METODA SUBSAMPLING A JEJÍ APLIKACE V ČASOVÝCH ŘADÁCH

ZUZANA PRÁŠKOVÁ

Abstrakt. In the paper, some basic results concerning subsampling in time series are reviewed. The method is then applied to estimate parameters of autoregressive model with heteroscedasticities and consistency of such estimators is proved.
Резюме: В этой статье изучается применение метода подвыборки к анализу временных рядов. Показывается состоятельность оценок параметров процесса авторегрессии в случае неодинакого распределённых ошибок.

1. Úvod

Mějme pozorování náhodných veličin $X_1, \ldots, X_n$, které nutně nemusí být nezávislé a stejně rozdělené; nechť $\widehat{\theta}_n$ je odhad neznámého parametru θ spočtený z pozorování $X_1, \ldots, X_n$, *tj.* $\widehat{\theta}_n = \widehat{\theta}_n(X_1, \ldots, X_n)$, a nechť $R_n = R_n(X_1, \ldots, X_n)$ je statistika, na které je založeno statistické usuzování o $\widehat{\theta}_n$.

V našem příspěvku budeme uvažovat statistiku $R_n = \tau_n(\widehat{\theta}_n - \theta)$, kde τ_n je konstanta závislá jen na n; nejčastěji se volí $\tau_n = \sqrt{n}$. Zabývat se budeme aproximací distribuční funkce statistiky R_n a odhadem rozptylu $\widehat{\theta}_n$ metodou subsampling.

Na rozdíl od metody bootstrap, která je založena na mnohonásobném opakování prostého náhodného výběru s vracením ze souboru $X_1, \ldots, X_n$ (čtenáře, který se chce více dozvědět o metodě bootstrap, zde odkazujeme na monografie Efron a Tibshirani (1993) nebo Shao a Tu (1995)), spočívá metoda subsampling v práci s menšími soubory dat, které vybereme určitým způsobem z $X_1, \ldots, X_n$. Způsob, jakým budou tyto podsoubory pořízeny, je jiný pro nezávislá pozorování a jiný pro závislá data, např. pro časové řady. Těmi se budeme zabývat podrobněji, nejdříve však uvedeme základní známé výsledky metody subsampling pro nezávislá pozorování.

2. Metoda subsampling pro nezávislá pozorování

Nechť $X_1, \ldots, X_n$ jsou nezávislé stejně rozdělené náhodné veličiny s distribuční funkcí F, $\theta = \theta(F)$ je neznámý parametr a $\widehat{\theta}_n$ jeho odhad. Nechť $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N$ jsou všechny možné podsoubory o rozsahu $b < n$ pořízené prostým náhodným výběrem *bez vracení* ze souboru $X_1, \ldots, X_n$, tj.

$$\boldsymbol{Y}_i = (X_1^{(i)}, \ldots, X_b^{(i)}), \quad i = 1, \ldots, N,$$

kde $N = N_{n,b} = \binom{n}{b}$. Nechť $\widehat{\theta}_{n,b,i}$ je odhad θ spočtený z výběru $\boldsymbol{Y}_i$. Potom rozdělení

2000 *Mathematics Subject Classification.* Primary 62F40; Secondary 62M10.
Klíčová slova. Subsampling; strong mixing.
Tato práce vznikla za podpory grantu GAČR č. 201/00/0769 a výzkumného záměru MŠMT č. MSM 113200008.

$$H_n(x) = P[\tau_n(\widehat{\theta}_n - \theta) \le x] \tag{1}$$

je aproximováno empirickým rozdělením založeným na hodnotách $\tau_b(\widehat{\theta}_{n,b,i} - \widehat{\theta}_n)$, tj. empirickou distribuční funkcí

$$L_{n,b}(x) = \frac{1}{N} \sum_{i=1}^{N} I[\tau_b(\widehat{\theta}_{n,b,i} - \widehat{\theta}_n) \le x], \tag{2}$$

kde $I[A]$ značí indikátor množiny A.

Poznámka 1. V náhodném výběru s vracením, který teoreticky připouští metoda bootstrap, je $N = n^n$.

Konzistence tohoto postupu je zaručena následující větou (Politis, Romano a Wolf (1999), Theorem 2.2.1).

Věta 1. *Nechť H je distribuční funkce taková, že*

$$H_n \to H \text{ slabě, když } n \to \infty,$$

dále nechť $b \to \infty$, $\frac{b}{n} \to 0$, $\frac{\tau_b}{\tau_n} \to 0$. Platí následující tvrzení.

(i) *Jestliže x je bod spojitosti H, potom*

$$L_{n,b}(x) \to H(x) \text{ v pravděpodobnosti.}$$

(ii) *Jestliže H je spojitá, potom*

$$\sup_x |L_{n,b}(x) - H_n(x)| \to 0 \text{ v pravděpodobnosti.}$$

(iii) *Nechť $\alpha \in (0,1)$ a nechť $c(1-\alpha)$ je $(1-\alpha)$-kvantil rozdělení H a $c_{n,b}(1-\alpha)$ je $(1-\alpha)$-kvantil rozdělení $L_{n,b}$. Jestliže H je spojitá v $c(1-\alpha)$, potom*

$$P[\tau_n(\widehat{\theta}_n - \theta) \le c_{n,b}(1-\alpha)] \to 1 - \alpha.$$

3. Subsampling ve stacionárních časových řadách

Nyní předpokládejme, že $X_1, \dots, X_n$ jsou pozorování striktně stacionární náhodné posloupnosti $\{X_t\}_{-\infty}^{\infty}$, která má vlastnost *strong mixing* (jinak též *α-mixing*), tj. platí $\alpha_X(m) \to 0$ pro $m \to \infty$, kde koeficient $\alpha_X(m)$ je definovaný předpisem

$$\alpha_X(m) = \sup_n \sup_{A \in \mathcal{F}_{-\infty}^n, B \in \mathcal{F}_{m+n}^{\infty}} |P(A \cap B) - P(A)P(B)|,$$

a $\mathcal{F}_n^m = \sigma\{X_n, \dots, X_m\}$ je σ-algebra generovaná náhodnými veličinami $X_n, \dots, X_m$. Z předpokladu striktní stacionarity plyne, že všechny náhodné veličiny X_t mají stejné rozdělení, nemusí však být nezávislé.

Pořadí náhodných veličin je v tomto případě důležité, proto chceme-li použít metodu subsampling, nelze postupovat jako v případě nezávislých stejně rozdělených náhodných veličin a uvažovat všechny možné výběry bez vracení o rozsahu b ze souboru $X_1, \dots, X_n$. V tomto případě se uvažují klouzavé bloky délky b, $(X_1, \dots, X_b), \dots, (X_{n-b+1}, \dots, X_n)$, v bloku $(X_t, \dots, X_{t+b-1})$ se stanoví odhad $\widehat{\theta}_{n,b,t}$ parametru θ a rozdělení H_n definované v (1) se aproximuje rozdělením

$$L_{n,b}(x) = \frac{1}{n-b+1} \sum_{t=1}^{n-b+1} I[\tau_b(\widehat{\theta}_{n,b,t} - \widehat{\theta}_n) \le x]. \tag{3}$$

Náhodné vektory $(X_t, \ldots, X_{t+b-1})$ pro $t = 1, \ldots, n-b+1$ mají vzhledem ke striktní stacionaritě stejné rozdělení a lze ukázat, že tvrzení $(i) - (iii)$ Věty 1 platí i pro striktně stacionární posloupnosti s vlastností strong mixing a pro distribuční funkci (3) (viz Politis, Romano a Wolf (1999), Theorem 3.2.1).

Rozptyl empirického rozdělení (3) je $(n-b+1)^{-1}\sum_{t=1}^{n-b+1}\tau_b^2(\widehat{\theta}_{n,b,t}-\overline{\theta}_b)^2$, kde

$$\overline{\theta}_b = \frac{1}{n-b+1}\sum_{t=1}^{n-b+1}\widehat{\theta}_{n,b,t},$$

je tedy přirozené odhadnout rozptyl $\widehat{\theta}_n$ jako

$$\widehat{V}_{n,b}^{(1)} = \frac{\tau_b^2}{\tau_n^2}\frac{1}{n-b+1}\sum_{t=1}^{n-b+1}(\widehat{\theta}_{n,b,t}-\overline{\theta}_b)^2. \tag{4}$$

Pro $n = kb$ a nepřekrývající se bloky délky b, tj. pro vektory $(X_{(i-1)b+1}, \ldots, X_{ib})$ a $i = 1, \ldots, k$, dostaneme odhad

$$\widehat{V}_{n,b}^{(2)} = \frac{\tau_b^2}{\tau_n^2}\frac{1}{k}\sum_{i=1}^{k}(\widehat{\theta}_{n,b,i}-\widetilde{\theta}_b)^2, \tag{5}$$

kde

$$\widetilde{\theta}_b = \frac{1}{k}\sum_{i=1}^{k}\widehat{\theta}_{n,b,i}$$

a $\widehat{\theta}_{n,b,i}$ je odhad v $i-$tém bloku. Podmínky konzistence těchto odhadů byly studovány různými autory a jsou shrnuty v Politis, Romano a Wolf (1999), Lemma 3.8.1.

Při použití metody subsampling vzniká problém jak stanovit délku bloku, neboť věty o konzistenci jsou formulovány za předpokladu $b \to \infty$ a $\frac{b}{n} \to \infty$ pro $n \to 0$.

Např. rozptyl výběrového průměru $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, kde $X_1, \ldots, X_n$ jsou pozorování striktně stacionární posloupnosti se střední hodnotou $\mathsf{E}\, X_1 = \theta$ a autokovarianční funkcí $R(k) = \mathsf{E}(X_{k+1}-\theta)(X_1-\theta)$, je

$$\mathsf{Var}\overline{X}_n = \frac{1}{n}\left[R(0) + 2\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)R(k)\right] \tag{6}$$

a $n\mathsf{Var}\overline{X}_n \to \sigma_\infty^2$ pro $n \to \infty$ a $\sum_{-\infty}^{\infty}|R(k)| < \infty$, kde $\sigma_\infty^2 = \sum_{-\infty}^{\infty} R(k)$ (Brockwell a Davis (1991), Věta 7.1.1). Pokud autokovarianční funkci neznáme, můžeme $\mathsf{Var}\overline{X}_n$ odhadnout tak, že do vzorce (6) dosadíme výběrové autokovariance

$$\widehat{R}(k) = \frac{1}{n}\sum_{t=1}^{n-k}(X_t-\overline{X}_n)(X_{t+k}-\overline{X}_n).$$

Pro velké hodnoty k (blízké n) však jsou odhady $\widehat{R}(k)$ značně nepřesné.

Pro odhady (4) a (5) zkráceně označené jako $\widehat{V}^{(1)}, \widehat{V}^{(2)}$ v tomto případě platí, že $n\widehat{V}^{(1)}$ a $n\widehat{V}^{(2)}$ jsou konzistentními odhady σ_∞^2, a délka bloku, která minimalizuje asymptotickou střední kvadratickou chybu těchto odhadů, je přibližně

$$b \approx \left(\frac{(\sum_{k=-\infty}^{\infty}|k|R(k))^2}{c\sigma_\infty^4}\right)^{\frac{1}{3}} n^{\frac{1}{3}}, \tag{7}$$

kde $c = \frac{2}{3}$ pro odhad $\widehat{V}^{(1)}$ a $c = 1$ pro odhad $\widehat{V}^{(2)}$ (Politis, Romano a Wolf (1999), kap. 9). Je zřejmé, že takto lze délku bloku stanovit jen při známé autokovarianční

n	$\mathsf{Var}\overline{X}_n$	$\widehat{V}^{(1)}$	$b^{(1)}$	$\widehat{V}^{(2)}$	$b^{(2)}$
50	0.04460	0.03300 (0.00910)	2	0.03333 (0.00936)	2
100	0.02240	0.01840 (0.00444)	3	0.01700 (0.00343)	2
200	0.01122	0.00975 (0.00178)	4	0.00972 (0.00194)	4
400	0.00562	0.00504 (0.00070)	5	0.00493 (0.00070)	4
500	0.00450	0.00406 (0.00051)	5	0.00396 (0.00050)	4
1000	0.00225	0.00208 (0.00022)	7	0.00203 (0.00020)	5
2000	0.00125	0.00106 (7.85e-005)	8	0.00105 (7.25e-005)	8
5000	0.00045	0.00043 (2.34e-005)	10	0.00043 (2.77e-005)	10

Tabulka 1. Odhad rozptylu průměru metodou subsampling při optimální délce bloku.

funkci $R(k)$.

V tabulce 1 jsou uvedeny výsledky simulací, které srovnávají skutečnou hodnotu rozptylu spočtenou podle vzorce (6) a jeho odhady $\widehat{V}^{(1)}$ a $\widehat{V}^{(2)}$ metodou subsampling pro různé počty pozorování n a odpovídající délky bloku $b^{(1)}, b^{(2)}$ počítané podle (7) a zaokrouhlené na celá čísla. Posloupnost $X_1, \dots, X_n$ byla generována jako MA(1) podle předpisu $X_t = Y_t + \frac{1}{2}Y_{t-1}$, kde $\{Y_t\}$ byl striktně stacionární bílý šum s rozdělením $\mathcal{N}(0,1)$. Směrodatné odchylky pro 1000 opakování jsou uvedeny v závorkách.

Dále jsme studovali chování odhadu rozptylu $\mathsf{Var}\overline{X}_n$ metodou subsampling v závislosti na délce bloku při pevném počtu pozorování.

Byly generovány posloupnosti

$$MA(1) : X_t = Y_t + \frac{1}{2}Y_{t-1} \tag{8}$$

$$MA(2) : X_t = Y_t + Y_{t-1} + Y_{t-2} \tag{9}$$

$$AR(1) : X_t = \beta X_{t-1} + Y_t, \quad \beta = 0.3 \text{ a } \beta = 0.8 \tag{10}$$

pro rozsahy $n = 100, n = 160, n = 200$ a $n = 400$. Posloupnost Y_t byla ve všech případech generována jako striktní bílý šum s rozdělením $\mathcal{N}(0,1)$.

Poznámka 2. Takto definované posloupnosti jsou striktně stacionární a mají vlastnost strong mixing (Davidson (1994), věta 14.9).

Pro každou řadu byl spočten odhad rozptylu průměru metodou subsampling pro klouzavé bloky (odhad $\widehat{V}^{(1)}$) a nepřekrývající se bloky (odhad $\widehat{V}^{(2)}$). Pro každou délku bloku bylo provedeno 1000 opakování a spočten průměr a směrodatná odchylka. Výsledky byly porovnány se skutečnou hodnotou rozptylu, která byla spočtena podle (6). Na obrázcích je skutečná hodnota rozptylu vyznačena $+$, hodnota odhadu $\widehat{V}^{(1)}$ jako $\bullet$ a hodnota odhadu $\widehat{V}^{(2)}$ jako $*$. U nepřekrývajících se bloků se rozsah, při kterém bylo dosaženo nejlepší shody se skutečnou hodnotou, pohyboval mezi $2-5$, u klouzavých bloků ve většině případů mezi $10-20$, ale odhady měly v některých případech velkou směrodatnou odchylku. Výsledky simulací pro rozsahy $n=100$ a $n=400$ jsou demonstrovány na obrázcích 1 a 2. Z obrázků je patrná určitá stabilita odhadů získaných metodou klouzavých bloků, které mají i jinou než optimální délku.

Na obrázku 3 je pro ilustraci také aproximace skutečného rozdělení metodou subsampling z jedné výběrové realizace.

Určitým řešením jak zvolit délku bloku, i když nedává optimální výsledky, je použít kalibrační metodu nebo metodu minimální volatility. Obě metody jsou popsány v Politis, Romano a Wolf (1999).

4. Subsampling v nestacionárních časových řadách

Nyní předpokládejme. že posloupnost $\{X_t\}_{-\infty}^{\infty}$ není striktně stacionární, ale má vlastnost strong mixing. Podobně jako v předchozím odstavci je metoda subsampling pro odhad obecného parametru θ založena na klouzavých blocích $(X_t,\ldots,X_{t+b-1})$ pro $t=1,\ldots,n-b+1$. Tyto náhodné vektory však již nemají stejné rozdělení. Nechť $\widehat{\theta}_n, \widehat{\theta}_{n,b,t}$ a $L_{n,b}$ jsou definovány stejně jako pro stacionární posloupnosti, H_n je definovaná v (1), nechť

$$H_{b,t}(x) = P[\tau_b(\widehat{\theta}_{n,b,t}-\theta) \le x]. \tag{11}$$

O konzistenci $L_{n,b}$ lze dokázat následující větu (Politis, Romano, Wolf (1999), Theorem 4.2.1).

Věta 2. *Nechť $\{X_t\}_{-\infty}^{\infty}$ je náhodná posloupnost s vlastností strong mixing. Předpokládejme, že existuje distribuční funkce H taková, že*

(a) *$H_n \to H$ slabě, když $n\to\infty$,*
(b) *V každém bodě spojitosti H a pro každou posloupnost indexů t_b platí*

$$H_{b,t_b}(x) \to H(x), \text{ když } b\to\infty.$$

Nechť dále $\frac{\tau_b}{\tau_n}\to 0$ a nechť $L_{n,b}$ je empirická distribuční funkce definovaná v (3). Potom platí tvrzení $(i)-(iii)$ Věty 1.

Poznámka 3. Předpoklad (b) lze nahradit slabším předpokladem

(b') *V každém bodě spojitosti H a pro $n\to\infty, b\to\infty$ a $\frac{b}{n}\to 0$*

$$\frac{1}{n-b+1}\sum_{t=1}^{n-b+1} H_{b,t}(x) \to H(x).$$

Výsledky lze zobecnit na vícerozměrný parametr.

Nyní uvažujme pozorování $X_1,\ldots,X_n$ autoregresní posloupnosti AR(1) definované jako

$$X_t = \beta X_{t-1} + Y_t,\ t=0,\pm1,\ldots, \tag{12}$$

kde $|\beta| < 1$ a Y_t jsou nezávislé náhodné veličiny s nulovou střední hodnotou a rozptyly σ_t^2. Posloupnost (12) není obecně stacionární. Nechť $\widehat{\beta}_n$ je odhad parametru β spočtený metodou nejmenších čtverců z pozorování $X_1, \ldots, X_n$, tj.

$$\widehat{\beta}_n = \frac{\sum_{t=2}^{n} X_t X_{t-1}}{\sum_{t=2}^{n} X_{t-1}^2}.$$

Ukážeme, že rozděleni $\sqrt{n}(\widehat{\beta}_n - \beta)$ je možné aproximovat metodou subsampling. Máme tedy $\tau_n = \sqrt{n},\ H_n(x) = P[\sqrt{n}(\widehat{\beta}_n - \beta) \le x]$. Nechť $\widehat{\beta}_{n,b,t}$ je odhad β z pozorování $X_t, \ldots, X_{t+b-1}$, tj.

$$\widehat{\beta}_{n,b,t} = \frac{\sum_{j=t+1}^{t+b-1} X_j X_{j-1}}{\sum_{j=t+1}^{t+b-1} X_{j-1}^2},$$

$$H_{b,t}(x) = P[\sqrt{b}(\widehat{\beta}_{n,b,t} - \beta) \le x],$$

a

$$L_{n,b}(x) = \frac{1}{n-b+1} \sum_{t=1}^{n-b+1} I[\sqrt{b}(\widehat{\beta}_{n,b,t} - \widehat{\beta}_n) \le x].$$

Nyní lze zformulovat podmínky konzistence.

Věta 3. *Nechť X_t je definovaná v (12), kde $|\beta| < 1$, Y_t jsou nezávislé s nulovou střední hodnotou, rozptyly σ_t^2 a stejnoměrně ohraničenými momenty 4. řádu a mají hustoty f_t pro které platí*

$$\sup_t \frac{1}{|a|} \int_{-\infty}^{\infty} |f_t(z+a) - f_t(z)| dz \le M \text{ pro všechna } |a| < \gamma \tag{13}$$

kde M, γ jsou kladné konstanty. Nechť dále platí

$$\frac{1}{k} \sum_{j=t}^{t+k-1} \sigma_j^2 \to \sigma^2 > 0 \text{ stejnoměrně v } t, \text{ když } k \to \infty, \tag{14}$$

$$\frac{1}{k} \sum_{j=t+1}^{t+k-1} \sigma_j^2 \mathsf{E} X_{j-1}^2 \to \overline{\sigma}^2 > 0 \text{ stejnoměrně v } t, \text{ když } k \to \infty. \tag{15}$$

Potom pro odhad parametru β metodou subsampling platí

$$\sup_x |L_{n,b}(x) - H_n(x)| \to 0 \text{ v pravděpodobnosti}$$

jestliže $n \to \infty, b \to \infty, \frac{b}{n} \to 0$ a $L_{n,b}$ a H_n jsou distribuční funkce definované výše.

Důkaz. Stačí ověřit předpoklady Věty 2. Podmínka $|\beta| < 1$, předpoklad o stejnoměrně ohraničených momentech a podmínka (13) zaručují, že posloupnost $\{X_t\}$ má vlastnost strong mixing (Davidson (1994), věta 14.9.)

Dále uvažujme schéma náhodných veličin

$$\begin{aligned}
\xi_{t_b,v} &= X_{t_b+v-1} & t_b &= 1, \ldots, n-b+1,\ v = 1, \ldots, b-1,\\
\nu_{t_b,v} &= Y_{t_b+v} & t_b &= 1, \ldots, n-b+1,\ v = 1, \ldots, b-1,\\
\zeta_{t_b,v} &= X_{t_b+v-1} Y_{t_b+v} & t_b &= 1, \ldots, n-b+1,\ v = 1, \ldots, b-1,
\end{aligned}$$

kde $b = b(n)$ a $b \to \infty$, když $n \to \infty$. Zřejmě je

$$\sqrt{b}(\widehat{\beta}_{n,b,t_b} - \beta) = \sqrt{b}\frac{\sum_{j=t_b+1}^{t_b+b-1} X_{j-1}Y_j}{\sum_{j=t_b+1}^{t_b+b-1} X_{j-1}^2} = \frac{\frac{1}{\sqrt{b}}\sum_{v=1}^{b-1}\zeta_{t_b,v}}{\frac{1}{b}\sum_{v=1}^{b-1}\xi_{t_b,v}^2}. \tag{16}$$

Pro každý index t_b jsou náhodné veličiny $\nu_{t_b,v}$ nezávislé, s nulovou střední hodnotou, s rozptyly $\mathsf{E}(\nu_{t_b,v})^2 = \sigma_{t_b+v}^2$ a se stejnoměrně ohraničenými momenty 4. řádu, tedy podle zákona velkých čísel a podmínky (14) platí

$$\frac{1}{b}\sum_{v=1}^{b-1}\nu_{t_b,v}^2 \to \sigma^2 \text{ pro } b \to \infty \text{ skoro jistě.} \tag{17}$$

Náhodné veličiny $\zeta_{t_b,v}$ tvoří posloupnost martingalových diferencí, pro které podle Prášková (1997), viz též Basu a Sen Roy (1990) a za platnosti podmínky (15) platí

$$\mathcal{L}\left(\frac{1}{\sqrt{b}}\sum_{v=1}^{b-1}\zeta_{t_b,v}\right) \to \mathcal{N}(0,\overline{\sigma}^2) \text{ pro } b \to \infty \tag{18}$$

pro každý index t_b.

Konečně, když využijeme (12), dostaneme

$$(1-\beta^2)\frac{1}{b}\sum_{v=1}^{b-1}\xi_{t_b,v}^2 = \frac{1}{b}(X_{t_b}^2 - X_{t_b+b-1}^2) + 2\beta\frac{1}{b}\sum_{v=1}^{b-1}\zeta_{t_b,v} + \frac{1}{b}\sum_{v=1}^{b-1}\nu_{t_b,v}^2. \tag{19}$$

První člen na pravé straně (19) konverguje v pravděpodobnosti k nule pro $b \to \infty$ vzhledem k předpokladu o stejnoměrně ohraničených čtvrtých momentech Y_t (lze snadno ukázat, že potom i X_t mají stejnoměrně ohraničené momenty 4. řádu.) Druhý člen konverguje v pravděpodobnosti k nule podle zákona velkých čísel pro martingalové diference $\zeta_{t_b,v}$, což spolu s (17) implikuje, že

$$\frac{1}{b}\sum_{v=1}^{b-1}\xi_{t_b,v}^2 \to \frac{\sigma^2}{1-\beta^2} \text{ v pravděpodobnosti pro } b \to \infty$$

pro každý index t_b. Odtud tedy plyne, že pro každý index t_b a $b \to \infty, n \to \infty$ má $\sqrt{b}(\widehat{\beta}_{n,b,t_b} - \beta)$ asymptoticky normální rozdělení $\mathcal{N}(0,\varDelta^2)$, kde

$$\varDelta^2 = \frac{(1-\beta^2)^2\overline{\sigma}^2}{\sigma^4},$$

tj. $H_{b,t_b}(x) \to H(x) = \Phi(\frac{x}{\varDelta})$ stejnoměrně v x a pro každý index t_b, což je podmínka (b) Věty 2. Podmínka (a) je splněna, položíme-li $b = n,\ t_b = 1$. □

Uvažujeme-li model $AR(1)$ s nenulovou střední hodnotou, tj. model

$$X_t - \mu = \beta(X_{t-1} - \mu) + Y_t, t = 0, \pm1, \dots, \tag{20}$$

potom momentové odhady parametrů jsou

$$\widehat{\mu}_n = \frac{1}{n}\sum_{t=1}^{n} X_t = \overline{X}_n, \quad \widehat{\beta}_n = \frac{\sum\limits_{t=2}^{n}(X_t - \overline{X}_n)(X_{t-1} - \overline{X}_n)}{\sum\limits_{t=2}^{n}(X_{t-1} - \overline{X})^2}$$

a jejich rozdělení lze opět aproximovat metodou subsampling. Označíme-li jako $\widehat{\mu}_{n,b,t}$ a $\widehat{\beta}_{n,b,t}$ odhady spočtené z bloku pozorování $X_t, \dots, X_{t+b-1}$, lze zformulovat následující tvrzení o konzistenci.

Věta 4. *Nechť* $X_1, \dots, X_n$ *jsou pozorování, která se řídí modelem (20) a nechť jsou splněny podmínky věty 3. Nechť* $n \to \infty, b \to \infty$ *a* $\frac{b}{n} \to 0$. *Potom platí*

$$\sup_x |P[\sqrt{n}(\widehat{\mu}_n - \mu) \le x] - L^{\mu}_{n,b}(x)| \to 0 \text{ v pravděpodobnosti}$$

a

$$\sup_x |P[\sqrt{n}(\widehat{\beta}_n - \beta) \le x] - L^{\beta}_{n,b}(x)| \to 0 \text{ v pravděpodobnosti}$$

kde

$$L^{\mu}_{n,b}(x) = \frac{1}{n-b+1} \sum_{t=1}^{n-b+1} I[\sqrt{b}(\widehat{\mu}_{n,b,t} - \widehat{\mu}_n) \le x],$$

a $L^{\beta}_{n,b}(x)$ *má stejný význam pro odhady* $\widehat{\beta}_n$ *a* $\widehat{\beta}_{n,bt}$.

Důkaz. Stejným způsobem jako v předchozí větě se ukáže, že za podmínky (14) platí stejnoměrně v x a pro každý index t_b, že $P[\sqrt{n}(\overline{\mu}_{n,b,t_b} - \mu) \le x] \to H^{\mu}(x)$, kde $H^{\mu}(x)$ je distribuční funkce normálního rozdělení $\mathcal{N}(0, \sigma^2(1-\beta)^{-2})$ (Prášková (1997)). Dále se aplikuje postup z věty 3 na pozorování $X_t - \mu$. □

Tvrzení věty 3 lze dále zobecnit na model autoregrese vyššího řádu

$$X_t = \sum_{j=1}^{p} \beta_j X_{t-j} + Y_t, \; t = 0, \pm 1 \dots \tag{21}$$

Model lze psát ve vektorovém zápisu jako

$$X_t = \boldsymbol{X}(t-1)^T \boldsymbol{\beta} + Y_t, \tag{22}$$

kde $\boldsymbol{X}(t)$ je vektor $(X_t, X_{t-1}, \dots, X_{t-p+1})^T$ a $\boldsymbol{\beta}$ je vektor autoregresních parametrů $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$. Odhad parametru $\boldsymbol{\beta}$ metodou nejmenších čtverců spočtený z pozorování $X_1, \dots, X_n$ je

$$\widehat{\boldsymbol{\beta}_n} = \left[\sum_{t=2}^{n} \boldsymbol{X}(t-1)\boldsymbol{X}(t-1)^T \right]^{-1} \sum_{t=2}^{n} \boldsymbol{X}(t-1) Y_t. \tag{23}$$

Věta 5. *Uvažujme model (21). Předpokládejme, že náhodné chyby* Y_t *jsou nezávislé s hustotami a momenty, které splňují podmínky věty 3 a dále nechť kořeny polynomu*

$$\lambda^p - \beta_1 \lambda^{p-1} + \dots + \beta_p$$

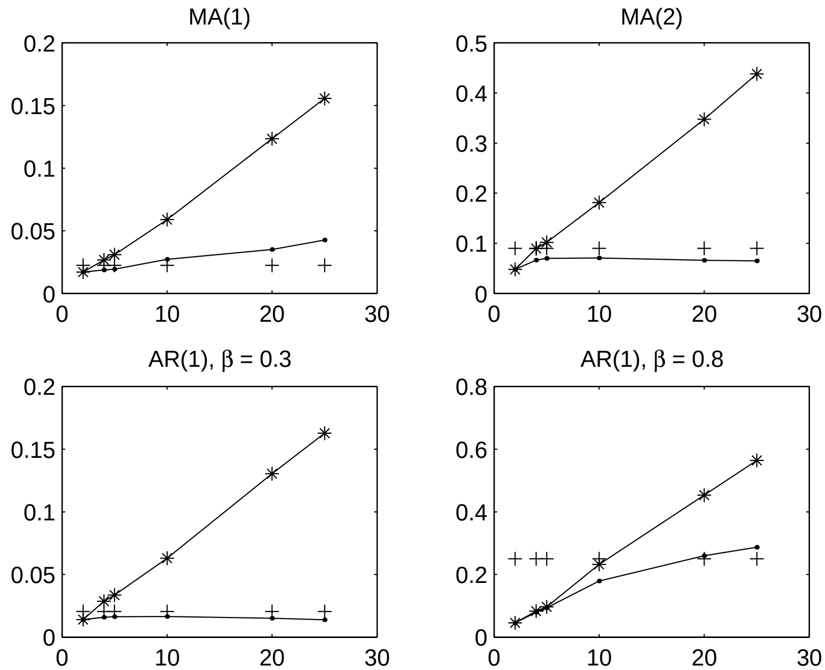
jsou v absolutní hodnotě ostře menší než 1. *Předpokládejme, že platí podmínka (14) a podmínka*

$$\frac{1}{k} \sum_{j=t+1}^{t+k-1} \sigma_j^2 \mathsf{E} \boldsymbol{X}(j-1)\boldsymbol{X}(j-1)^T \to \overline{\boldsymbol{\Sigma}} \text{ stejnoměrně v } t, \text{ když } k \to \infty, \tag{24}$$

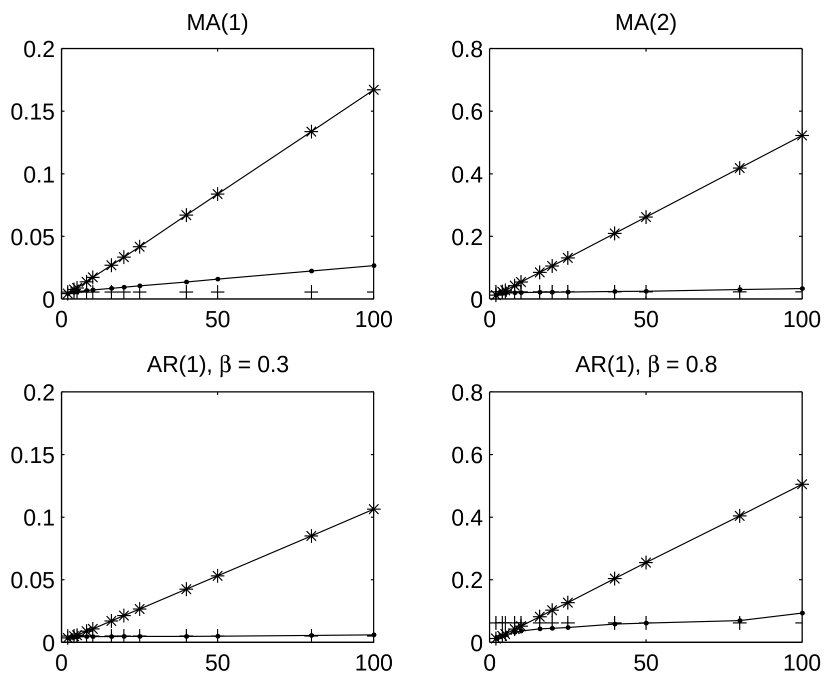
kde $\overline{\boldsymbol{\Sigma}}$ *je pozitivně definitní nenulová matice.*

Potom rozdělení náhodného vektoru $\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ *lze aproximovat empirickým rozdělením spočteným z hodnot* $\sqrt{b}(\widehat{\boldsymbol{\beta}}_{n,b,t} - \widehat{\boldsymbol{\beta}}_n), \; t = 1, \dots, n-b+1$.

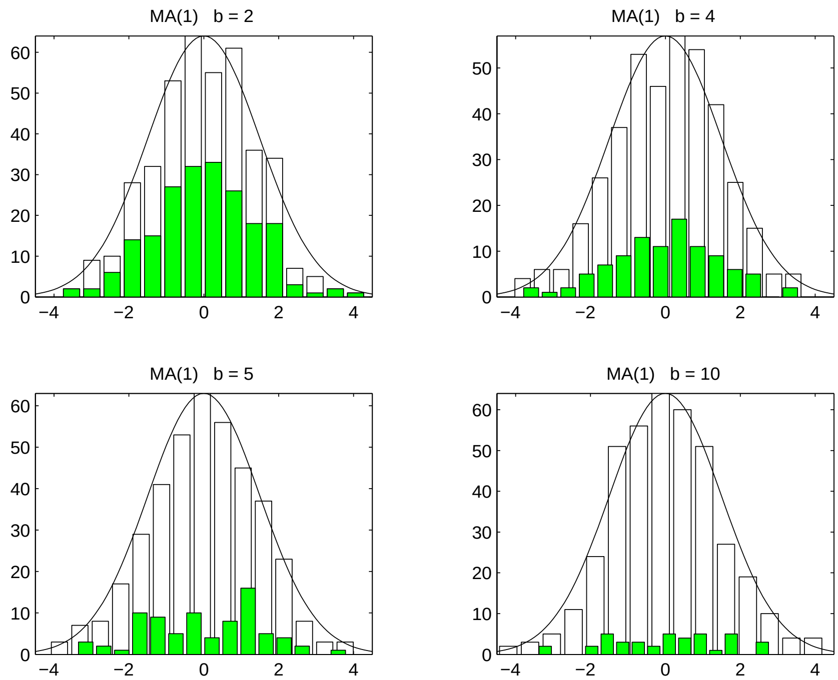
Důkaz. Postupuje se stejně jako v jednorozměrném případě. Podmínka na kořeny spolu s podmínkou (13) zaručují, že posloupnost (21) má vlastnost strong mixing (Davidson (1994), Corollary 14.13). Dále se využije vlastnosti vektoru martingalových diferencí $\boldsymbol{X}(t-1)Y_t$ a jeho asymptotické normality podle Tjøstheim a Paulssen (1985) (tam dokázané Lemma 3.1 platí i za našich předpokladů) a ukáže se, že jsou splněny předpoklady 4.3.2 z Romano, Politis a Wolf (1999) (str. 106-107). □



Obrázek 1. Závislost odhadu rozptylu průměru na délce bloku, 100 pozorování.



Obrázek 2. Závislost odhadu rozptylu průměru na délce bloku, 400 pozorování.



Obrázek 3. Aproximace rozdělení odhadu rozptylu metodou subsampling, 400 pozorování. Tmavě jsou označeny nepřekrývající se bloky.

Literatura

1 Basu, A. K., Sen Roy, S. (1990): On rates of convergence in the central limit theorem for parameter estimation in general autoregressive model. Statistics, 21, 461-470

2 Brockwell, P. J., Davis, R. A. (1991): Time Series: Theory and Methods. Springer, New York

3 Davidson, J. (1994): Stochastic Limit Theory. Oxford University Press, Oxford

4 Efron, B., Tibshirani, R. J. (1993): An Introduction to the Bootstrap. Chapmann & Hall, New York, London

5 Politis, D. N., Romano, J. P., Wolf, M. (1999): Subsampling. Springer, New York

6 Prášková, Z. (1997): Bootstrapping in autoregression with heteroscedasticity. Submitted.

7 Shao, J., Tu, D. (1995): The Jackknife and Bootstrap. Springer, New York

8 Tjøstheim, D., Paulsen, J. (1985): Least squares estimates and order determination procedures for autoregressive process with a time dependent variance. J. Time Ser. Anal., 6, 117-138

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