ROBUST'2000, 149 – 163

PROCES ODHADOVÁNÍ PARAMETRŮ MODELU

Petr Lachout

Abstract. Tento pspvek souhrn rekapituluje a diskutuje proces odhadovn parametr modelu. Nejdve je vysvtleno, co se mysl modelem njak reln situace a co jsou jeho parametry. Pak je zaveden pojem odhadu parametr, lpe eeno, odhad t sti parametr, kter je pro chovn modelu podstatn. Diskutuje se kvalita tchto odhad a jejich optimln volba. Je pipomenuta metoda maximln vrohodnosti a zbyl kapitoly pinej konstrukce odhad zaloen na minimalizaci empirick chyby modelu linern regrese. V prci je teoreticky diskutovn M-odhad a jsou zmnny jeho speciln volby, jako je OLS-odhad a L_1-odhad. V zvren kapitole je zmnn LMS-odhad.

Abstract: The paper presents an overview on parameter estimation process. An explanation of the concept of mathematical models and model parameters starts the discussion. After that, the idea of parameter estimation is presented together with a discussion of convenient and optimal properties an estimator should fulfill. Hence, we proceed to constructions of estimators. We present maximal likelihood estimator and estimators based on minimization of empirical errors. The last section considers a linear regression model. M-estimators are treated as a generalization of OLS- and L_1-estimator. The text finishes with a note on LMS-estimator.

Резюме:Статя подытоживает и дискутирует процесс оценки параметров модели. С начала понятие модели, параметров модели и оценки параметров пояснены. После етого введения дискутирована проблематика выбора оценки имеющей полезные или оптималные свойства. Другие главы посвящённы возможностям постраения оценок. Представлен метод максималной вероятности и метод минимализации емпирической ошибки, каторый показан на моделы линейной регресии. Теоретически дискутирована М-оценка и её частные случаи, как ОЛС-оценка и L_1-оценка. Последная глава посвящённая ЛМС-оценке.

1. Pozorování, model, odhad parametrů

Představme si, že studujeme nějaký konkrétní systém, např. pražskou burzu, ekonomiku nějakého průmyslového odvětví, rodinný rozpočet, atd. U tohoto systému jsme schopni opakovaně měřit nějaké veličiny, které jsou tímto procesem ovlivněny, vstupují do něj nebo jsou systémem plně určeny. Na burze sledujeme ceny akcií, objem jejich obchodování atd. V průmyslovém odvětví máme informaci o zásobách výrobních surovin, o kapitálu uloženém v budovách a strojích, o výrobních nákladech, o nákladech na změnu výroby, atd. V rodině známe příjmy jednotlivých členů, jejich stáří, výdaje a celkové úspory. Označme tato opakovaná pozorování $X_1, X_2, \ldots, X_n$ a množinu, kde se nacházejí všechny možné hodnoty měření, označme $\mathbb{X}$. Aby bylo možno studovat stochastické chování, musí být množina $\mathbb{X}$ opatřena σ-algebrou, označme ji $\mathcal{X}$. Poznamenejme zde, že na základě samotných pozorování nedokážeme určit nebo popsat chování pozorovaného systému. Tento problém je principiální, zjednodušeně řečeno „nemáme se oč opřít“.

1991 *Mathematics Subject Classification.* 62F10, 62J99.
Key words and phrases. Odhady, regrese.
Vzniklo za podpory grantu MSM 113200008

Na pomoc musí přijít znalost mechaniky sledované situace, zkušenost a také intuice. Na jejich základě navrhneme matematický model sledovaného systému, který teoreticky vysvětluje vznik a vnitřní strukturu pozorování $X_1, X_2, \ldots, X_n$, to znamená, že známe rozdělení $\mathcal{L}(X_1, X_2, \ldots, X_n) = \mu_n(.;\theta_o)$. Tento model obsahuje neznámý parametr θ_o, o kterém máme apriorní informaci $\theta_o \in \Theta$. Prostor parametrů Θ musí být také opatřen σ-algebrou, budeme ji označovat $\mathcal{Q}$. Takovým parametrem mohou být technické koeficienty, koeficienty ve vztazích mezi proměnnými, rozptyly náhodných chyb obsažených v modelu. Mohou to však také být polynomy do řádu 5, spliny 3 stupně, nebo i funkce mající 3.derivaci nebo nějakou jinou předepsanou vlastnost, např. v modelu obecné regrese, při odhadu hustoty. Parametrem však může být i celé rozdělení náhodných vlivů. Pokud je dimenze Θ konečná, pak se často mluví o parametrickém modelu. V opačném případě se mluví o modelu semiparametrickém.

Model sice závisí na parametru $\theta_o \in \Theta$, ale z praktického hlediska nebo ze zadání úlohy nás nemusí nezajímat přímo parametr θ_o. Často je důležitá pouze jeho „část“ nebo, lépe řečeno, nějaká jeho transformace. Dobře je toto členění vidět na modelu lineární regrese $Y_i = X_i\beta_o + e_i$, kde (X_i, e_i) jsou i.i.d., X_i je nezávislé s e_i a chyby e_i mají nulovou střední hodnotu a konečný nenulový rozptyl. Zde jsou parametrem regresní koeficienty $\beta_o \in \mathbb{R}^k$, rozdělení regresorů $G_o \in \mathcal{G}$ a rozdělení chyb $F_o \in \mathcal{F}$. Pro praktické účely nás však zajímá pouze odhad regresních koeficientů a pro odhad kvality jejich odhadu ještě potřebujeme odhadnout rozptyl chyby. Místo parametru $\theta_o = (\beta_o, G_o, F_o) \in \mathbb{R}^k \times \mathcal{G} \times \mathcal{F}$ nás proto zajímá pouze jeho transformace $\iota(\theta_o) = \left(\beta_o, \int_{\mathbb{R}} e^2 \,\mathsf{d}F_o(e)\right) \in \mathbb{R}^k \times \mathbb{R}_+$.

Uvažujme tedy dále, že je dána funkce $\iota : \Theta \to \Psi$ a pro každé $n \in \mathbb{N}$ hledáme měřitelné zobrazení $\tau_n : \mathbb{X}^n \to \Psi$, které budeme nazývat odhad parametru $\iota(\theta)$. Naším cílem je nalézt odhad tak, aby byl „rozumný“ a případně i v nějakém smyslu „optimální“. Základním požadavkem je, aby odhad v limitě konvergoval ke správné hodnotě parametru modelu, t.j. buď

slabá konzistence $\sim \forall\theta \in \Theta : \tau_n(X_1, X_2, \ldots, X_n) \xrightarrow[n\to+\infty]{\mathrm{P}_\theta} \iota(\theta)$,

nebo

(silná) konzistence $\sim \forall\theta \in \Theta : \tau_n(X_1, X_2, \ldots, X_n) \xrightarrow[n\to+\infty]{\mathrm{P}_\theta - s.j.} \iota(\theta)$.

Pokud chceme testovat nějaké vlastnosti parametru θ_o, například nulovost některých složek, pak musíme ještě pro každé $\theta \in \Theta$ znát (alespoň vhodný odhad) rozdělení $\mathcal{L}_\theta(\tau_n(X_1, X_2, \ldots, X_n) - \iota(\theta))$. Pokud toto neumíme nebo je to nevyčíslitelné, využíváme k testování asymptotické rozdělení tohoto rozdílu, t.j.

existují $0 < \eta_n \nearrow +\infty$ tak, že $\forall\theta \in \Theta : \eta_n\left(\tau_n(X_1, X_2, \ldots, X_n) - \iota(\theta)\right) \xrightarrow[n\to+\infty]{d} \xi(\theta)$,

kde rozdělení $\mathcal{L}_\theta(\xi(\theta))$ je známo a závisí pouze na parametru modelu. Často se jedná o regulární normální rozdělení. Normalizační konstanty η_n se nazývají řád konzistence odhadu a je známo, že závisí na mohutnost množiny Ψ. Přesný vztah i teorie kolem něho nejsou jednoduché; viz ??? , ??? .

Odhady často hledáme pouze mezi funkcemi splňujícími nějakou speciální vlastnost. Budeme tedy uvažovat pouze odhady $\tau_n \in \Gamma_n$, kde $\Gamma_n \subset \{T : \mathbb{X}^n \to \Psi\}$, například uvažujeme pouze lineární kombinace pozorování, nebo pouze nestranné odhady, t.j.

nestrannost $\sim \forall\theta \in \Theta : \mathsf{E}_\theta[T(X_1, X_2, \ldots, X_n)] = \iota(\theta)$.

2. Odhad minimalizující riziko

Naší snahou je nalézt odhad tak, aby byl v nějakém smyslu optimální. V této kapitole vyložíme teorii odhadů minimalizujících riziko. Budeme při tom vycházet ze skript ??? , kde je potřebná teorie pěkně a přehledně vysvětlena. Tuto teorii však aplikujeme na odhad transformovaného parametru.

Kvalitu odhadu měříme ztrátovou funkcí $L : \Psi \times \Psi \to \mathbb{R}_+$, což je vlastně zadaná „vzdálenost“ na Ψ. Od ztrátové funkce kromě nezápornosti ještě vyžadujeme, aby rozlišovala body z Ψ, t.j.

$\forall\, \psi, \varphi \in \Psi : L(\psi, \varphi) = 0 \iff \psi = \varphi.$

Naším úkolem je v nějakém smyslu minimalizovat ztrátu $L(\tau_n(X_1, X_2, \ldots, X_n), \iota(\theta))$ Po odstranění vlivu náhody tedy minimalizujeme rizikovou funkci

$$R(\tau_n, \theta) = \mathsf{E}_\theta[L(\tau_n(X_1, X_2, \ldots, X_n), \iota(\theta))] =$$

$$\int_{\mathbb{X}^n} L(\tau_n(x_1, x_2, \ldots, x_n), \iota(\theta))\, \mathsf{d}\mu_n(x_1, x_2, \ldots, x_n; \theta)\,.$$

V případě i.i.d. pozorování bude riziková funkce mít tvar

$$R(\tau_n, \theta) = \int_{\mathbb{X}^n} L(\tau_n(x_1, x_2, \ldots, x_n), \iota(\theta))\, \mathsf{d}\mu_1(x_1; \theta) \cdot \mathsf{d}\mu_1(x_2; \theta) \cdot \ldots \cdot \mathsf{d}\mu_1(x_n; \theta)\,.$$

Chtěli bychom nalézt odhad $\tau_n \in \Gamma_n$ tak, aby pro každé $\theta \in \Theta$ minimalizoval riziko $R(T, \theta)$ přes všechny $T \in \Gamma_n$. Takovému odhadu říkáme stejnoměrně eficientní odhad v třídě odhadů Γ_n. Takový odhad však obecně nemusí existovat. Je tomu tak například, když Γ_n obsahuje supereficientní odhady, t.j. odhady, které mají pro nějakou hodnotu parametru nulové riziko. Nejjednodušším příkladem supereficientních odhadů jsou konstantní odhady $T_\psi \equiv \psi$ pro $\psi \in \Psi$. Následující tvrzení ukazuje, že stačí, aby Γ_n obsahovala dva konstantní odhady, a již se můžeme s existencí stejnoměrně eficientního odhadu rozloučit.

Tvrzení 1: *Nechť $\mathbb{X} = \mathbb{R}^q$ a $\xi, \eta \in \Theta$ jsou takové, že $\iota(\xi) \neq \iota(\eta)$ a $T_{\iota(\xi)}, T_{\iota(\eta)} \in \Gamma_n$. Když pro každé $A \in \mathcal{X}^n$, $\mathsf{P}_\xi((X_1, X_2, \ldots, X_n) \in A) = 1$ je $\mathsf{P}_\eta((X_1, X_2, \ldots, X_n) \in A)$ > 0, pak stejnoměrně eficientní odhad v třídě odhadů Γ_n neexistuje.*

Důkaz: Předpokládejme, že odhad $\tau_n \in \Gamma_n$ je stejnoměrně eficientní v třídě odhadů Γ_n. Pak ovšem musí platit $R(\tau_n, \xi) \leq R\big(T_{\iota(\xi)}, \xi\big) = 0$ a $R(\tau_n, \eta) \leq R\big(T_{\iota(\eta)}, \eta\big) = 0$.
Odtud $\mathsf{P}_\xi(L(\tau_n(X_1, X_2, \ldots, X_n), \iota(\xi)) = 0) = 1$ a
$\mathsf{P}_\eta(L(\tau_n(X_1, X_2, \ldots, X_n), \iota(\eta)) = 0) = 1$.
Ztrátová funkce rozlišuje body z Ψ a tak $\mathsf{P}_\xi(\tau_n(X_1, X_2, \ldots, X_n) = \iota(\xi)) = 1$ a zároveň také
$\mathsf{P}_\eta(\tau_n(X_1, X_2, \ldots, X_n) = \iota(\eta)) = 1$.
To však ale není možné neboť $\iota(\xi) \neq \iota(\eta)$ a $\tau_n : \mathbb{X}^n \to \Psi$ je měřitelné zobrazení. Předpokládaná vlastnost pak totiž implikuje, že

$$\mathsf{P}_\eta(\tau_n(X_1, X_2, \ldots, X_n) \neq \iota(\eta)) \geq \mathsf{P}_\eta(\tau_n(X_1, X_2, \ldots, X_n) = \iota(\xi)) > 0.$$

Stejnoměrně eficientní odhad tedy nemůže v této třídě existovat.

Q.E.D.

Tento problém se odráží i v lidové slovesnosti:

Otázka: Které hodiny jdou nejpřesněji?
Odpověď: Ty které stojí.
Zdůvodnění: Takové hodiny ukazují dvakrát denně přesný čas!

Máme v podstatě dvě možnosti jak tento nedostatek odstranit. První možností je uvažovat jen takové třídy odhadů, pro které stejnoměrně eficientní odhad existuje. Například třídu lineárních nestranných odhadů při odhadování střední hodnoty i.i.d. posloupnosti, viz ??? .

Druhou možností je zjistit, nebo přibližně odhadnout, frekvenci výskytu parametrů v praxi. To znamená mít k dispozici míru Q na $(\Theta, \mathcal{Q})$ a minimalizovat průměrné riziko $\int_\Theta R(T, \theta)\ \mathsf{d}Q(\theta)$.

3. Extrakce informace ukryté v pozorováních

V pozorováních $X_1, X_2, \ldots, X_n$ je skryta „informace“ o skutečné hodnotě parametru. Otázkou je, jak tuto „informaci“ z výběru extrahovat a jak ji využít pro odhadování.

Cílem je vlastně zredukovat rozměrnost pozorování a místo nich používat pouze nějakou statistiku $T_n : \mathbb{X}^n \to \mathbb{Y}_n$, která by obsahovala stejnou informaci o parametru jako vlastní pozorování a byla, co „nejmenší“. Matematicky je tento požadavek vyjádřen pojmy postačující statistika a úplná statistika.

Statistiku $T_n : \mathbb{X}^n \to \mathbb{Y}_n$ nazveme postačující, jestliže pro každou dvojici parametrů $\theta, \eta \in \Theta$ platí
$\mathcal{L}_\theta(X_1, X_2, \ldots, X_n \,|\, T_n(X_1, X_2, \ldots, X_n)) = \mathcal{L}_\eta(X_1, X_2, \ldots, X_n \,|\, T_n(X_1, \ldots, X_n))$

Statistiku $T_n : \mathbb{X}^n \to \mathbb{Y}_n$ nazveme úplnou, jestliže pro každou měřitelnou funkci $h : \mathbb{Y}_n \to \mathbb{R}$ vlastnost $\forall \theta \in \Theta$ je $\mathsf{E}_\theta[h(T_n(X_1, X_2, \ldots, X_n))] = 0$ implikuje $\forall \theta \in \Theta$ je $\mathsf{P}_\theta(h(T_n(X_1, X_2, \ldots, X_n)) = 0) = 1$.

Pro každý problém existuje postačující statistika, např. $T_n(X_1, X_2, \ldots, X_n) = (X_1, X_2, \ldots, X_n)$, a také vždy existuje úplná statistika, např. $T_n(X_1, X_2, \ldots, X_n) \equiv \psi$ pro nějaké $\psi \in \Psi$. Pro odhadování je však důležité nalézt statistiku, která ma obě tyto vlastnosti. Její důležitost ukazuje následující věta.

Věta 2: *Když $\Psi \subset \mathbb{R}^k$ a pro studovaný problém existuje postačující úplná statistika T_n a nestranný odhad η_n pro $\iota(\theta)$, pak podmíněná střední hodnota*

$$\tau_n(X_1, \ldots, X_n) = \mathsf{E}_\theta[\,\eta_n(X_1, \ldots, X_n) |\, T_n(X_1, \ldots, X_n)]$$

nezávisí na parametru $\theta \in \Theta$ a je stejnoměrně eficientním odhadem pro $\iota(\theta)$ ve třídě všech nestranných odhadů při ztrátové funkci $L(\theta, \rho) = \|\theta - \rho\|^2$.

Důkaz: Obecně je podmíněná střední hodnota funkcí podmínky a tak

$$\tau_n(X_1, \ldots, X_n; \theta) = \mathsf{E}_\theta[\,\eta_n(X_1, \ldots, X_n) |\, T_n(X_1, \ldots, X_n)] \,.$$

Statistika T_n je postačující a proto $\tau_n(X_1, \ldots, X_n; \theta)$ nezávisí na parametru $\theta \in \Theta$ a můžeme psát pouze $\tau_n(X_1, \ldots, X_n)$. Tudíž τ_n je odhad pro $\iota(\theta)$ a je nestranným odhadem neboť je podmíněnou střední hodnotou nestranného odhadu.
Uvažujme t_n nějaký jiný nestranný odhad pro $\iota(\theta)$. Při kvadratické ztrátové funkci pro něj dostáváme rizikovou funkci

$$R(t_n, \theta) = \mathsf{E}_\theta\left[\|t_n(X_1, \ldots, X_n) - \iota(\theta)\|^2\right].$$

Z Jensenovy nerovnosti pro podmíněnou střední hodnotu dostáváme

$$\begin{aligned}R(t_n,\theta) =&\mathsf{E}_\theta\Big[\mathsf{E}_\theta\Big[\,\|t_n(X_1,\ldots,X_n)-\iota(\theta)\|^2\Big|\,T_n\Big]\Big] \geq \\ &\geq\mathsf{E}_\theta\Big[\|\mathsf{E}_\theta[\,t_n(X_1,\ldots,X_n)|\,T_n]-\iota(\theta)\|^2\Big] = R(\mathsf{E}_\theta[\,t_n|\,T_n]\,,\theta)\,.\end{aligned}$$

Postačitelnost statistiky T_n říká, že existuje funkce

$$h(T_n(X_1,\ldots,X_n)) = \tau_n(X_1,\ldots,X_n) - \mathsf{E}_\theta[\,t_n(X_1,\ldots,X_n)|\,T_n(X_1,\ldots,X_n)]\,.$$

Z nestrannosti odhadů η_n, t_n plyne

$$\mathsf{E}_\theta[h(T_n(X_1,\ldots,X_n))] = 0 \;\; \forall\theta\in\Theta.$$

Úplnost statistiky T_n proto dává

$$\tau_n(X_1,\ldots,X_n) = \mathsf{E}_\theta[\,t_n(X_1,\ldots,X_n)|\,T_n(X_1,\ldots,X_n)] \;\; s.j. \;\; \forall\theta\in\Theta.$$

Tudíž zjišťujeme, že $\forall\theta\in\Theta\;R(t_n,\theta)\geq R(\tau_n,\theta)$.
Jinými slovy, odhad τ_n je stejnoměrně eficientní v třídě všech nestranných odhadů.

Q.E.D.

Pokud existuje dominující σ-konečná míra pro sledovanou úlohu, pak lze z tvaru hustoty vyčíst některé postačující statistiky.

Tvrzení 3: *Nechť μ_n je σ-konečná míra na $\mathcal{X}^n$ a $(X_1,\ldots,X_n)$ má hustotu $f_n(x_1,...,x_n;\theta)$ vzhledem k μ_n. Když $f_n(x_1,\ldots,x_n;\theta) = g_n(T_n(x_1,\ldots,x_n);\theta)h_n(x_1,\ldots,x_n)$, kde $g_n(.\,;\theta):\mathbb{Y}_n\to\mathbb{R}_+$, $T_n:\mathbb{X}^n\to\mathbb{Y}_n$ a $h_n:\mathbb{X}^n\to\mathbb{R}_+$ jsou měřitelné funkce, a $\int_{\mathbb{X}} h_n(x)\,\mathsf{d}\mu_n(x) < +\infty$, potom T_n je postačující statistika.*

Důkaz: Pro každé $A\in\mathcal{X}^n$, $B\in\mathcal{Y}_n$ definujme

$$\rho(B,A) = \int_{[T_n(x)\in B, x\in A]} h_n(x)\,\mathsf{d}\mu_n(x).$$

Díky předpokladu $\int_{\mathbb{X}} h_n(x)\,\mathsf{d}\mu_n(x) < +\infty$ jsou $\rho(.\,,A)$ konečné míry.
Evidentně pro každé $A\in\mathcal{X}^n$ je míra $\rho(.\,,A)$ absolutně spojitá vzhledem k míře $\rho(.\,,\mathbb{X})$.
Tudíž podle Radon-Nikodýmovi věty existuje $s:\mathbb{Y}_n\times\mathcal{X}^n\to\mathbb{R}_+$ tak, že pro každé $A\in\mathcal{X}^n$, $B\in\mathcal{Y}_n$ platí $\rho(B,A) = \int_B s(y,A)\rho(\,\mathsf{d}y,\mathbb{X})$.
Pro $A\in\mathcal{X}^n$, $B\in\mathcal{Y}_n$ platí

$$\int_{[T_n(X)\in B]} s(T_n(X),A)\,\mathsf{dP}_\theta = \int_{[T_n(x)\in B]} s(T_n(x),A)g_n(T_n(x);\theta)h_n(x)\,\mathsf{d}\mu_n(x) =$$

$$= \int_{[y\in B]} s(y,A)g_n(y;\theta)\rho(\,\mathsf{d}y,\mathbb{X}) = \int_{[y\in B]} g_n(y;\theta)\rho(\,\mathsf{d}y,A) = \int_{[T_n(X)\in B]} \mathbb{I}_{[X\in A]}\,\mathsf{dP}_\theta.$$

Ověřili jsme, že $\mathsf{P}_\theta(\,X\in A|\,T_n(X)) = s(T_n(X),A)$.
Tudíž rozdělení vektoru pozorování podmíněné statistikou T_n nezávisí na parametru $\theta\in\Theta$ a T_n je postačující statistikou.

Q.E.D.

Ještě si uveďme příklad.

Příklad 1: Nechť X_i, $i = 1, 2, \ldots, n$ jsou i.i.d. n.v. s normálním rozdělením $\Phi\left(.\,; \mu, \sigma^2\right)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$.

Pak vektor $(X_1, \ldots, X_n)$ má sdruženou hustotu

$$\begin{aligned} f_n(x; \mu, \sigma^2) &= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\} \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}x_i^2 - 2\mu\sum_{i=1}^{n}x_i + n\mu^2\right)\right\}. \end{aligned}$$

Podle předchozího tvrzení je statistika $(\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2)$ postačující. Zároveň lze také dokázat, že je tato statistika úplná.

Snadno se přesvědčíme, že $\left(X_1, \frac{(X_1-X_2)^2}{4}\right)$ je nestranný odhad parametru (μ, σ^2). Tudíž

$$\mathsf{E}_{\mu,\sigma^2}\left[\left(X_1, \frac{(X_1-X_2)^2}{4}\right)\middle| \sum_{i=1}^{n}X_i, \sum_{i=1}^{n}X_i^2\right] =$$

$$= \left(\frac{1}{n}\sum_{i=1}^{n}X_i, \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \frac{1}{n}\sum_{j=1}^{n}X_j\right)^2\right) \quad s.j.$$

je stejnoměrně eficientní odhad parametru (μ, σ^2) ve třídě nestranných odhadů a při ztrátové funkci $L\big((\mu_1, \sigma_1^2), (\mu_2, \sigma_2^2)\big) = (\mu_1 - \mu_2)^2 + (\sigma_1^2 - \sigma_2^2)^2$.

$\triangle$

Teorie postačujících a úplných statistik je vyložena také v kapitole XV.5, ??? .

4. Maximálně věrohodné odhady

Nyní již přistupujeme k obecným návodům pro hledání odhadů. Zde již není obecné kritérium, které chceme minimalizovat. Zde přistupujeme rovnou k datům a naším cílem je minimalizovat nějakou ztrátu rovnou pro ně.

Metoda maximální věrohodnosti je založena na rozdělení pozorovaných veličin na část s rozdělením spojitým vzhledem k Lebesguevě míře a na zbytek nabývající pouze spočetně hodnot. V jiných případech tuto ideu nelze aplikovat.

Předpokládejme, že pozorování lze zapsat ve tvaru $X_i = (Z_i, D_i)$, $i \in \{1, \ldots, n\}$, kde $Z_i \in \mathbb{R}^k$ a $D_i \in \mathcal{D}$ pro nějakou nejvýše spočetnou množinu $\mathcal{D}$. Dále pro každý parametr $\theta \in \Theta$ předpokládáme existenci hustoty $f_n(z_1, d_1; z_2, d_2; \ldots; z_n, d_n; \theta)$ vzhledem k součinové míře $(\lambda_k \otimes \check{c}_{\mathcal{D}})^{\otimes n}$. Poznamenejme, že čítací míra $\check{c}_A$ je definována pro každou množinu $B \subset A$, jako počet bodů obsažených v množině B, t.j. $\check{c}_A(B) = \mathsf{card}(B)$.

Principem ML-odhadu (Maximal Likelihood) je „maximalizace“ věrohodnosti $f_n(X_1, X_2, \ldots, X_n; \theta)$. Odhad konstruujeme následovně.

Odhadneme původní parametr

$$\widehat{\theta}_n \in \underset{\theta\in\Theta}{\mathsf{argmax}}\, f_n(X_1, X_2, \ldots, X_n; \theta) = \underset{\theta\in\Theta}{\mathsf{argmax}}\, f_n(Z_1, D_1; Z_2, D_2; \ldots; Z_n, D_n; \theta)$$

a položíme $\tau_n = \iota\left(\widehat{\theta}_n\right)$.

Tento postup však není vždy výpočetně realizovatelný. Často se proto volí modifikace tohoto postupu využívající odhadu části parametrů, t.j. máme k dispozici informaci (odhad), že hledaný parametr leží v množině $\widehat{\Theta}_n \subset \Theta$. V tomto případě odhadneme původní parametr

$$\widehat{\theta}_n \in \underset{\theta\in\widehat{\Theta}_n}{\operatorname{argmax}} f_n(X_1, X_2, \ldots, X_n; \theta) = \underset{\theta\in\widehat{\Theta}_n}{\operatorname{argmax}} f_n(Z_1, D_1; Z_2, D_2; \ldots; Z_n, D_n; \theta)$$

a položíme $\tau_n = \iota\left(\widehat{\theta}_n\right)$.

Na závěr kapitoly uveďme ještě příklady.

Příklad 2: Nechť X_i, $i = 1, 2, \ldots, n$ jsou i.i.d. n.v. s normálním rozdělením $\Phi\left(.\,; \mu, \sigma^2\right)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$.

Pak vektor $(X_1, \ldots, X_n)$ má sdruženou hustotu

$$f_n(x_1, x_2, \ldots, x_n; \mu, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\} =$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}\left(x_i - \frac{1}{n}\sum_{i=1}^{n}x_i\right)^2 + n\left(\frac{1}{n}\sum_{i=1}^{n}x_i - \mu\right)^2\right)\right\}.$$

Při hledání ML-odhadu budeme maximalizovat věrohodnostní funkci

$$f_n(X_1, X_2, \ldots, X_n; \mu, \sigma^2) =$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}\left(X_i - \frac{1}{n}\sum_{i=1}^{n}X_i\right)^2 + n\left(\frac{1}{n}\sum_{i=1}^{n}X_i - \mu\right)^2\right)\right\}.$$

Nejprve budeme odhadovat oba parametry naráz. To znamená $\Theta = \Psi = \mathbb{R} \times \mathbb{R}_+$, $\iota\left(\mu, \sigma^2\right) = (\mu, \sigma^2)$. Minimalizací věrohodnostní funkce získáme ML-odhad

$$\widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n}X_i,\ \widehat{\sigma^2}_n = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \frac{1}{n}\sum_{j=1}^{n}X_j\right)^2.$$

Povšimněme si, že $\widehat{\mu}_n$ je stejnoměrně eficientní odhad parametru μ ve třídě nestranných odhadů při L_2-kritériu, jak je ukázáno v předchozí kapitole. Ale $\widehat{\sigma^2}_n$ není nestranným odhadem pro parametr σ^2. Víme, že $\frac{n}{n-1}\widehat{\sigma^2}_n$ je stejnoměrně eficientní odhad parametru σ^2 ve třídě nestranných odhadů při L_2-kritériu, jak je ukázáno v předchozí kapitole.

Odhadujme nyní pouze střední hodnotu μ.

- Když je rozptyl σ^2 známý, pak $\Theta = \Psi = \mathbb{R}$, $\iota(\mu) = \mu$ a ML-odhad vyjde $\widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n}X_i$.
- Když je rozptyl σ^2 neznámý, pak $\Theta = \mathbb{R} \times \mathbb{R}_+$, $\Psi = \mathbb{R}$, $\iota\left(\mu, \sigma^2\right) = \mu$ a ML-odhad opět vyjde $\widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n}X_i$.
- Když je rozptyl σ^2 neznámý, ale máme jeho odhad s_n^2, pak $\Theta = \mathbb{R} \times \mathbb{R}_+$, $\Psi = \mathbb{R}$ a $\iota\left(\mu, \sigma^2\right) = \mu$. Použijeme modifikovaný postup s $\widehat{\Theta}_n = \mathbb{R} \times \left\{s_n^2\right\}$ a ML-odhad vyjde opět $\widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n}X_i$.

Odhadujme pouze rozptyl σ^2.

- Když je střední hodnota μ známá, pak $\Theta = \Psi = \mathbb{R}_+$, $\iota\big(\sigma^2\big) = \sigma^2$ a ML-odhad vyjde
$$\widehat{\sigma^2}_n = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \mu\right)^2 .$$
- Když je střední hodnota μ neznámá, pak $\Theta = \mathbb{R} \times \mathbb{R}_+$, $\Psi = \mathbb{R}_+$, $\iota\big(\mu, \sigma^2\big) = \sigma^2$ a ML-odhad vyjde
$$\widehat{\sigma^2}_n = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \frac{1}{n}\sum_{j=1}^{n} X_j\right)^2 .$$
- Když je střední hodnota μ neznámá, ale máme k dispozici její odhad m_n, pak $\Theta = \mathbb{R} \times \mathbb{R}_+$, $\Psi = \mathbb{R}_+$ a $\iota\big(\mu, \sigma^2\big) = \sigma^2$. Použijeme modifikovaný postup s $\widehat{\Theta}_n = \{m_n\} \times \mathbb{R}_+$ a ML-odhad vyjde
$$\widehat{\sigma^2}_n = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - m_n\right)^2 .$$

$\triangle$

Příklad 3: Regrese 0-1 náhodných veličin

Pozorujeme dvojice (Y_i, x_i), $i \in \mathbb{N}$ svázané 0-1 regresí
$$Y_i = F\left(x_i^T \beta\right) + e_i, \ \ i \in \mathbb{N},$$
kde $Y_i \in \{0, 1\}$, $i \in \mathbb{N}$ jsou náhodné veličiny, $x_i \in \mathbb{R}^k$, $i \in \mathbb{N}$ jsou nenáhodné vektory, $e_i \in \mathbb{R}^k$, $i \in \mathbb{N}$ jsou náhodné chyby, které jsou nezávislé a mají nulovou střední hodnotu, $\beta \in \mathbb{R}^k$ neznámé parametry a F je známá d.f. Pokud $F = \Phi$, pak mluvíme o probit modelu. Pokud F je d.f. logistického rozdělení, t.j. $F(t) = \frac{1}{1+\mathrm{e}^{-t}}$ $\forall\, t \in \mathbb{R}$, pak mluvíme o logit modelu.

Tento model má jediný parametr a to $\beta \in \mathbb{R}^k$. Na základě n pozorování sestavíme věrohodnostní funkci
$$L(\beta) = \prod_{i=1}^{n}\left(1 - F\left(x_i'\beta\right)\right)^{(1-Y_i)} F\left(x_i'\beta\right)^{Y_i} .$$
Maximalizací věrohodnostní funkce L získáme ML-odhad pro parametr β. Diskusi o jeho výpočtu a vlastnostech můžeme nalézt v kapitole 16.4 v ???

$\triangle$

Příklad 4: Tobit model

Povšimněme si nyní modelu z článku ??? a navrženého odhadu z článků ??? , ??? . Pozorujeme trojice (Y_i, x_i, δ_i) svázané modelem cenzorované regrese:
$$\begin{aligned} Y_i &= x_i'\beta + e_i \text{ když } x_i'\beta + e_i \geq 0, \\ &= 0 \quad \text{jinak}, \end{aligned} \qquad \delta_i = \mathbb{I}_{[x_i'\beta + e_i \geq 0]}, \ \ i \in \mathbb{N},$$
kde $x_i \in \mathbb{R}^k$, $i \in \mathbb{N}$ jsou nenáhodná a e_i, $i \in \mathbb{N}$ jsou i.i.d. s rozdělením $F(.;\theta)$, $\theta \in \Theta$, $\mathsf{E}[e_i] = 0$.

Tento model má parametr $(\beta, \theta) \in \mathbb{R}^k \times \Theta$. Máme-li kdispozici n pozorování, můžeme se pokusit odhadnout parametr β pouze na základě těch pozorování, u nichž víme, že jsme regresi skutečně pozorovali. Máme tedy náhodnou množinu indexů $N^* = \{i \in \{1, 2, \ldots, n\} : \delta_i = 1\}$ a regresi

$$Y_i = x_i'\beta + e_i \text{ pro } i \in \mathbb{N}^*.$$

Když pro odhad parametru β použijeme nějaký odhad, který je nestranný (případně asymptoticky nestranný) v běžném modelu regrese, získáme vychýlený odhad. Je to proto, že střední hodnota Y_i za podmínky, že pozorujeme regresi, není regresní přímka $x_i^T\beta$. Spočtěme si tuto podmíněnou střední hodnotu při pevném $\theta \in \Theta$ přesně:

$$\begin{aligned} \mathsf{E}[Y_i|\, \delta_i = 1] &= \mathsf{E}\big[\,(x_i'\beta + e_i)\,\mathbb{I}_{[x_i'\beta+e_i\geq 0]}\big|\, x_i'\beta + e_i \geq 0\big] = \\ &= \mathsf{E}[\,(x_i'\beta + e_i)|\, x_i'\beta + e_i \geq 0]\,\mathbb{I}_{[x_i'\beta+e_i\geq 0]} = \\ &= \left(x_i'\beta + \frac{1}{\mathsf{P}(x_i'\beta + e_i \geq 0)}\int_{-x_i\beta}^{+\infty} e\,\mathsf{d}F(e;\theta)\right)\mathbb{I}_{[x_i'\beta+e_i\geq 0]} = \\ &= \left(x_i'\beta + \frac{1}{1 - F(-x_i'\beta;\theta)}\int_{-x_i\beta}^{+\infty} e\,\mathsf{d}F(e;\theta)\right)\mathbb{I}_{[x_i'\beta+e_i\geq 0]}. \end{aligned}$$

Pokud by pro každé $i \in \mathbb{N}$ bylo $\int\limits_{-x_i\beta}^{+\infty} e\,\mathsf{d}F(e;\theta) = 0$, pak by odhad založený pouze na očištěném výběru byl konzistentní. To však je možné pouze v případě $F(-x_i'\beta;\theta) = \mathsf{P}(x_i'\beta + e_i < 0) = 0$ pro všechna $i \in \mathbb{N}$. To znamená, k cenzorování s.j. nedojde a my stále pozorujeme regresi.

Pro odhad regrese založeného na očištěném souboru musíme mít k dispozici odhad druhého členu v podmíněné střední hodnotě. Pro normální rozdělení s neznámým rozptylem je situace jednodušší, ale hlavně názornější.

Uvažujme tedy $\Theta = \mathbb{R}_+$ a $F(e;\sigma) = \Phi\left(\frac{e}{\sigma}\right)$, kde Φ označuje d.f. standardního normálního rozdělení a φ je jeho hustota. V tomto případě dostáváme

$$\int_{-x_i\beta}^{+\infty} e\,\mathsf{d}F(e;\sigma) = \sigma\int_{\frac{-x_i\beta}{\sigma}}^{+\infty} e\,\mathsf{d}\Phi\,(e) = \sigma\int_{\frac{-x_i\beta}{\sigma}}^{+\infty} e\varphi\,(e)\,\mathsf{d}e = \sigma\varphi\left(\frac{-x_i\beta}{\sigma}\right) = \sigma\varphi\left(\frac{x_i\beta}{\sigma}\right)$$

$$1 - F(-x_i'\beta;\theta) = 1 - \Phi\left(\frac{-x_i'\beta}{\sigma}\right) = \Phi\left(\frac{x_i'\beta}{\sigma}\right).$$

Tudíž

$$\mathsf{E}[Y_i|\, x_i'\beta + e_i \geq 0] = \left(x_i'\beta + \sigma\frac{\varphi\left(\frac{x_i'\beta}{\sigma}\right)}{\Phi\left(\frac{x_i'\beta}{\sigma}\right)}\mathbb{I}_{[x_i'\beta+e_i\geq 0]}\right).$$

Pokud by jsme znali podíly $\frac{\varphi\left(\frac{x_i'\beta}{\sigma}\right)}{\Phi\left(\frac{x_i'\beta}{\sigma}\right)}$, nebo měli alespoň konzistentní odhady, pak by jsme dokázali konzistentně odhadnout oba parametry modelu, t.j. β, $\sigma > 0$, třeba pomocí OLS-odhadu.

Pozorování δ_i, $i = 1, 2, \ldots, n$ nabývají pouze hodnot 0 nebo 1 a splňují

$$\mathsf{E}[\delta_i] = \mathsf{P}(x_i'\beta + e_i \geq 0) = 1 - \Phi\left(\frac{-x_i'\beta}{\sigma}\right) = \Phi\left(\frac{x_i'\beta}{\sigma}\right).$$

Jsou tedy řízeny probit modelem. Na základě teorie probit modelu získáme konzistentní odhad $\widehat{\lambda}_n$ pro podíl $\frac{\beta}{\sigma}$; třeba ML-odhadem viz předchozí příklad, jiné možnosti jsou uvedeny v kapitole 16 v ??? .

Dosazením do předchozího získáme

$$\mathsf{E}[Y_i|\, x_i'\beta + e_i \geq 0] = \left(x_i'\beta + \sigma \frac{\varphi\left(x_i'\widehat{\lambda}_n\right)}{\Phi\left(x_i'\widehat{\lambda}_n\right)} \right) \mathbb{I}_{[x_i'\beta + e_i \geq 0]}.$$

Nyní z očištěného výběru spočteme odhad (třeba OLS-odhad) pro parametry regrese

$$Y_i = x_i'\beta + \sigma \frac{\varphi\left(x_i'\widehat{\lambda}_n\right)}{\Phi\left(x_i'\widehat{\lambda}_n\right)} + e_i,\ i \in N^*.$$

Konzistence a asymptotická normalita je ukázána v článcích ??? , ??? .

Druhou možností je sestavit věrohodnostní funkci

$$L(\beta, \sigma) = \prod_{i=1}^{n} \left(1 - \Phi\left(\frac{x_i'\beta}{\sigma} \right) \right)^{1-\delta_i} \varphi\left(Y_i - \frac{x_i'\beta}{\sigma} \right)^{\delta_i}.$$

Odhad parametrů pak získáme metodou maximální věrohodnosti, t.j. maximalizací věrohodnostní funkce L. Konzistence a asymptotická normalita tohoto odhadu je ukázána v článku ??? .

Tento postup je tedy teoreticky dobře zdůvodněn. Potíže jsou však s nalezením maxima věrohodnostní funkce. Funkce je velmi nelineární a dělá problémy při numerickém řešení. V článku ??? je proto navržen jiný odhad a je zde také ukázána jeho konzistence a asymptotická linearita. Tento odhad se dá zlepšit jedním, případně více, kroky Newtonovy metody. Konstrukce tohoto odhadu je také uvedena v kapitole 16.6.2 v ??? .

$\triangle$

5. Odhady v modelu lineární regrese

Věnujme se nyní odhadování parametrů v modelu lineární regrese. Zde pozorujeme dvojice

$$(Y_1, X_1), (Y_2, X_2), \ldots, (Y_n, X_n),\ Y_i \in \mathbb{R}, X_i \in \mathbb{R}^k,$$

které jsou svázány předpisem

$$Y_i = X_i\beta_o + \varepsilon_i,\ \varepsilon_i \in \mathbb{R},\ \forall i = 1, 2, \ldots, n.$$

Parametry modelu jsou $\beta_o \in \mathbb{R}^k$ a sdružené rozdělení $\mathcal{L}((X_1, \varepsilon_1), \ldots, (X_n, \varepsilon_n))$.

Označíme-li $Y = (Y_1, Y_2, \ldots, Y_n)'$, $X = (X_1', X_2', \ldots, X_n')'$ a $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$, pak lze model zkráceně zapsat

$$Y = X\beta_o + \varepsilon.$$

Aby byl model dobře určen musíme učinit některé předpoklady.

Předpoklad: Matice X má plnou sloupcovou hodnost.

Jinak by existovalo $\delta \in \mathbb{R}^k$ tak, že $X\delta = 0$. Pak ovšem regresní koeficienty nejsou určeny jednoznačně, neboť $Y = X\beta_o + \varepsilon = X(\beta_o + \delta) + \varepsilon$.

Předpoklad: Chyby musí být nějakým způsobem „centrovány“.

V opačném případě by parametry modelu nebyly jednoznačně určeny. K chybám by se jinak dal přičíst libovolný vektor z lineárního obalu řádků matice X a dostali bychom stejný model se změněnými parametry: $Y = X\beta_o + \varepsilon = X(\beta_o - v) + (\varepsilon + Xv)$. Jak takové „centrování“ vypadá, uvidíme později u jednotlivých odhadů.

Předpoklad: Chyby mají nesingulární rozdělením; t.j. neexistuje $a \in \mathbb{R}^n$, $a \neq \mathbf{0}$ takové, aby $a^T\varepsilon$ byla nenáhodná konstanta.

Singulárnost rozdělení znamená, že soustava obsahuje deterministickou závislost. Tuto je třeba nalézt a používat jako omezující podmínku na regresní koeficienty.

Předpoklad: Chyby nejsou ovlivněny regresory. To znamená ε a X jsou nezávislé nebo alespoň nekorelované.

Matice X bývá často nenáhodná a při některých experimentech lze její hodnoty předem nastavit, nebo řídit.

Naším cílem je odhadnout parametr β_o. Popřípadě ještě odhadnout nějakou charakteristiku rozdělení tohoto odhadu, například varianční matici.

Odhad založíme na reziduích $e_i(b) = Y_i - X_i b$ pro každé $i \in \{1, 2, \ldots, n\}$, $b \in \mathbb{R}^k$. Rezidua můžeme vyjádřit také jako $e_i(b) = \varepsilon_i + X_i(\beta_o - b)$, což je výhodné pro teoretické vyšetření asymptotického chování.

Proberme nyní několik nejpoužívanějších postupů odvození odhadů.

5.1. OLS-odhad (Ordinary Least Squares)

OLS-odhad je založen na minimalizaci součtu čtverců reziduí, t.j. hledá se minimum

$$\widehat{\beta}_n \in \underset{b \in \mathbb{R}^k}{\mathsf{argmin}} \sum_{i=1}^{n} e_i(b)^2 = \underset{b \in \mathbb{R}^k}{\mathsf{argmin}} \sum_{i=1}^{n} (Y_i - X_i b)^2 .$$

Pokud je matice X plné hodnosti dostaneme $\widehat{\beta}_n = \left(X^T X\right)^{-1} X^T Y$. OLS-odhad je pak možno zapsat ve tvaru

$$\widehat{\beta}_n = \left(X^T X\right)^{-1} X^T Y = \left(X^T X\right)^{-1} X^T (X\beta_o + \varepsilon) = \beta_o + \left(X^T X\right)^{-1} X^T \varepsilon.$$

Z tohoto rozpisu okamžitě vyplývá nestrannost a silná i slabá konzistence.

Věta 4: *Když X, ε jsou nezávislé, $\mathsf{E}[\varepsilon] = \mathbf{0}$ a $\mathsf{E}\left[\left(X^T X\right)^{-1} X^T\right]$ existuje, pak $\widehat{\beta}_n$ je nestranným odhadem parametru β_o.*

Věta 5: *Když $\frac{1}{n} X^T X \xrightarrow[n \to +\infty]{s.j.} Q$, kde Q je regulární matice, a $\frac{1}{n} X^T \varepsilon \xrightarrow[n \to +\infty]{s.j.} \mathbf{0}$, potom $\widehat{\beta}_n \xrightarrow[n \to +\infty]{s.j.} \beta_o$.*

Věta 6: *Když $\frac{1}{n} X^T X \xrightarrow[n \to +\infty]{p} Q$, kde Q je regulární matice, a $\frac{1}{n} X^T \varepsilon \xrightarrow[n \to +\infty]{p} \mathbf{0}$, potom $\widehat{\beta}_n \xrightarrow[n \to +\infty]{p} \beta_o$.*

Pro OLS-odhad je tedy přirozené chtít, aby chyby měli nulovou střední hodnotu. To je vlastně to avizované „centrování“ chyb modelu.

5.2. M-odhady v modelu lineární regrese

V této kapitole si stručně vysvětlíme, co je to M-odhad. Uvedeme také jeho základní vlastnosti. Podrobnou informaci o M-odhadech je možno nalézt v monografii ??? .

Je dána funkce $\rho : \mathbb{R} \to \mathbb{R}_+$. Pro každé $b \in \mathbb{R}^k$ definujeme

$$M_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} \rho(e_i(b))$$

. Pak M-odhadem nazveme každé řešení minimalizace této funkce, t.j.

$$\widehat{\beta}_n \in \operatorname*{argmin}_{b \in \mathbb{R}^k} M_n(\beta) .$$

Aby byl odhad „rozumný“ musí být splněny některé předpoklady.

Předpoklad: Existuje nenáhodná funkce $M : \mathbb{R}^k \to \mathbb{R}_+$ taková, že pro každé $b \in \mathbb{R}^k$ platí

$$\mathsf{E}[M_n(b) - M_n(\beta_o)] = \frac{1}{n} \sum_{i=1}^{n} \mathsf{E}[\rho(e_i(b)) - \rho(\varepsilon_i)] \xrightarrow[n \to +\infty]{} M(b) ,$$

a pro každé $D \in \mathbb{R}$ platí

$$\sup_{\|b\| \le D} |M_n(b) - M_n(\beta_o) - M(b)| \xrightarrow[n \to +\infty]{s.j.} 0.$$

Funkce M má jediné globální minimum v bodě $\tilde{\beta} \in \mathbb{R}^k$, t.j. $\{\tilde{\beta}\} = \operatorname*{argmin}_{b \in \mathbb{R}^k} M(b)$.

Za celkem mírných předpokladů lze dokázat konvergenci M-odhadu k bodu $\tilde{\beta}$.

Věta 7: *Když ρ je spojitá funkce na $\mathbb{R}^k$ a existuje kompakt $K \subset \mathbb{R}^k$, $\Delta > 0$ a $n_o \in \mathbb{N}$ tak, že pro každé $n \ge n_o$ je $\inf_{b \notin K} M_n(b) - M_n(\beta_0) \ge M\left(\tilde{\beta}\right) + \Delta$, potom pro každé $n \ge n_o$ odhad $\widehat{\beta}_n$ existuje a platí $\widehat{\beta}_n \xrightarrow[n \to +\infty]{s.j.} \tilde{\beta}$.*

Tudíž, aby jsme odhadovali správně, musí být $\beta_o = \tilde{\beta}$. V jiném případě bude náš odhad systematicky vychýlen. Připojme proto další dva předpoklady, které nám tuto rovnost zaručí.

Předpoklad: Náhodné veličiny X a ε jsou nezávislé.

Předpoklad: Pro každé $i \in \{1, 2, \ldots, n\}$ a pro každé $u \in \mathbb{R}$ platí $\mathsf{E}[\rho(\varepsilon_i + u) - \rho(\varepsilon_i)] \ge 0$.

Pak totiž platí

$$\begin{aligned}
\mathsf{E}[M_n(b) - M_n(\beta_o)] &= \frac{1}{n} \sum_{i=1}^{n} \mathsf{E}[\rho(\varepsilon_i + X_i(\beta_o - b)) - \rho(\varepsilon_i)] = \\
&= \frac{1}{n} \sum_{i=1}^{n} \mathsf{E}[\mathsf{E}[\,\rho(\varepsilon_i + X_i(\beta_o - b)) - \rho(\varepsilon_i) |\, X_i]] = \\
&= \frac{1}{n} \sum_{i=1}^{n} \mathsf{E}[R_i(X_i(\beta_o - b))] \ge 0,
\end{aligned}$$

kde $R_i(u) = \mathsf{E}[\rho(\varepsilon_i + u) - \rho(\varepsilon_i)]$.

Limitním přechodem zjistíme, že β_o je globálním minimem funkce M. Z předpokládané jednoznačnosti tohoto minima okamžitě plyne, že $\beta_o = \tilde{\beta}$.

Uvědomme si, že druhá podmínka je vlastně tím „centrováním" chyb, které jsme inzerovali na začátku kapitoly o modelu lineární regrese.

Nyní si povšimněme, co tato podmínka znamená pro některé konkrétní volby funkce ρ.

Lemma 8: *Nechť existuje $\psi : \mathbb{R} \to \mathbb{R}$ taková, že pro každé $u, v \in \mathbb{R}$ platí $\rho(u) - \rho(v) = \int_v^u \psi(t)\,\mathsf{d}t$ a nechť existuje $\delta > 0$ tak, že $\sup_{|t|\le\delta} \mathsf{E}[|\psi(\varepsilon + t)|] < +\infty$. Potom*

$$\limsup_{u\to 0+} \frac{1}{u}\left(\mathsf{E}[\rho(\varepsilon_i + u) - \rho(\varepsilon_i)]\right) \le \limsup_{u\to 0+} \mathsf{E}[\psi(\varepsilon_i + u)]\,,$$
$$\liminf_{u\to 0+} \frac{1}{u}\left(\mathsf{E}[\rho(\varepsilon_i + u) - \rho(\varepsilon_i)]\right) \ge \liminf_{u\to 0+} \mathsf{E}[\psi(\varepsilon_i + u)]\,,$$
$$\limsup_{u\to 0+} \frac{1}{u}\left(\mathsf{E}[\rho(\varepsilon_i - u) - \rho(\varepsilon_i)]\right) \le -\liminf_{u\to 0+} \mathsf{E}[\psi(\varepsilon_i - u)]\,,$$
$$\liminf_{u\to 0+} \frac{1}{u}\left(\mathsf{E}[\rho(\varepsilon_i - u) - \rho(\varepsilon_i)]\right) \ge -\limsup_{u\to 0+} \mathsf{E}[\psi(\varepsilon_i - u)]\,.$$

Důkaz: Pro $i \in \{1, 2, \ldots, n\}$ máme rozpis

$$\mathsf{E}[\rho(\varepsilon_i + u) - \rho(\varepsilon_i)] = \mathsf{E}\left[\int_0^u \psi(\varepsilon_i + t)\,\mathsf{d}t\right].$$

Předpoklad $\sup_{|t|\le\delta} \mathsf{E}[|\psi(\varepsilon + t)|] < +\infty$ umožňuje pro $|u| \le \delta$ zaměnit pořadí integrace:

$$\mathsf{E}[\rho(\varepsilon_i + u) - \rho(\varepsilon_i)] = \int_0^u \mathsf{E}[\psi(\varepsilon_i + t)]\,\mathsf{d}t.$$

Limitní přechodem pro integrál vydělený délkou u dostáváme

$$\limsup_{u\to 0+} \frac{1}{u}\left(\mathsf{E}[\rho(\varepsilon_i + u) - \rho(\varepsilon_i)]\right) = \limsup_{u\to 0+} \frac{1}{u}\int_0^u \mathsf{E}[\psi(\varepsilon_i + t)]\,\mathsf{d}t \le \limsup_{u\to 0+} \mathsf{E}[\psi(\varepsilon_i + u)]$$

Obdobně získáme i zbylé nerovnosti

$$\liminf_{u\to 0+} \frac{1}{u}\left(\mathsf{E}[\rho(\varepsilon_i + u) - \rho(\varepsilon_i)]\right) \ge \liminf_{u\to 0+} \mathsf{E}[\psi(\varepsilon_i + u)]\,,$$
$$\limsup_{u\to 0+} \frac{1}{u}\left(\mathsf{E}[\rho(\varepsilon_i - u) - \rho(\varepsilon_i)]\right) \le -\liminf_{u\to 0+} \mathsf{E}[\psi(\varepsilon_i - u)]\,,$$
$$\liminf_{u\to 0+} \frac{1}{u}\left(\mathsf{E}[\rho(\varepsilon_i - u) - \rho(\varepsilon_i)]\right) \ge -\limsup_{u\to 0+} \mathsf{E}[\psi(\varepsilon_i - u)]$$

Q.E.D.

Tvrzení 9: *Nechť jsou splněny předpoklady lemmatu ??? a nechť pro každé $i \in \{1, 2, \ldots, n\}$ a pro každé $u \in \mathbb{R}$ platí $\mathsf{E}[\rho(\varepsilon_i + u) - \rho(\varepsilon_i)] \ge 0$, potom pro každé $i \in \{1, 2, \ldots, n\}$ je*

$$\limsup_{u\to 0+} \mathsf{E}[\psi(\varepsilon_i + u)] \ge 0,\ \liminf_{u\to 0+} \mathsf{E}[\psi(\varepsilon_i - u)] \le 0.$$

Důkaz: Tvrzení je přímým důsledkem lemmatu ??? .

Q.E.D.

Tvrzení 10: *Nechť jsou splněny předpoklady lemmatu ??? , funkce ρ je konvexní a nechť pro každé $i \in \{1, 2, \ldots, n\}$ je*

$$\liminf_{u \to 0+} \mathsf{E}[\psi(\varepsilon_i + u)] \geq 0, \ \limsup_{u \to 0+} \mathsf{E}[\psi(\varepsilon_i - u)] \leq 0$$

potom pro každé $i \in \{1, 2, \ldots, n\}$ a pro každé $u \in \mathbb{R}$ platí $\mathsf{E}[\rho(\varepsilon_i + u) - \rho(\varepsilon_i)] \geq 0$.

Důkaz: Když ρ je konvexní, pak je konvexní i střední hodnota $\mathsf{E}[\rho(\varepsilon_i + u) - \rho(\varepsilon_i)]$, jako funkce proměnné u. Z této konvexity a z lemmatu ??? již plyne vlastní tvrzení.

Q.E.D.

1) <u>OLS-odhad</u>

OLS-odhad je speciálním případem M-odhadu při volbě $\rho(t) = t^2$. Tato funkce je konvexní a má hustotu $\psi(t) = 2t$. Pak podle tvrzení ??? , je vhodné předpokládat $\mathsf{E}[\varepsilon_i] = 0$ pro každé $i \in \{1, 2, \ldots, n\}$.

2) <u>L_1-odhad</u>

Zde $\rho(t) = |t|$. Tato funkce je konvexní a má hustotu $\psi(t) = \begin{cases} 1 & pro\text{t¿0} \\ 0 & pro\text{t=0} \\ -1 & pro\text{t¡0} \end{cases}$ a tak, podle tvrzení ??? , je vhodné předpokládat $\mathsf{E}[\psi(\varepsilon_i+)] = \mathsf{P}_\theta(\varepsilon \geq 0) - \mathsf{P}_\theta(\varepsilon < 0) \geq 0$, $\mathsf{E}[\psi(\varepsilon_i-)] = \mathsf{P}_\theta(\varepsilon > 0) - \mathsf{P}_\theta(\varepsilon \leq 0) \leq 0$ pro každé $i \in \{1, 2, \ldots, n\}$.

Jinými slovy 0 je mediánem chyby ε_i pro každé $i \in \{1, 2, \ldots, n\}$.

5.3. LMS-odhad (Least Median of Squares)

Uveďme ještě příklad odhadu regresních koeficientů, který není M-odhadem. Jedním z takových odhadů je LMS-odhad navržený v knize ??? . Jedná se o odhad minimalizující medián empirické chyby.

Tento odhad je založen na absolutních reziduích $|e_1(b)|, |e_2(b)| \ldots |e_n(b)|$, které uspořádáme podle velikosti, od nejmenšího po největší. Tyto uspořádané statistiky označme $e_{[1]}(b) \leq e_{[2]}(b) \leq \cdots \leq e_{[n]}(b)$.

Minimalizací j-té uspořádané statistiky získáme odhad $M_{j,n} \in \operatorname*{argmin}_{b \in \mathbb{R}^k} e_{[j]}(b)$, zde $j \in \{1, 2, \ldots, n\}$.

Tento odhad můžeme popsat také jiným způsobem. Z pozorování vybereme všemi způsoby j pozorování na jejichž základě uděláme odhad tak, že minimalizujeme maximum všech reziduí. Pak vybereme tu j-tici pozorování, která dává minimální rezidua.

Odhad je rozumným odhadem pro β_o pokud $j > \frac{n}{2}$. Při $j < \frac{n}{2}$ odhadujeme vlastně charakteristiku kontaminace původních dat.

Odhad $M_{\left[\frac{n+k+1}{2}\right],n}$ se nazývá LMS-odhad (Least Median of Squares) a je konzistentním odhadem parametru β_o, viz ??? .

Literatura

[Amemiya] T. Amemiya: Regression analysis when the dependent variable is truncated normal. Econometrica **42**, 999-1012, 1973.

[Anděl] J. Anděl: *Matematická statistika.* SNTL/ALFA, Praha 1978.

[Fomby, Hill, Johnson] T.B. Fomby and R.C. Hill and S.R. Johnson: *Advanced Econometric Methods.* Springer-Verlag, Berlin 1984.

[Heckman 1976] J. Heckman: The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. Annals of Econometric and Social Measurament **5**, 475-492, 1976.

[Heckman 1979] J. Heckman: Sample bias as specification error. Econometrica **47 ?**, 153-162, 1979.

[Jurečková, Sen] Jurečková, J. and P.K. Sen: *Robust Statistical Procedures.* John Wiley & Sons, Inc., New York 1996.

[Leroy, Rouseeuw] Robust Regression and Outlier Detection: *A.M. Leroy, P.J. Rousseeuw.* John Wiley & Sons, New York 1987.

[Machek] Josef Machek: *Teorie odhadu.* KPMS MFF UK, Praha 1980.

[van de Geer] Sara van de Geer: *Regression Analysis and Empirical Processes.* Ph.D. thesis of State University of Leiden, 1987.

[Tobin] J. Tobin: Estimation of relationships for limited dependent variables. Econometrica **26**, 24-36, 1958.

[van der Vaart, Wellner] Aad W. van der Vaart and Jon A. Wellner: *Weak Convergence and Empirical Processes.* Springer, New York 1996.

UK MFF, KPMS, Sokolovská 83, 186 75 Praha
E-mail: lachout@karlin.mff.cuni.cz