j.} \boldsymbol{\beta}_0 \\
\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) &\xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}_p(\boldsymbol{0}_p, W^{-1}).
\end{aligned}$$

Vidíme, že podmínky P2 (i) a (ii) jsou obdobou podmínek (2.13) a (2.14), které jsme potřebovali pro ukázání konzistence a asymptotické normality v případě normálního lineárního modelu s pevnými regresory.

2.4.2 Logistická regrese

Uvažujme jako další model logistickou regresi s pevnými regresory. Pozorujeme tedy nezávislé náhodné veličiny s alternativním rozdělením, pro které pravděpodobnost úspěchu je dána předpisem

$$P[Y_i = 1|\boldsymbol{x}_i] = \frac{\exp(\boldsymbol{\beta}^\top \boldsymbol{x}_i)}{1 + \exp(\boldsymbol{\beta}^\top \boldsymbol{x}_i)} =: \pi_i,$$

kde $\boldsymbol{\beta} \in \mathbb{R}^p$ je vektor regresních parametrů, $\boldsymbol{x}_i \in \mathbb{R}^p$ je vektor regresorů. Podle zavedeného značení pak píšeme $P[Y_i = 0|\boldsymbol{x}_i] = 1 - \pi_i$. Zajímají nás vlastnosti odhadu parametru $\boldsymbol{\beta}$. Jako v předchozím příkladu, použijeme větu 8.

Příprava k použití věty 8.

Začneme tím, že se podíváme na konkrétní tvary výrazů, které budeme potřebovat. Uvažujeme opět funkci $g_i(y_i,\boldsymbol{\beta}) = -\log f_i(y_i|\boldsymbol{\beta})$ a v tomto případě, kde $f_i(y_i|\boldsymbol{\beta})$ je hustota alternativního rozdělení, tedy hustota vzhledem k čítací míře, dostáváme

$$g_i(y_i,\boldsymbol{\beta}) = \log(1 + \exp(\boldsymbol{\beta}^\top \boldsymbol{x}_i)) - y_i \boldsymbol{\beta}^\top \boldsymbol{x}_i.$$

Ověřme nejprve, že je tato funkce konvexní v $\boldsymbol{\beta}$. Vypočítáme tedy postupně vektor prvních parciálních derivací a matici druhých parciálních derivací

$$\begin{aligned} -\frac{\partial \log f_i(y_i|\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= y_i \boldsymbol{x}_i - \pi_i \boldsymbol{x}_i, \\ -\frac{\partial \log f_i(y_i|\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} &= \frac{\exp(\boldsymbol{\beta}^\top \boldsymbol{x}_i)}{(1 + \exp(\boldsymbol{\beta}^\top \boldsymbol{x}_i))^2} \boldsymbol{x}_i \boldsymbol{x}_i^\top. \end{aligned} \tag{2.15}$$

Z (2.15) vidíme, že matice druhých parciálních derivací funkce g_i podle $\boldsymbol{\beta}$, ve které vystupují pouze exponenciální funkce, které jsou kladné, a poté matice $\boldsymbol{x}_i \boldsymbol{x}_i^\top$, která je pozitivně semidefinitní, je celkově pozitivně semidefinitní a z toho plyne, že daná funkce je v $\boldsymbol{\beta}$ konvexní. Podívejme se dále, jak vypadá rozvoj (2.1) v tomto případě,

$$\begin{aligned} g_i(y_i,\boldsymbol{\beta}_0 + \mathbf{t}) - g_i(y_i,\boldsymbol{\beta}_0) &= \log f(y_i,\boldsymbol{\beta}_0) - \log f(y_i,\boldsymbol{\beta}_0 + \mathbf{t}) = \\ &= \log(1 + \exp((\boldsymbol{\beta}_0 + \mathbf{t})^\top \boldsymbol{x}_i)) - \log(1 + \exp(\boldsymbol{\beta}_0^\top \boldsymbol{x}_i)) - y_i \mathbf{t}^\top \boldsymbol{x}_i. \end{aligned} \tag{2.16}$$

Vzhledem k tomu, že funkce g_i je v tomto případě funkcí více proměnných a bylo by obtížné zapsat její Taylorův rozvoj (třetí parciální derivace), využijeme toho, že se nám podařilo přejít k funkci $\log(1 + \exp((\boldsymbol{\beta}_0 + \mathbf{t})^\top \boldsymbol{x}_i))$ a použijeme tedy dále Taylorův rozvoj pro tuto funkci, tedy pro funkci $\log(1 + \exp(u + h))$, kde později budeme uvažovat $u = \boldsymbol{\beta}_0^\top \boldsymbol{x}_i$ a $h = \boldsymbol{t}^\top \boldsymbol{x}_i$. Vypočteme nejprve první tři derivace funkce $\log(1 + \exp(u))$ podle u a vyjádříme je pomocí $\pi(u) = \frac{\exp(u)}{1+\exp(u)}$.

$$\begin{aligned} \Big(\log(1 + \exp(u))\Big)' &= \frac{\exp(u)}{1 + \exp(u)} = \pi(u) \\ \Big(\log(1 + \exp(u))\Big)'' &= \frac{\exp(u)}{(1 + \exp(u))^2} = \pi(u)(1 - \pi(u)) \\ \Big(\log(1 + \exp(u))\Big)''' &= \frac{\exp(u)(1 - \exp(u))}{(1 + \exp(u))^3} = \pi(u)(1 - \pi(u))(1 - 2\pi(u)) \end{aligned}$$

Taylorův rozvoj $\log(1+\exp(u+h))$ okolo bodu u tak můžeme vyjádřit jako

$$\begin{aligned}\log(1+\exp(u+h)) = \log(1+\exp(u)) + \pi(u)h + \frac{1}{2}\pi(u)(1-\pi(u))h^2 + \\ + \frac{1}{6}\pi(\tilde{u})(1-\pi(\tilde{u}))(1-2\pi(\tilde{u}))h^3,\end{aligned} \tag{2.17}$$

kde třetí člen rozvoje tvoří Lagrangeův zbytek v bodě $\tilde{u}$, který leží mezi body u a $u+h$.

Výraz $\pi(\tilde{u})(1-\pi(\tilde{u}))(1-2\pi(\tilde{u}))$ nahradíme výrazem $\gamma(u,h)$, kde $|\gamma(u,h)| < \exp(|h|)$. Toto můžeme provést, neboť lze ukázat, že

$$\pi(\tilde{u})(1-\pi(\tilde{u})) \leq \pi(u)(1-\pi(u))\exp(|h|),$$

bez ohledu na to, jaké jsou hodnoty u a h (Hjort a Pollard, 2011, str.12) a také platí $|1-2\pi(\tilde{u})| < 1$. Poslední nerovnost platí, neboť $\pi(\tilde{u})$ vyjadřuje pravděpodobnost a nabývá tedy hodnot z intervalu $\langle 0,1\rangle$ a rozdíl $1-2\pi(\tilde{u})$ se tak pohybuje v intervalu $\langle -1,1\rangle$.

Rozvoj (2.17) nyní použijeme k přepsání (2.16) do následujícího tvaru

$$\begin{aligned}&\log\frac{f(y_i,\boldsymbol{\beta}_0)}{f(y_i,\boldsymbol{\beta}_0+\mathbf{t})} = \\ &= \pi_i\mathbf{t}^\top\boldsymbol{x}_i + \frac{1}{2}\pi_i(1-\pi_i)(\mathbf{t}^\top\boldsymbol{x}_i)^2 + \frac{1}{6}\pi_i(1-\pi_i)\gamma(\boldsymbol{\beta}_0^\top\boldsymbol{x}_i,\boldsymbol{t}^\top\boldsymbol{x}_i)(\mathbf{t}^\top\boldsymbol{x}_i)^3 - y_i\mathbf{t}^\top\boldsymbol{x}_i \\ &\quad = (\pi_i-y_i)\boldsymbol{x}_i^\top\boldsymbol{t} + \frac{1}{2}\pi_i(1-\pi_i)(\mathbf{t}^\top\boldsymbol{x}_i)^2 + \frac{1}{6}\pi_i(1-\pi_i)\gamma(\boldsymbol{\beta}_0^\top\boldsymbol{x}_i,\boldsymbol{t}^\top\boldsymbol{x}_i)(\mathbf{t}^\top\boldsymbol{x}_i)^3,\end{aligned}$$

ze kterého vidíme, jak vypadají funkce $\boldsymbol{D}_i(y_i)$ a $R_i(y_i,\mathbf{t})$ z rozvoje (2.1) pro tento případ:

$$\boldsymbol{D}_i(Y_i,\boldsymbol{\beta}_\mathbf{0}) = (\pi_i - Y_i)\boldsymbol{x}_i, \tag{2.18}$$

$$R_i(Y_i,\mathbf{t}) = \frac{1}{2}\mathbf{t}^\top\pi_i(1-\pi_i)\boldsymbol{x}_i\boldsymbol{x}_i^\top\mathbf{t} + \frac{1}{6}\pi_i(1-\pi_i)\gamma(\boldsymbol{\beta}_0^\top\boldsymbol{x}_i,\boldsymbol{t}^\top\boldsymbol{x}_i)(\mathbf{t}^\top\boldsymbol{x}_i)^3. \tag{2.19}$$

Vzhledem k tomu, že $R_i(Y_i,\mathbf{t})$ v tomto případě už nezávisí na Y_i, není tedy náhodnou veličinou. Můžeme tedy přímo ze vztahu (2.19) zjistit tvar A_i a $v_{i,0}(\mathbf{t})$ (definovány v (2.2))

$$A_i = \pi_i(1-\pi_i)\boldsymbol{x}_i\boldsymbol{x}_i^\top \tag{2.20}$$

$$v_{i,0}(\mathbf{t}) = \frac{1}{6}\pi_i(1-\pi_i)\gamma(\boldsymbol{\beta}_0^\top\boldsymbol{x}_i,\boldsymbol{t}^\top\boldsymbol{x}_i)(\mathbf{t}^\top\boldsymbol{x}_i)^3. \tag{2.21}$$

Využijeme je v další části k ověření předpokladů věty 8.

Ověření předpokladů věty 8.

S pomocí (2.18) a (2.19) můžeme nyní přejít k ověření předpokladů potřebných k použití věty 8. Podívejme se nejprve na matice J_n a K_n, jejichž předpisy jsou definovány v části 2.2:

$$\begin{aligned} J_n &= \sum_{i=1}^{n} A_i = \sum_{i=1}^{n} \pi_i(1-\pi_i)\boldsymbol{x}_i\boldsymbol{x}_i^\top \\ K_n &= \sum_{i=1}^{n} B_i = \sum_{i=1}^{n} \mathrm{var}(\boldsymbol{D}_i(Y_i,\boldsymbol{\beta_0})) = \sum_{i=1}^{n} \mathrm{var}\left((\pi_i - Y_i)\boldsymbol{x}_i\right) \\ &= \sum_{i=1}^{n} \pi_i(1-\pi_i)\boldsymbol{x}_i\boldsymbol{x}_i^\top \end{aligned}$$

Vidíme, že matice J_n a K_n jsou stejné. Požadujeme, aby $\frac{J_n}{n} \xrightarrow[n\to\infty]{} J$, která je pozitivně definitní. Musíme tedy předpokládat, že $\frac{1}{n}\sum_{i=1}^{n} \pi_i(1-\pi_i)\boldsymbol{x}_i\boldsymbol{x}_i^\top \xrightarrow[n\to\infty]{} J$, kde J je pozitivně definitní matice.

Dále chceme ověřit, že $\sum_{i=1}^{n} v_i(\frac{\mathbf{s}}{\sqrt{n}}) \xrightarrow[n\to\infty]{} 0$, kde $v_i(\frac{\mathbf{s}}{\sqrt{n}}) = \mathrm{var}\, R_i(Y_i,\frac{\mathbf{s}}{\sqrt{n}})$. Vzhledem k tomu, že $R_i(Y_i,\frac{\mathbf{s}}{\sqrt{n}})$ zde nezávisí na náhodné veličině Y_i, je rozptyl nulový a tento předpoklad je tedy splněn.

Předposledním předpokladem, který zbývá ověřit, je $\sum_{i=1}^{n} v_{i,0}(\frac{\mathbf{s}}{\sqrt{n}}) \xrightarrow[n\to\infty]{} 0$. Podle předpisu pro $v_{i,0}(\boldsymbol{t})$ z (2.21) a s využitím dříve získané nerovnosti

$$|\gamma(\boldsymbol{\beta}_0^\top \boldsymbol{x}_i,\boldsymbol{t}^\top \boldsymbol{x}_i)| \leq \exp(|\boldsymbol{t}^\top \boldsymbol{x}_i|)$$

dostáváme pro $\boldsymbol{t} = \frac{\mathbf{s}}{\sqrt{n}}$

$$\left|\sum_{i=1}^{n} v_{i,0}\left(\frac{\mathbf{s}}{\sqrt{n}}\right)\right| \leq \sum_{i=1}^{n} \frac{1}{6}\pi_i(1-\pi_i)\exp\left(\left|\left(\frac{\mathbf{s}}{\sqrt{n}}\right)^\top \boldsymbol{x}_i\right|\right) \cdot \left|\left(\frac{\mathbf{s}}{\sqrt{n}}\right)^\top \boldsymbol{x}_i\right|^3, \tag{2.22}$$

dále můžeme $|(\frac{\mathbf{s}}{\sqrt{n}})^\top \boldsymbol{x}_i|^3$ rozepsat jako $\frac{1}{n}\mathbf{s}^\top \boldsymbol{x}_i\boldsymbol{x}_i^\top \mathbf{s}|(\frac{\mathbf{s}}{\sqrt{n}})^\top \boldsymbol{x}_i|$ a použít také Cauchy-ho-Schwartzovu nerovnost a dostaneme $|(\frac{\mathbf{s}}{\sqrt{n}})^\top \boldsymbol{x}_i| \leq \frac{1}{\sqrt{n}}\|\mathbf{s}\| \cdot \|\boldsymbol{x}_i\|$. Dále můžeme $\|\boldsymbol{x}_i\|$ shora omezit výrazem $\max\limits_{i=1,\dots,n}\|\boldsymbol{x}_i\|$, čímž zbavíme tento člen závislosti na i. Pravou stranu nerovnosti (2.22) můžeme tedy shora omezit výrazem

$$\sum_{i=1}^{n} \frac{1}{6}\pi_i(1-\pi_i)\exp\left(\frac{1}{\sqrt{n}}\|\mathbf{s}\| \cdot \max_{i=1,\dots,n}\|\boldsymbol{x}_i\|\right)\frac{1}{n}\mathbf{s}^\top \boldsymbol{x}_i\boldsymbol{x}_i^\top \mathbf{s}\frac{1}{\sqrt{n}}\|\mathbf{s}\| \cdot \max_{i=1,\dots,n}\|\boldsymbol{x}_i\|,$$

vzhledem k tomu, že některé ze členů už nezávisí na indexu i, dostáváme po přeuspořádání výraz

$$\frac{1}{6}\exp\left(\|\mathbf{s}\|\frac{1}{\sqrt{n}}\max_{i=1,\dots,n}\|\boldsymbol{x}_i\|\right)\|\mathbf{s}\|\frac{1}{\sqrt{n}}\max_{i=1,\dots,n}\|\boldsymbol{x}_i\|\mathbf{s}^\top\frac{1}{n}\sum_{i=1}^{n}(\pi_i(1-\pi_i)\boldsymbol{x}_i\boldsymbol{x}_i^\top)\mathbf{s},$$

kde $\sum_{i=1}^{n}(\pi_i(1-\pi_i)\boldsymbol{x}_i\boldsymbol{x}_i^\top) = J_n$. Pro $n \to \infty$ konverguje $\frac{J_n}{n}$ k matici J, která je pozitivně definitní. Vzhledem k tomu, že požadujeme, aby $\sum_{i=1}^{n} v_{i,0}\left(\frac{\mathbf{s}}{\sqrt{n}}\right)$ konvergoval k 0 pro každé $\boldsymbol{s} \in \mathbb{R}^p$, potřebujeme, aby platilo, že $\frac{1}{\sqrt{n}}\max\limits_{i=1,\dots,n}\|\boldsymbol{x}_i\| \xrightarrow[n\to\infty]{} 0$. Je-li tento požadavek splněn, pak také $\sum_{i=1}^{n} v_{i,0}(\frac{\mathbf{s}}{\sqrt{n}}) \xrightarrow[n\to\infty]{} 0$ pro každé $\boldsymbol{s} \in \mathbb{R}^p$.

Asymptotická normalita

Nakonec ukážeme, že pro posloupnost náhodných vektorů $\frac{1}{\sqrt{n}}\boldsymbol{D}_i(Y_i,\boldsymbol{\beta_0})$ pro $i \in \{1,\dots,n\}$ platí Feller-Lindebergova centrální věta (v příloze, věta A.2). Použijeme Feller-Lindebergovu podmínku pro lineární kombinace náhodných vektorů

$\frac{1}{\sqrt{n}}\boldsymbol{\lambda}^\top \boldsymbol{D}_i(Y_i,\boldsymbol{\beta_0})$. Po dosazení za $\boldsymbol{D}_i(Y_i,\boldsymbol{\beta_0})$ z (2.18) vypadá F.-L. podmínka takto

$$\forall\varepsilon > 0 \lim_{n\to\infty} \sum_{i=1}^{n} \mathsf{E}\left[\frac{1}{n}(\pi_i - Y_i)^2(\boldsymbol{\lambda}^\top \boldsymbol{x}_i)^2 \mathbb{I}\left\{\left|\frac{1}{\sqrt{n}}(\pi_i - Y_i)\boldsymbol{\lambda}^\top \boldsymbol{x}_i\right| \geq \varepsilon\right\}\right] = 0.$$

Označme

$$F_n(\varepsilon) := \sum_{i=1}^{n} \mathsf{E}\left[\frac{1}{n}(\pi_i - Y_i)^2(\boldsymbol{\lambda}^\top \boldsymbol{x}_i)^2 \mathbb{I}\left\{\left|\frac{1}{\sqrt{n}}(\pi_i - Y_i)\boldsymbol{\lambda}^\top \boldsymbol{x}_i\right| \geq \varepsilon\right\}\right].$$

Pokud $F_n(\varepsilon)$ dokážeme shora omezit výrazem, který konverguje k 0, bude F.-L. podmínka splněna. Ukažme nejprve, že

$$\begin{aligned}\left|(\pi_i - Y_i)\boldsymbol{\lambda}^\top \boldsymbol{x}_i\right| &\leq |(\pi_i - Y_i)| \cdot \left|\boldsymbol{\lambda}^\top \boldsymbol{x}_i\right| \\ &\leq 2\|\boldsymbol{\lambda}\| \cdot \max_{i=1,\dots,n} \|\boldsymbol{x}_i\|,\end{aligned}$$

s použitím Cauchyho-Schwartzovy nerovnosti a také toho, že náhodná veličina Y_i má alternativní rozdělení, a tedy nabývá pouze hodnot 0 a 1. Nahradíme-li tímto výrazem výraz v indikátoru $F_n(\varepsilon)$ dostáváme, že

$$\begin{aligned}F_n(\varepsilon) &\leq \sum_{i=1}^{n} \mathsf{E}\left[\frac{1}{n}(\pi_i - Y_i)^2(\boldsymbol{\lambda}^\top \boldsymbol{x}_i)^2 \mathbb{I}\left\{2\frac{1}{\sqrt{n}}\|\boldsymbol{\lambda}\| \cdot \max_{i=1,\dots,n}\|\boldsymbol{x}_i\| \geq \varepsilon\right\}\right] \\ &\leq \mathbb{I}\left\{2\frac{1}{\sqrt{n}}\|\boldsymbol{\lambda}\| \cdot \max_{i=1,\dots,n}\|\boldsymbol{x}_i\| \geq \varepsilon\right\} \sum_{i=1}^{n} \mathsf{E}\left[\frac{1}{n}(\pi_i - Y_i)^2(\boldsymbol{\lambda}^\top \boldsymbol{x}_i)^2\right],\end{aligned}$$

neboť indikátor už nezávisí na náhodné veličině Y_i. Dále můžeme upravit

$$\begin{aligned}\sum_{i=1}^{n} \mathsf{E}\left[\frac{1}{n}(\pi_i - Y_i)^2(\boldsymbol{\lambda}^\top \boldsymbol{x}_i)^2\right] &= \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{\lambda}^\top \boldsymbol{x}_i)^2\, \mathsf{E}\left[(\pi_i - Y_i)^2\right] \\ &= \frac{1}{n}\sum_{i=1}^{n}\left((\boldsymbol{\lambda}^\top \boldsymbol{x}_i)^2(1-\pi_i)\pi_i\right) \\ &= \frac{1}{n}\boldsymbol{\lambda}^\top \sum_{i=1}^{n}\left((1-\pi_i)\pi_i \boldsymbol{x}_i \boldsymbol{x}_i^\top\right)\boldsymbol{\lambda} \\ &= \frac{1}{n}\boldsymbol{\lambda}^\top J_n \boldsymbol{\lambda}.\end{aligned}$$

$F_n(\varepsilon)$ tedy můžeme omezit shora výrazem

$$F_n(\varepsilon) \leq \mathbb{I}\left\{2\|\boldsymbol{\lambda}\|\frac{1}{\sqrt{n}}\max_{i=1,\dots,n}\|\boldsymbol{x}_i\| \geq \varepsilon\right\}\boldsymbol{\lambda}^\top \frac{J_n}{n}\boldsymbol{\lambda}.$$

Pokud předpokládáme, že $\frac{1}{\sqrt{n}}\max_{i=1,\dots,n}\|\boldsymbol{x}_i\| \xrightarrow[n\to\infty]{} 0$ (což jsme potřebovali již dříve), potom

$$\mathbb{I}\left\{\left(2\|\boldsymbol{s}\|\frac{1}{\sqrt{n}}\max_{i=1,\dots,n}\|\boldsymbol{x}_i\|\right) \geq \varepsilon\right\} \xrightarrow[n\to\infty]{} 0$$

a Feller-Lindebergova podmínka je splněna, z čehož plyne

$$\boldsymbol{U_n} = \sum_{i=1}^{n}\frac{1}{\sqrt{n}}((\pi_i - Y_i)\boldsymbol{x}_i) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}_p(\boldsymbol{0}_p, J).$$

Tedy $-J^{-1}\boldsymbol{U_n}$ konverguje k $\mathcal{N}_p(\boldsymbol{0}_p, J^{-1})$.

Závěr

Pro logistický regresní model s pevnými regresory jsme ukázali konzistenci a asymptotickou normalitu odhadu regresních koeficientů $\boldsymbol{\beta}$ za následujících podmínek

(i) $\frac{1}{n}\sum_{i=1}^{n}\pi_i(1-\pi_i)\boldsymbol{x}_i\boldsymbol{x}_i^\top \xrightarrow[n\to\infty]{} J$, kde J je pozitivně definitní matice,

(ii) $\frac{1}{\sqrt{n}}\max_{i=1,\dots,n}\|\boldsymbol{x}_i\| \xrightarrow[n\to\infty]{} 0.$

Dostali jsme tedy obdobné podmínky jako pro normální lineární model. Podmínku (i) jsme ale v tomto případě získali v pěknějším tvaru díky alternativnímu rozdělení odezev Y_i, které můžeme jednoduše omezit. Můžeme poznamenat, že pro model s náhodnými regresory bychom pro zajištění konzistence a asymptotické normality museli předpokládat $\mathsf{E}\ \pi_1(1-\pi_1)\boldsymbol{X}_1\boldsymbol{X}_1^\top = W$, kde W je pozitivně definitní matice a $\boldsymbol{X}_1$ je náhodný vektor regresorů.

2.4.3 Poissonovská regrese

Uvažujme další z často používaných modelů, kde předpokládáme, že pozorování $Y_1,\dots,Y_n$ pocházejí z Poissonova rozdělení. Střední hodnota tohoto rozdělení závisí na regresorech $\boldsymbol{x}_i \in \mathbb{R}^p$ následujícím způsobem

$$Y_i \sim Po(\mu_i),\ \text{kde}\ \mu_i = \exp(\boldsymbol{\beta}^\top\boldsymbol{x}_i)$$

a $\boldsymbol{\beta} \in \mathbb{R}^p$ je vektor regresních koeficientů, pro jehož odhad chceme ukázat konzistenci a asymptotickou normalitu. Jeho skutečnou hodnotu označíme $\boldsymbol{\beta}_0$. Pracujeme tedy s hustotou vzhledem k čítací míře. Hustota je tvaru

$$f_i(y_i|\boldsymbol{\beta}) = \frac{\mu_i}{y_i!}\exp(-\mu_i) \quad \text{pro}\ y_i = 0{,}1{,}2,\dots.$$

Uvažujeme nadále maximálně věrohodný odhad parametru $\boldsymbol{\beta}$, který získáme minimalizací výrazu $\sum_{i=1}^{n} g_i(Y_i,\boldsymbol{\beta})$. Funkce $g_i(Y_i,\boldsymbol{\beta})$ má v tomto případě následující tvar

$$\begin{aligned} g_i(Y_i,\boldsymbol{\beta}) &= -Y_i\log\mu_i + \log(Y_i!) + \mu_i \\ &= -Y_i\boldsymbol{\beta}^\top\boldsymbol{x}_i + \log(Y_i!) + \exp(\boldsymbol{\beta}^\top\boldsymbol{x}_i). \end{aligned}$$

Ověřme nejprve, že je funkce g_i v $\boldsymbol{\beta}$ konvexní. Matice druhých parciálních derivací je tvaru $\exp(\boldsymbol{\beta}^\top\boldsymbol{x}_i)\boldsymbol{x}_i\boldsymbol{x}_i^\top$. Matice $\boldsymbol{x}_i\boldsymbol{x}_i^\top$ je pozitivně semidefinitní matice, což se nezmění po vynásobení kladným číslem $\exp(\boldsymbol{\beta}^\top\boldsymbol{x}_i)$, a tedy podmínka potřebná pro konvexitu funkce g_i je splněna. Výraz $\log(Y_i!)$ ve funkci g_i nezávisí na parametru $\boldsymbol{\beta}$, a proto jej při minimalizaci nemusíme brát v úvahu a minimum z $\sum_{i=1}^{n} g_i(Y_i,\boldsymbol{\beta})$ se tím nezmění. Uvažujeme tedy dále jen funkci $-Y_i\boldsymbol{\beta}^\top\boldsymbol{x}_i + \exp(\boldsymbol{\beta}^\top\boldsymbol{x}_i)$.

Pro důkaz konzistence a asymptotické normality maximálně věrohodného odhadu $\boldsymbol{\beta}$ využijeme opět větu 8.

Příprava k použití věty 8.

Přejděme nyní k rozvoji (2.1)

$$g_i(y_i,\boldsymbol{\beta}_0+\mathbf{t})-g_i(y_i,\boldsymbol{\beta}_0)=\exp((\boldsymbol{\beta}_0+\mathbf{t})^\top\boldsymbol{x}_i))-y_i(\boldsymbol{\beta}_0+\mathbf{t})^\top\boldsymbol{x}_i \tag{2.23}$$
$$-\exp(\boldsymbol{\beta}_0^\top\boldsymbol{x}_i)+y_i\boldsymbol{\beta}_0^\top\boldsymbol{x}_i \tag{2.24}$$
$$=-y_i\boldsymbol{t}^\top\boldsymbol{x}_i+\exp(\boldsymbol{\beta}_0^\top\boldsymbol{x}_i)[\exp(\boldsymbol{t}^\top\boldsymbol{x}_i)-1]. \tag{2.25}$$

Dále využijeme Taylorova rozvoje pro exponenciální funkci, ze kterého dostáváme $\exp(u)=1+u+\frac{1}{2}u^2+\frac{1}{6}\tilde{u}^3\exp(\tilde{u})$ pro $\tilde{u}$ ležící mezi body 0 a u. Část rozvoje $\tilde{u}^3\exp(\tilde{u})$ můžeme nahradit funkcí $\rho(u)$, pro kterou platí, že

$$|\rho(u)|\leq|u|^3\exp(|u|), \tag{2.26}$$

což můžeme jednoduše ukázat

$$\begin{aligned}|\rho(u)|&=|\tilde{u}^3\exp(\tilde{u})|\\&=|\tilde{u}^3|\cdot|\exp(\tilde{u})|\\&\leq|\tilde{u}|^3\exp(|\tilde{u}|),\end{aligned} \tag{2.27}$$

neboť exponenciální funkce je kladná a rostoucí funkce. Vzhledem k tomu, že také platí, že $\tilde{u}$ leží mezi 0 a u a obě funkce z (2.27) jsou v $|\tilde{u}|$ rostoucí, můžeme je tedy shora omezit výrazem na pravé straně nerovnosti (2.26).

Použitím Taylorova rozvoje pro exponenciálu a označením $\mu_i=\exp(\boldsymbol{\beta}_0^\top\boldsymbol{x}_i)$ (nyní ve skutečném parametru) můžeme psát

$$\begin{aligned}(2.25)&=-y_i\boldsymbol{t}^\top\boldsymbol{x}_i+\mu_i\left(\boldsymbol{t}^\top\boldsymbol{x}_i+\frac{1}{2}\boldsymbol{t}^\top\boldsymbol{x}_i\boldsymbol{x}_i^\top\boldsymbol{t}+\frac{1}{6}\mu_i\rho(\boldsymbol{t}^\top\boldsymbol{x}_i)\right)\\&=(\mu_i-Y_i)\boldsymbol{x}_i^\top\boldsymbol{t}+\frac{1}{2}\boldsymbol{t}^\top\mu_i\boldsymbol{x}_i\boldsymbol{x}_i^\top\boldsymbol{t}+\frac{1}{6}\mu_i\rho(\boldsymbol{t}^\top\boldsymbol{x}_i),\end{aligned}$$

odkud vidíme, jak v tomto případě vypadají výrazy, které nás zajímají

$$\boldsymbol{D_i}(y_i,\boldsymbol{\beta}_0)=(\mu_i-Y_i)\boldsymbol{x}_i, \tag{2.28}$$
$$R_i(y_i,\mathbf{t})=\frac{1}{2}\boldsymbol{t}^\top\mu_i\boldsymbol{x}_i\boldsymbol{x}_i^\top\boldsymbol{t}+\frac{1}{6}\mu_i\rho(\boldsymbol{t}^\top\boldsymbol{x}_i). \tag{2.29}$$

Výraz $R_i(y_i,\mathbf{t})$ nezávisí na náhodné veličině Y_i, pro výrazy z (2.2) tedy platí, že $A_i=\mu_i\boldsymbol{x}_i\boldsymbol{x}_i^\top$ a $v_{i,0}(\boldsymbol{t})=\frac{1}{6}\mu_i\rho(\boldsymbol{t}^\top\boldsymbol{x}_i)$.

Ověření předpokladů věty 8.

Postupujeme stejně jako v předchozím příkladu pro logistickou regresi (část 2.4.2). Začněme maticemi J_n a K_n. Z předpisu pro A_i v tomto případě máme, že

$$J_n=\sum_{i=1}^n\mu_i\boldsymbol{x}_i\boldsymbol{x}_i^\top.$$

Snadno lze také vidět, že matice K_n je stejného tvaru (získáme ji jako součet rozptylů náhodných vektorů $(\mu_i-Y_i)\boldsymbol{x}_i$). Pro splnění předpokladu pro matici J_n je tedy nutné předpokládat $\frac{1}{n}\sum_{i=1}^n\mu_i\boldsymbol{x}_i\boldsymbol{x}_i^\top\xrightarrow[n\to\infty]{}J$, kde J je pozitivně definitní.

Dále ověříme, že $\sum_{i=1}^n v_i(\frac{\mathbf{s}}{\sqrt{n}}) \xrightarrow[n\to\infty]{} 0$, kde $v_i(\frac{\mathbf{s}}{\sqrt{n}}) = \operatorname{var} R_i\left(Y_i, \frac{\mathbf{s}}{\sqrt{n}}\right)$. Vzhledem k tomu, že $R_i\left(Y_i, \frac{\mathbf{s}}{\sqrt{n}}\right)$ ani v tomto případě nezávisí na náhodné veličině Y_i, je rozptyl nulový a tento předpoklad je tedy splněn. Přejdeme k ověření předpokladu $\sum_{i=1}^n v_{i,0}(\frac{\mathbf{s}}{\sqrt{n}}) \xrightarrow[n\to\infty]{} 0$, kde $v_{i,0}(\frac{\mathbf{s}}{\sqrt{n}}) = \frac{1}{6}\mu_i\rho((\frac{\mathbf{s}}{\sqrt{n}})^\top \boldsymbol{x}_i)$. Postupujeme obdobně jako v příkladu s logistickou regresí

$$\begin{aligned}
\left|\sum_{i=1}^n v_{i,0}\left(\frac{\mathbf{s}}{\sqrt{n}}\right)\right| &\leq \sum_{i=1}^n \frac{1}{6}\mu_i \left|\rho\left(\left(\frac{\mathbf{s}}{\sqrt{n}}\right)^\top \boldsymbol{x}_i\right)\right| \\
&\leq \sum_{i=1}^n \frac{1}{6}\mu_i \left|\left(\frac{\mathbf{s}}{\sqrt{n}}\right)^\top \boldsymbol{x}_i\right|^3 \exp\left(\left|\left(\frac{\mathbf{s}}{\sqrt{n}}\right)^\top \boldsymbol{x}_i\right|\right) \\
&\leq \frac{1}{6}\|\mathbf{s}\| \frac{1}{\sqrt{n}} \max_{i=1,\dots,n} \|\boldsymbol{x}_i\| \exp\left(\|\mathbf{s}\| \frac{1}{\sqrt{n}} \max_{i=1,\dots,n} \|\boldsymbol{x}_i\|\right) \mathbf{s}^\top \frac{1}{n}\sum_{i=1}^n \mu_i \boldsymbol{x}_i \boldsymbol{x}_i^\top \mathbf{s},
\end{aligned}$$

kde $\frac{1}{n}\sum_{i=1}^n \mu_i \boldsymbol{x}_i \boldsymbol{x}_i^\top \xrightarrow[n\to\infty]{} J$. Předpoklad $\frac{1}{\sqrt{n}} \max_{i=1,\dots,n} \|\boldsymbol{x}_i\| \xrightarrow[n\to\infty]{} 0$ tak znovu zaručí splnění podmínky $\sum_{i=1}^n v_{i,0}(\frac{\mathbf{s}}{\sqrt{n}}) \xrightarrow[n\to\infty]{} 0$.

Asymptotická normalita

Nakonec zbývá ověřit Lindebergovu podmínku pro posloupnost $\frac{1}{\sqrt{n}}(\mu_i - Y_i)\boldsymbol{x}_i$ k zajištění asymptotické normality. Stejně jako v případě normálního lineárního modelu přejdeme k ověření Ljapunovovy podmínky, která je tohoto tvaru

$$\forall \boldsymbol{\lambda} \in \mathbb{R}^p \exists \delta > 0 \lim_{n\to\infty} \sum_{k=1}^n \mathsf{E}\left[\left|\frac{1}{\sqrt{n}}(\mu_i - Y_i)\boldsymbol{\lambda}^\top \boldsymbol{x}_i\right|^{2+\delta}\right] = 0.$$

Označme

$$L_n(\delta) := \sum_{i=1}^n \mathsf{E}\left[\left|\frac{1}{\sqrt{n}}(\mu_i - Y_i)\boldsymbol{\lambda}^\top \boldsymbol{x}_i\right|^{2+\delta}\right].$$

Podívejme se dále na výraz $L_n(\delta)$

$$\begin{aligned}
L_n(\delta) &= \sum_{i=1}^n \mathsf{E}\left[\left(\frac{1}{n}\right)^{1+\delta/2} |(\mu_i - Y_i)|^{2+\delta} (\boldsymbol{\lambda}^\top \boldsymbol{x}_i)^2 \left|\boldsymbol{\lambda}^\top \boldsymbol{x}_i\right|^\delta\right] \\
&\leq \frac{1}{n^{\delta/2}} \|\boldsymbol{\lambda}\|^\delta \max_{i=1,\dots,n} \|\boldsymbol{x}_i\|^\delta \boldsymbol{\lambda}^\top \left(\frac{1}{n}\sum_{i=1}^n \mathsf{E}\, |(\mu_i - Y_i)|^{2+\delta} \boldsymbol{x}_i \boldsymbol{x}_i^\top\right) \boldsymbol{\lambda},
\end{aligned}$$

kde jsme využili Cauchyho-Schwartzovy nerovnosti a nahrazení norem z vektorů $\boldsymbol{x}_i$ jejich maximem přes všechna pozorování jako v předchozích příkladech.

Předpokládáme-li, že

$$\frac{1}{\sqrt{n}} \max_{i=1,\dots,n} \|\boldsymbol{x}_i\| \xrightarrow[n\to\infty]{} 0 \tag{2.30}$$

$$\frac{1}{n}\sum_{i=1}^n \mathsf{E}\, |(\mu_i - Y_i)|^{2+\delta} \boldsymbol{x}_i \boldsymbol{x}_i^\top \xrightarrow[n\to\infty]{} C, \tag{2.31}$$

kde C je konečná matice. Potom je Ljapunovova podmínka splněna, a tedy jsme ukázali asymptotickou normalitu.

Závěr

Jsou-li splněny předpoklady (2.30) a (2.31), máme zaručenu konzistenci pro odhad vektoru regresních koeficientů $\boldsymbol{\beta}$ a také jeho asymptotickou normalitu.

Závěr

V práci jsme se seznámili s metodou maximální věrohodnosti pro pozorování, která jsou nezávislá, ale nejsou stejně rozdělená. Zabývali jsme se podmínkami, které musí být splněny, abychom i pro odhady parametrů v tomto případě získali konzistenci a asymptotickou normalitu. Příkladem, na kterém jsme použití tohoto postupu a ověření podmínek ilustrovali, byl K-výběrový problém. Je zřejmé, že pro jednoduché úlohy tohoto typu bychom upřednostnili k ukázání požadovaných vlastností použití základnějších výsledků a nevolili bychom obecnou metodu. Příklad tak naznačil, že pro komplikovanější situace, kdy nemůžeme konzistenci a asymptotickou normalitu ukázat snáze, může být ověřování podmínek pro konzistenci a asymptotickou normalitu značně obtížné.

Lépe srozumitelné a ověřitelné podmínky jsme dostali pro případ, kdy měla naše pozorování logkonkávní hustotu. V tomto případě jsme mohli využít asymptotické vlastnosti speciální třídy M-odhadů, které jsou založeny na konvexních ztrátových funkcích. Ve třech příkladech různých regresních modelů, na které jsme tuto metodu použili, jsme dostali podmínky, které nejsou pro běžné aplikace svazující a lze je jednodušeji ověřit.

Zabývali jsme se pouze porušením předpokladu stejné rozdělenosti. Rozšíření práce by tak směřovalo ke studiu situací, kdy je porušen předpoklad nezávislosti jednotlivých pozorování, tedy například využití metody maximální věrohodnosti pro časové řady. Dalším možným rozšířením práce by bylo použití asymptotické teorie pro odhady získané minimalizací konvexních ztrátových funkcí, kterou jsme se zabývali v druhé části práce, na další zobecněné lineární modely nebo pro cenzorovaná data.

Seznam použité literatury

BILLINGSLEY, P. (2008). *Probability and Measure, 3rd ed.* Wiley series in probability and mathematical statistics. Wiley India Pvt. Limited. ISBN 9788126517718.

HJORT, N. L. a POLLARD, D. (2011). Asymptotics for minimisers of convex processes. arXiv:1107.3806.

HOADLEY, B. (1971). Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case. *The Annals of Mathematical Statistics*, **42**(6), 1977–1991. ISSN 00034851. URL http://www.jstor.org/stable/2240126.

KOMÁREK, A. (2016). Course notes, NMSA407 Linear Regression. Poznámky k přednášce Lineární regrese, MFF UK. URL http://msekce.karlin.mff.cuni.cz/~komarek/vyuka/2015_16/nmsa407/2015-NMSA407-notes.pdf.

LEHMANN, E. a CASELLA, G. (2003). *Theory of Point Estimation.* Springer Texts in Statistics. Springer New York. ISBN 9780387985022.

PAWITAN, Y. (2001). *In All Likelihood: Statistical Modelling and Inference Using Likelihood.* Oxford science publications. OUP Oxford. ISBN 9780198507659.

Přílohy

Centrální limitní věty

Následuje mnohorozměrná Ljapunovova centrální limitní věta, která je hlavním klíčem v důkazu asymptotické normality v případě, kdy nemáme stejně rozdělená pozorování.

Věta A. 1 (Ljapunovova centrální limitní věta)**.** *(Hoadley, 1971, Věta A.6) Nechť* $\{\boldsymbol{X}_i : i = 1,2,\ldots\}$ *jsou nezávislé p-rozměrné náhodné vektory, pro které platí* $\mathsf{E}\,\boldsymbol{X}_i = 0$ *a* $\operatorname{var}(\boldsymbol{X}_i) = \Gamma_i$ *a předpokládejme dále, že*

- *(i)* $\overline{\Gamma}_n \xrightarrow[n\to\infty]{} \overline{\Gamma}$ *a* $\overline{\Gamma}$ *je pozitivně definitní, kde* $\overline{\Gamma}_n = \frac{1}{n}\sum_{i=1}^n \Gamma_i$.
- *(ii) Pro nějaké* $\delta > 0$, $\sum_{i=1}^n \mathsf{E}\,|\boldsymbol{\lambda}^\top \boldsymbol{X}_i|^{2+\delta}/n^{(2+\delta)/2} \xrightarrow[n\to\infty]{} 0$ *pro všechna* $\boldsymbol{\lambda} \in \mathbb{R}^p$,

potom

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{X}_i \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0,\overline{\Gamma}).$$

Dále uvedeme mnohorozměrnou Feller-Lindebergovu centrální větu upravenou pro její použití v této práci. Dokazujeme asymptotickou normalitu pro posloupnost náhodných vektorů pomocí Cramérovy-Woldovy věty, díky které můžeme přejít od náhodných vektorů k náhodným veličinám, které jsou lineární kombinací původních náhodných vektorů.

Věta A. 2 (Feller-Lindebergova centrální limitní věta)**.** *Nechť* $\boldsymbol{X}_i$, $i \in \mathbb{N}$ *je posloupnost p-rozměrných náhodných vektorů, pro které platí* $\mathsf{E}\,\boldsymbol{X}_i = \boldsymbol{0}_p$ *a* $\operatorname{var}\boldsymbol{X}_i = \Gamma_i$, *kde* Γ_i *je konečná matice pro všechna* $i = 1,\ldots,n$. *Předpokládejme dále, že* $\lim\limits_{n\to\infty}\sum_{i=1}^n \operatorname{var}\boldsymbol{X}_i = \Gamma$, *kde* Γ *je konečná matice. Jestliže je splněna Feller-Lindebergova podmínka pro lineární kombinace těchto vektorů, tj.*

$$\forall \boldsymbol{\lambda} \in \mathbb{R}^p\ \forall \varepsilon > 0 : \lim_{n\to\infty}\sum_{i=1}^n \mathsf{E}\,[(\boldsymbol{\lambda}^\top \boldsymbol{X}_i)^2 \mathbb{I}\{|\boldsymbol{\lambda}^\top \boldsymbol{X}_i| \geq \varepsilon\}] = 0, \tag{1}$$

potom

$$\sum_{i=1}^n \boldsymbol{X}_i \xrightarrow[n\to\infty]{\mathcal{D}} \boldsymbol{Y},\ \boldsymbol{Y} \sim \mathcal{N}_p(\boldsymbol{0}_p,\Gamma).$$

Důkaz. Využijeme Cramér-Woldovy věty (Billingsley, 2008, Věta 29.4., str. 383) a od náhodných vektorů přejdeme k jejich lineárním kombinacím, které jsou náhodnými veličinami a ukážeme asymptotickou normalitu pro ně pomocí Feller-Lindebergovy věty (Billingsley, 2008, Věta 27.2., str. 359) s podmínkou (2), ve které s_n^2 je součet rozptylů veličin $\boldsymbol{\lambda}^\top \boldsymbol{X}_i$.

$$\forall \varepsilon > 0 : \lim_{n\to\infty}\sum_{i=1}^n \frac{1}{s_n^2}\,\mathsf{E}\,[(\boldsymbol{X}_i)^2 \mathbb{I}\{|\boldsymbol{X}_i| \geq s_n\varepsilon\}] = 0, \tag{2}$$

Vzhledem k tomu, že s_n^2 konverguje ke konečné konstantě $\boldsymbol{\lambda}^\top\Gamma\boldsymbol{\lambda}$ (neboť Γ je konečná matice), stačí ukázat, že platí podmínka (1), ve které neuvažujeme rozptyl veličin $\boldsymbol{\lambda}^\top \boldsymbol{X}_i$.

□