 © JČMF 2004

METODA BOOTSTRAP

Zuzana Prášková

Klíčová slova: Bootstrap, wild bootstrap, blokový bootstrap.

Abstrakt: V tomto článku jsou shrnuty základní vlastnosti metody bootstrap. Je popsán klasický přístup založený na nezávislých stejně rozdělených náhodných veličinách a také modifikace této metody pro časové řady.

1 Základní myšlenka metody bootstrap

Metoda bootstrap patří mezi tzv. intenzivní počítačové metody pro statistickou analýzu dat. První článek o bootstrapu [4] vyvolal velký ohlas a brzy po jeho zveřejnění byla publikována řada dalších teoretických i simulačních studií, které měly za cíl zkoumat použití, účinnost a spolehlivost této metody v nejrůznějších aplikacích.

V tomto článku uvedeme základní vlastnosti metody bootstrap a zmíníme se i o současných trendech.

Uvažujme nezávislé stejně rozdělené (*iid*) náhodné veličiny $X_1, \ldots, X_n$, jejichž distribuční funkce F není blíže specifikována. Nechť $\theta = \theta(F)$ je nějaká charakteristika rozdělení; je to pro nás neznámý parametr, který má být odhadnut na základě realizace náhodného výběru.

Nechť $T_n = T_n(X_1, \ldots, X_n)$ je statistika pro odhad parametru θ, nechť $R_n = R_n(X_1, \ldots, X_n)$ je její vhodně standardizovaná verze, např. $R_n = \sqrt{n}(T_n - \theta)$, nebo nějaká její funkce. Nechť

$$H_n(x) = P[R_n(X_1, \ldots, X_n, F) \leq x]$$

značí distribuční funkci statistiky R_n.

Explicitní odvození rozdělení H_n i výpočet číselných charakteristik mohou být v jednotlivých případech značně obtížné, či dokonce analyticky neproveditelné a to i tehdy, když je distribuční funkce F známá. V takovém případě (při známé distribuční funkci) lze postupovat metodou Monte Carlo, generovat dlouhou sérii nezávislých náhodných výběrů z rozdělení s danou distribuční funkcí (tj. mnohokrát uměle opakovat experiment), pro každé opakování spočítat hodnotu příslušné charakteristiky a její skutečné rozdělení aproximovat empirickým rozdělením získaným z řady takto uměle získaných hodnot.

Je-li skutečná distribuční funkce F neznámá, což je mnohem častější případ, je možné aproximovat H_n asymptotickým rozdělením, které lze odvodit na základě limitních vět teorie pravděpodobnosti. Přesnost takové aproximace však je ovlivněna a omezena počtem pozorování, která jsou skutečně k dispozici.

Metoda *bootstrap* nabízí řešení, které kombinuje tzv. *substituční princip* a *metodu Monte Carlo.*

Vysvětleme nejdříve substituční princip. Nechť $F_n(x)$ je nějaký odhad distribuční funkce. Nejčastěji se uvažuje empirická distribuční funkce založená na náhodném výběru $X_1, \ldots, X_n$, tj.

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I[X_i \leq x],$$

kde $I[A]$ značí indikátor množiny A. Při daných hodnotách $X_1, \ldots, X_n$ je F_n *známá* funkce.

Nechť $X_1^*, \ldots, X_n^*$ je nezávislý náhodný výběr z F_n, tj. při daných pozorováních $X_1, \ldots, X_n$ jsou $X_1^*, \ldots, X_n^*$ (podmíněně) nezávislé, stejně rozdělené náhodné veličiny, z nichž každá nabývá hodnot $X_1, \ldots, X_n$ s pravděpodobností $\frac{1}{n}$. Soubor $X_1^*, \ldots, X_n^*$ se nazývá *bootstrapový výběr.* V dalších úvahách původní výběr nahradíme bootstrapovým výběrem a neznámou distribuční funkci F známou distribuční funkcí F_n. Dostaneme parametr $\theta^* = \theta(F_n)$ a statistiky $T_n^* = T_n(X_1^*, \ldots, X_n^*)$ a $R_n^* = R_n(X_1^*, \ldots, X_n^*, F_n)$.

Nyní můžeme definovat charakteristiky jako

$$E^* T_n^* = \int T_n(x_1, \ldots, x_n) d(F_n(x_1) \ldots F_n(x_n)),$$

$$\mathrm{var}^* T_n^* = \int [T_n(x_1, \ldots, x_n) - E^* T_n^*]^2 d(F_n(x_1), \ldots F_n(x_n))$$

a distribuční funkci

$$\begin{aligned} H_n^*(x) &= P^*(R_n(X_1^*, \ldots, X_n^*, F_n) \leq x) \\ &= P(R_n(X_1^*, \ldots, X_n^*, F_n) \leq x | X_1, \ldots, X_n); \end{aligned}$$

jsou to tzv. *teoretické charakteristiky* a *teoretická distribuční funkce* získané metodou bootstrap.

Řekneme, že H_n^* je konsistentní odhad H_n, jestliže

$$\rho(H_n^*, H_n) \to 0 \text{ při } n \to \infty$$

v pravděpodobnosti (slabá konzistence) nebo skoro jistě (silná konzistence), kde ρ je nějaká metrika na prostoru distribučních funkcí. Nejčastěji používané metriky v tomto případě jsou *supremální metrika*

$$\rho_\infty(G, H) = \sup_{x \in \mathbb{R}} |G(x) - H(x)|$$

nebo tzv.*Mallowsova vzdálenost,* která je pro distribuční funkce G a H z rodiny distribučních funkcí s konečnými $r-$tými momenty definovaná předpisem

$$\widetilde{\rho}_r(G, H) = \inf_{\mathcal{T}_{X,Y}} (E|X - Y|^r)^{1/r},$$

kde $\mathcal{T}_{X,Y}$ je množina všech možných sdružených rozdělení vektorů (X, Y), jejichž marginální rozdělení jsou G a H. O vlastnostech Mallowsovy vzdálenosti a souvislosti s konvergencí v distribuci náhodných veličin viz např. [2].

Konzistence bootstrapových charakteristik se definuje přirozeným způsobem. Řekneme např., že $\operatorname{var}{}^*T_n^*$ je konzistentní odhad rozptylu $\operatorname{var} T_n$, jestliže

$$\operatorname{var}{}^*T_n^* / \operatorname{var} T_n \to 1 \text{ při } n \to \infty$$

buď v pravděpodobnosti nebo skoro jistě.

Pro praktické použití jsou teoretické bootstrapové charakteristiky vhodné jen v případě, že jsou explicitními funkcemi pozorování $X_1, \ldots, X_n$. Přesné stanovení bootstrapového rozdělení by vyžadovalo provedení všech n^n možných výběrů s vracením z populace pozorovaných hodnot $X_1, \ldots, X_n$. To je však uskutečnitelné jen pro výběry o malém rozsahu. I kdybychom se omezili jen na vzájemně různé výběry, máme takových výběrů stále ještě $\binom{2n-1}{n}$, což již pro $n = 10$ je hodnota 92 378.

Nejčastěji se proto na bootstrapový výběr $X_1^*, \ldots X_n^*$ a známou distribuční funkci F_n aplikuje metoda Monte Carlo, kdy se mnohokrát (B-krát) generuje nezávislý náhodný výběr z rozdělení F_n, při každém opakování se spočtou hodnoty T_n^*, R_n^* a z nich se stanoví aritmetický průměr. Dostaneme tak *bootstrapové odhady* původního rozdělení a původních charakteristik.

Např. bootstrapový odhad rozptylu T_n dostaneme tak, že opakujeme nezávislý náhodný výběr z rozdělení F_n celkem $B-$krát a spočteme vždy hodnotu statistiky T_n^*. Dostáváme tak hodnoty $T_{n,1}^*, \ldots, T_{n,B}^*$, ze kterých spočteme

$$\widehat{\operatorname{var}{}^*}T_n^* = \frac{1}{B}\sum_{b=1}^{B}\left(T_{n,b}^* - \frac{1}{B}\sum_{k=1}^{B}T_{n,k}^*\right)^2.$$

Podobně odhadneme distribuční funkci statistiky R_n jako

$$\widehat{H_n^*}(x) = \frac{1}{B}\sum_{b=1}^{B} I\{R_n(X_{1,b}^*, \ldots, X_{n,b}^*, F_n) \le x\},$$

kde $\{X_{1,b}^*, \ldots, X_{n,b}^*\}, b = 1, \ldots, B$, jsou nezávislé výběry z F_n. Pro některé účely se lépe hodí histogram pořízený z hodnot R_n^*.

1.0.1 Příklad. Nechť $X_1, \ldots, X_n$ je náhodný výběr z rozdělení se střední hodnotou μ a rozptylem σ^2. Přirozeným odhadem parametru $\theta = e^\mu$ je statistika $T_n = e^{\overline{X}_n}$, kde $\overline{X}_n$ je výběrový průměr. Zabývejme se odhadem směrodatné odchylky $s_n = \sqrt{\operatorname{var} T_n}$.

V případě, že $X_i \sim \mathcal{N}(\mu, \sigma^2)$, má statistika T_n logaritmicko-normální rozdělení s parametry μ a $\frac{\sigma^2}{n}$, tedy

$$s_n = \left[e^{2\mu+\frac{\sigma^2}{n}}\left(e^{\frac{\sigma^2}{n}} - 1\right)\right]^{\frac{1}{2}}. \tag{1}$$

n	s_n	$\widetilde{s}_n$	s_{boot}
50	9,760	9,798	10,492
200	4,407	4,384	4,484
500	2,731	2,736	2,748

Tabulka 1: Porovnání odhadů směrodatné odchylky $T_n = e^{\bar{X}_n}$; s_n, $\widetilde{s}_n$, resp. s_{boot} značí skutečnou, simulovanou, resp. bootstrapovou směrodatnou odchylku.

V tabulce 1 jsou porovnány odhady skutečné hodnoty s_n ze vzorce (1) pro různé rozsahy náhodného výběru z normálního rozdělení $\mathcal{N}(3,9)$ jednak metodou Monte Carlo, tj. opakováním náhodného výběru $X_1,\dots,X_n$, jednak metodou bootstrap. Hodnota $\widetilde{s}_n$ je hodnota spočtená metodou Monte Carlo z 10 000 opakování náhodného výběru $\mathcal{N}(3,9)$ o rozsahu n. Hodnota s_{boot} značí průměrnou hodnotu bootstrapového odhadu založeného na $B = 500$ realizacích bootstrapového výběru o rozsahu n spočtenou pro 10 000 simulačních experimentů.

Dále se zabývejme odhadem distribuční funkce statistiky

$$R_n = \sqrt{n}(T_n - \theta) = \sqrt{n}(e^{\overline{X}_n} - e^{\mu}). \tag{2}$$

Předpokládáme-li stále, že $X_i \sim \mathcal{N}(\mu,\sigma^2)$, zjistíme snadno, že R_n má distribuční funkci

$$H_n(x) = \Phi\Big(\big(\ln\big(\frac{x}{\sqrt{n}} + e^{\mu}\big) - \mu\big)\frac{\sqrt{n}}{\sigma}\Big) \tag{3}$$

a hustotu

$$h_n(x) = \frac{\sqrt{n}}{\sigma(x + e^{\mu}\sqrt{n})}\varphi\Big(\big(\ln\big(\frac{x}{\sqrt{n}} + e^{\mu}\big) - \mu\big)\frac{\sqrt{n}}{\sigma}\Big), \tag{4}$$

kde Φ a φ značí distribuční funkci a hustotu $\mathcal{N}(0,1)$.

Pokud bychom rozdělení náhodných veličin $X_1,\dots,X_n$ neznali, mohli bychom se pokusit nalézt asymptotické rozdělení. Z Taylorova rozvoje dostaneme, že

$$R_n = \sqrt{n}(e^{\overline{X}_n} - e^{\mu}) = \sqrt{n}(\overline{X}_n - \mu)e^{\mu} + o_p(1),$$

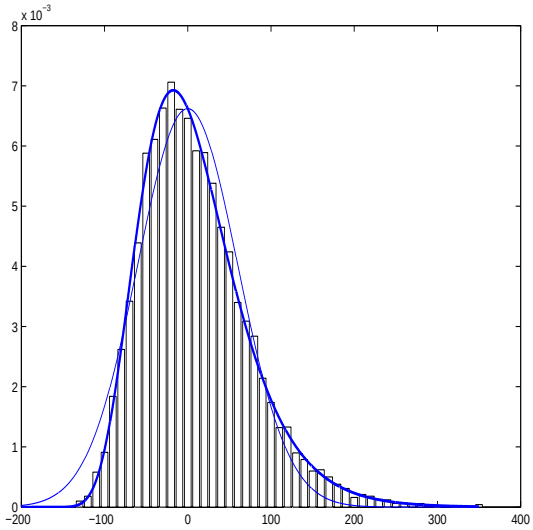
tedy R_n má asymptoticky normální rozdělení s nulovou střední hodnotou a rozptylem $e^{2\mu}\sigma^2$.

Další z možností je použít bootstrap. Nechť $X_1^*,\dots,X_n^*$ je bootstrapový výběr pořízený z pozorování $X_1,\dots,X_n$. Potom $X_1^*\dots,X_n^*$ jsou *iid,* pro které platí

$$E^*X_1^* = \mu^* = \frac{1}{n}\sum_{j=1}^{n} X_j = \overline{X_n}, \quad \mathrm{var}^*X_1^* = \sigma^{2\,*} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2,$$

a bootstrapová statistika je

$$R_n^* = \sqrt{n}(T_n^* - \theta^*) = \sqrt{n}(e^{\overline{X_n^*}} - e^{\mu^*}) = \sqrt{n}(e^{\overline{X_n^*}} - e^{\overline{X_n}}). \tag{5}$$



Obrázek 1: Hustota statistiky $\sqrt{n}(e^{\overline{X}_n} - e^{\mu})$: skutečná (tučná čára), asymptotická (slabá čára) a bootstrapová (histogram).

Na obrázku 1 je znázorněna hustota skutečného rozdělení statistiky (2) spočtená podle (4) pro náhodný výběr o rozsahu $n = 100$ z normálního rozdělení $\mathcal{N}(3, 9)$ (silnou čarou), hustota asymptotického normálního rozdělení (slabou čarou) a histogram odpovídající bootstrapové statistiky (5) pro $B = 10\,000$ bootstrapových výběrů.

Úvahy, které jsme dosud provedli, se dají zobecnit na vícerozměrný případ. Jsou-li $\mathbf{X}_1, \ldots, \mathbf{X}_n$ nezávislé stejně rozdělené náhodné vektory s distribuční funkcí F, lze spočítat empirickou distribuční funkci a definovat bootstrapový výběr $\mathbf{X}_1^*, \ldots, \mathbf{X}_n^*$ jako nezávislý náhodný výběr z rozdělení s touto empirickou distribuční funkcí. Bootstrapové odhady lze potom definovat zcela analogicky jako v jednorozměrném případě.

2 Vlastnosti bootstrapových aproximací

2.1 Přesnost aproximace rozdělení

Teoretické výsledky zkoumající přesnost bootstrapové aproximace rozdělení jsou založeny na centrálních limitních větách a jejich zpřesněních pomocí Berryovy- Esséenovy nerovnosti a Edgeworthova rozvoje pro normované, případně studentizované statistiky. Základní výsledky lze nalézt v pracech [16], [2], [1], [7], jejich shrnutí např. v [15].

Uveďme nejprve základní výsledky týkající se výběrového průměru. Uvažujme distribuční funkce výběrových statistik a jejich bootstrapové verze

$$H_n(x) = P(\sqrt{n}(\overline{X}_n - \mu) \le x), \quad H_n^*(x) = P^*(\sqrt{n}(\overline{X}_n^* - \mu^*) \le x),$$

$$\widetilde{H}_n(x) = P\Big(\sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} \le x\Big), \quad \widetilde{H}_n^*(x) = P^*\Big(\sqrt{n}\frac{\overline{X}_n^* - \mu^*}{\sigma^*} \le x\Big),$$

kde μ, σ^2 jsou střední hodnota a rozptyl a $\mu^* = \overline{X}_n$ a $\sigma^{2\,*} = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2$ jsou příslušné bootstrapové protějšky. Potom lze zformulovat následující velmi důležité tvrzení.

2.1.1 Věta Nechť $X_1, \dots, X_n$ jsou nezávislé stejně rozdělené náhodné veličiny s distribuční funkcí F.

(i) Jestliže $EX_1^2 < \infty$, potom

$$\rho_\infty(H_n^*, H_n) \xrightarrow[n\to\infty]{\text{s. j.}} 0.$$

(ii) Jestliže $EX_1^4 < \infty$, potom

$$\overline{\lim_{n\to\infty}} \frac{\sqrt{n}\rho_\infty(H_n^*, H_n)}{\sqrt{\log\log n}} = \frac{\sqrt{\mathrm{var}(X_1 - \mu)^2}}{2\sigma^2\sqrt{2\pi e}} \quad \text{s. j.}$$

(iii) Jestliže $E|X_1|^3 < \infty$ a F je řešetovitá, tj. existují konstanty c, h takové, že $\sum_{k=-\infty}^{\infty} P(X_1 = c + kh) = 1$, potom

$$\overline{\lim_{n\to\infty}} \sqrt{n}\rho_\infty(\widetilde{H}_n^*, \widetilde{H}_n) = \frac{h}{\sqrt{2\pi}\sigma} \quad \text{s. j.}$$

(iv) Jestliže $E|X_1|^3 < \infty$ a F není řešetovitá, potom

$$\sqrt{n}\rho_\infty(\widetilde{H}_n^*, \widetilde{H_n}) \xrightarrow[n\to\infty]{\text{s. j.}} 0.$$

Důkaz. Viz [16].

Tvrzení (i) je důkaz konzistence pro nestandardizované statistiky, tvrzení (ii) udává rychlost této konvergence. Tvrzení (iii) a (iv) udávají rychlost konvergence bootstrapové aproximace rozdělení standardizovaného výběrového průměru. Porovnejme ji s rychlostí aproximace normálním rozdělením. Ta je pro řešetovitá i neřešetovitá rozdělení podle Berryovy-Esséenovy nerovnosti řádu $O(n^{-\frac{1}{2}})$ a nedá se zlepšit. Pro neřešetovitá rozdělení je tudíž podle (iv) bootstrapová aproximace lepší.

Dále uveďme podobné výsledky pro hladké funkce výběrového průměru. Lze totiž ukázat, že mnoho výběrových statistik je možno přepsat jako funkci výběrového průměru nějakých náhodných vektorů.

Mějme nezávislé stejně rozdělené p-rozměrné náhodné vektory $\mathbf{X}_1, \ldots, \mathbf{X}_n$ s distribuční funkcí F, se střední hodnotou $\boldsymbol{\mu}$ a varianční maticí $\boldsymbol{\Sigma}$. Dále uvažujme funkci g z $\mathbb{R}^p$ do $\mathbb{R}$, $\nabla g(\mathbf{x}) := l(\mathbf{x})$ nechť je gradient g spočtený v bodě $\mathbf{x}$. Označme

$$s^2 = l^T(\boldsymbol{\mu})\boldsymbol{\Sigma}\, l(\boldsymbol{\mu}), \qquad s^{2*} = l^T(\boldsymbol{\mu}^*)\boldsymbol{\Sigma}^* l(\boldsymbol{\mu}^*),$$

$$S_n^2 = l^T(\overline{\boldsymbol{X}}_n)\boldsymbol{\Sigma}_n l(\overline{\boldsymbol{X}}_n), \quad S_n^{2*} = l^T(\overline{\boldsymbol{X}}_n{}^*)\boldsymbol{\Sigma}_n^* l(\overline{\boldsymbol{X}}_n{}^*),$$

kde

$$\boldsymbol{\mu}^* = \overline{\boldsymbol{X}}_n, \quad \boldsymbol{\Sigma}^* = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_n)(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_n)^T = \boldsymbol{\Sigma}_n.$$

Nyní uvažujme distribuční funkce statistik

$$H_n(x) = P(\sqrt{n}(g(\overline{\boldsymbol{X}}_n) - g(\boldsymbol{\mu})) \le x),$$
$$H_n^*(x) = P^*(\sqrt{n}(g(\overline{\boldsymbol{X}}_n^*) - g(\boldsymbol{\mu}^*)) \le x),$$

$$\widetilde{H}_n(x) = P\Big(\sqrt{n}\frac{g(\overline{\boldsymbol{X}}_n) - g(\boldsymbol{\mu})}{s} \le x\Big), \quad \widetilde{H}_n^*(x) = P^*\Big(\sqrt{n}\tfrac{g(\overline{\boldsymbol{X}}_n^*) - g(\boldsymbol{\mu}^*)}{s^*} \le x\Big),$$
$$\widehat{H}_n(x) = P\Big(\sqrt{n}\frac{g(\overline{\boldsymbol{X}}_n) - g(\boldsymbol{\mu})}{S_n} \le x\Big), \quad \widehat{H}_n^*(x) = P^*\Big(\sqrt{n}\tfrac{g(\overline{\boldsymbol{X}}_n^*) - g(\boldsymbol{\mu}^*)}{S_n^*} \le x\Big).$$

2.1.2 Věta Nechť $\mathbf{X}_1, \ldots, \mathbf{X}_n$ jsou *iid* p-rozměrné vektory se střední hodnotou $\boldsymbol{\mu}$ a varianční maticí $\boldsymbol{\Sigma}$. Nechť g je funkce z $\mathbb{R}^p$ do $\mathbb{R}$.

(i) Nechť $\mathrm{E}\|\mathbf{X}_1\|^2 < \infty$, nechť g je spojitě diferencovatelná v $\boldsymbol{\mu}$ a $l(\boldsymbol{\mu}) \neq \mathbf{0}$. Potom
$$\rho_\infty(H_n^*, H_n) \xrightarrow[n\to\infty]{\text{s. j.}} 0.$$

(ii) Nechť $\mathrm{E}\|\mathbf{X}_1\|^3 < \infty$, nechť pro charakteristickou funkci $\phi(\boldsymbol{t})$ náhodného vektoru $\boldsymbol{X}_1$ platí $|\phi(\boldsymbol{t})| < 1$ pro $\boldsymbol{t} \neq \mathbf{0}$. Nechť g je třikrát spojitě diferencovatelná na okolí $\boldsymbol{\mu}$ a $l(\boldsymbol{\mu}) \neq 0$. Potom pro distribuční funkce standardizovaných statistik platí
$$\sqrt{n}\rho_\infty(\widetilde{H}_n^*, \widetilde{H}_n) \xrightarrow[n\to\infty]{\text{s. j.}} 0.$$

Pro distribuční funkce studentizovaných statistik platí
$$\sqrt{n}\rho_\infty(\widehat{H}_n^*, \widehat{H}_n) \xrightarrow[n\to\infty]{\text{s. j.}} 0.$$

Důkaz. Viz [1].

2.2 Redukce vychýlení odhadu bootstrapem

Nechť $X_1, \ldots, X_n$ je náhodný výběr z rozdělení s konečnou střední hodnotou μ a rozptylem σ^2. Nechť g je spojitá funkce, taková, že $E\,|g(\overline{X}_n)| < \infty$; uvažujme parametr $\theta = g(\mu)$. Víme, že výběrový průměr $\overline{X}_n$ je nestranný a konzistentní odhad parametru μ, tj. $E\,\overline{X}_n = \mu$ a $\overline{X}_n \to \mu$ skoro jistě. Potom $g(\overline{X}_n) \to g(\mu)$ skoro jistě, tj. $g(\overline{X}_n)$ je konzistentní odhad $g(\mu)$, ale obecně je vychýlený, neboť $Eg(\overline{X}_n) \neq g(\mu)$, pokud g není lineární.

Studujme velikost vychýlení $b_n = Eg(\overline{X}_n) - g(\mu)$. Nadále předpokládejme, že g je dostatečně hladká funkce a X_i mají konečné momenty takového řádu, že platí Taylorův rozvoj

$$g(\overline{X}_n) - g(\mu) = (\overline{X}_n - \mu)g'(\mu) + \frac{1}{2}(\overline{X}_n - \mu)^2 g''(\mu) + R_n, \tag{6}$$

kde $ER_n = O(n^{-2})$ (viz např. [5, kap. 5.4]). Označíme-li

$$B_n = \frac{1}{2}\frac{\sigma^2}{n} g''(\mu),$$

potom

$$b_n = E(g(\overline{X}_n) - g(\mu)) = \frac{1}{2}\frac{\sigma^2}{n} g''(\mu) + O(n^{-2}) = B_n + O(n^{-2}) = O(n^{-1}).$$

Nyní uvažujme bootstrap. Je-li $X_1^*, \ldots, X_n^*$ bootstrapový výběr, potom bootstrapová verze (6) je

$$g(\overline{X}_n^*) - g(\overline{X}_n) = (\overline{X}_n^* - \overline{X}_n)g'(\overline{X}_n) + \frac{1}{2}(\overline{X}_n^* - \overline{X}_n)^2 g''(\overline{X}_n) + R_n^*,$$

kde díky silnému zákonu velkých čísel platí $E^* R_n^* = O(n^{-2})$ skoro jistě. Pro bootstrapové vychýlení $b_n^* = E^* g(\overline{X}_n^*) - g(\overline{X}_n)$ tak máme

$$E^*(g(\overline{X}_n^*) - g(\overline{X}_n)) = \frac{1}{2}\frac{\sigma^{*2}}{n} g''(\overline{X}_n) + O(n^{-2}) \quad \text{s. j.}$$

Z dalšího Taylorova rozvoje dostaneme $g''(\overline{X}_n) = g''(\mu) + \frac{1}{2}(\overline{X}_n - \mu)g'''(\mu) + \widetilde{R}_n$ a odtud spočteme

$$Eb_n^* = E\Big(\frac{1}{2n^2}\sum_{j=1}^{n}(X_i - \overline{X}_n)^2 g''(\mu)\Big) + O(n^{-2}) = B_n + O(n^{-2}).$$

Uvažujme nyní místo odhadu $g(\overline{X}_n)$ jeho opravu $g(\overline{X}_n) - b_n^*$. Vychýlení tohoto opraveného odhadu je

$$\begin{aligned} E[g(\overline{X}_n) - b_n^*] - g(\mu) &= b_n - Eb_n^* \\ &= B_n + O(n^{-2}) - B_n + O(n^{-2}) = O(n^{-2}), \end{aligned}$$

což je ve srovnání s b_n řádově lepší výsledek.

odhad	B_n	redukce (%)	$rmse$
$\widehat{\theta}_{20}$	5,186	25,821	19,675
$\widehat{\theta}^c_{20}$	-1,011	5,034	14,712
$\widehat{\theta}_{50}$	1,863	9,275	9,941
$\widehat{\theta}^c_{50}$	-0,145	0,721	8,893
$\widehat{\theta}_{100}$	0,940	4,682	6,514
$\widehat{\theta}^c_{100}$	-0,041	0,207	6,195

Tabulka 2: Redukce vychýlení odhadu $T_n = e^{\bar{X}_n}$ metodou bootstrap.

V tabulce 2 jsou uvedeny výsledky simulační studie, která srovnává vychýlení $b_n = E\widehat{\theta}_n - \theta$, kde $\theta = e^{\mu}, \widehat{\theta}_n = e^{\overline{X}_n}$, a vychýlení opraveného odhadu $\widehat{\theta}^c_n = \widehat{\theta}_n - b^*_n$, kde $b^*_n = E^*\widehat{\theta}^*_n - \widehat{\theta}_n$, v závislosti na rozsahu náhodného výběru. Náhodné výběry byly generovány z normálního rozdělení $\mathcal{N}(3, 9)$, skutečná hodnota $\theta = 20,086$. Pro každý výběr o rozsahu n bylo generováno 500 bootstrapových výběrů. Ve sloupci B_n je uveden vždy rozdíl odhadnuté a skutečné hodnoty, ve sloupci *redukce* je podíl $\frac{|B_n|}{\theta}$, ve sloupci *rmse* odmocnina ze střední kvadratické chyby pro 10 000 simulačních experimentů (převzato z [18]).

2.3 Intervaly spolehlivosti

2.3.1 Studentizované intervaly spolehlivosti. Označme jako $\widehat{\theta}_n$ odhad parametru θ a uvažujme studentizovanou statistiku

$$R_n = \frac{\widehat{\theta}_n - \theta}{S_n}$$

a její bootstrapový protějšek

$$R^*_n = \frac{\widehat{\theta}^*_n - \theta_n}{S^*_n}.$$

Je-li H_n distribuční funkce R_n a γ_p je příslušný $p-$kvantil, potom interval spolehlivosti pro θ s koeficientem $1 - \alpha$ je

$$\big(\widehat{\theta}_n - \gamma_{1-\frac{\alpha}{2}} S_n,\ \widehat{\theta}_n - \gamma_{\frac{\alpha}{2}} S_n\big).$$

Bootstrapový interval spolehlivosti je

$$\big(\widehat{\theta}_n - \gamma^*_{1-\frac{\alpha}{2}}\, S_n,\ \widehat{\theta}_n - \gamma^*_{\frac{\alpha}{2}}\, S_n\big),$$

kde γ^*_p je $p-$kvantil distribuční funkce H^*_n statistiky R^*_n spočtený jako $\gamma^*_p = R^*_{([B\,p])}$, tj. výběrový kvantil spočtený z uspořádaného výběru $R^*_{(1)}, \ldots, R^*_{(B)}$.

skutečný	(5,4634; 30,4683)
asymptotický	(9,7289; 33,3495)
hybridní	(5,3061; 30,6092)
percentilový	(12,4691; 37,7722)

Tabulka 3: 95%−ní intervaly spolehlivosti parametru e^{μ} pro výběr o rozsahu 100 z $\mathcal{N}(3,9)$, použito 10 000 bootstrapových výběrů.

Lze ukázat [7, kap. 3 a 5], že takto sestrojené intervaly pokrývají neznámý parametr s přesností, která je lepší než klasické intervaly využívající asymptotickou normalitu statistiky R_n. Nevýhodou tohoto postupu je velký počet výpočetních operací. Bootstrapový odhad S_n^* totiž většinou hledáme metodou Monte Carlo; k tomu potřebujeme B_1 bootstrapových výběrů. Dalších B_2 výběrů potřebujeme pro odhad kvantilů, tedy celkem $B_1 \cdot B_2$ výběrů. Je-li $B_1 = 200$ a $B_2 = 1000$, budeme potřebovat 100 000 bootstrapových výběrů.

2.3.2 Percentilové intervaly. Tato metoda počítá intervalový odhad parametru θ pomocí kvantilů distribuční funkce nestandardizované statistiky $\widehat{\theta}_n^*$, tj. z distribuční funkce $G_n^*(x) = P^*(\widehat{\theta}_n^* \le x)$. Dostaneme intervalový odhad

$$\Big(G_n^{*-1}(\frac{\alpha}{2}),\ G_n^{*-1}\big(1-\frac{\alpha}{2}\big)\Big).$$

Tato metoda se hodí v případě, že neznámý parametr je kvantil distribuční funkce uvažovaného náhodného výběru.

2.3.3 Hybridní intervaly. Je-li $H_n(x) = P(\sqrt{n}(\widehat{\theta}_n - \theta) \le x)$, je interval spolehlivosti pro θ s koeficientem $1-\alpha$

$$\Big(\widehat{\theta}_n - c_{1-\frac{\alpha}{2}}\frac{1}{\sqrt{n}},\ \widehat{\theta}_n - c_{\frac{\alpha}{2}}\frac{1}{\sqrt{n}}\Big),$$

kde c_p je $p-$ kvantil distribuční funkce H_n.

Je-li $H_n(x) \approx H_n^*(x) = P^*(\sqrt{n}(\widehat{\theta}_n^* - \widehat{\theta}_n) \le x)$, můžeme místo neznámých kvantilů c_p uvažovat odpovídající kvantily c_p^* distribuční funkce H_n^* a potom dostaneme interval spolehlivosti

$$\Big(\widehat{\theta}_n - c^*_{1-\frac{\alpha}{2}}\frac{1}{\sqrt{n}},\ \widehat{\theta}_n - c^*_{\frac{\alpha}{2}}\frac{1}{\sqrt{n}}\Big).$$

V tabulce 3 jsou pro ilustraci uvedeny intervaly spolehlivosti pro parametr $\theta = \mathrm{e}^{\mu}$ z příkladu 1.0.1. Skutečné intervaly jsou založeny na kvantilech distribuční funkce statistiky $R_n = \sqrt{n}(e^{\overline{X}_n} - e^{\mu})$.

Více podrobností o konstrukci intervalových odhadů lze nalézt např. v knize [15].

2.4 Některé výpočetní aspekty

2.4.1 Volba počtu bootstrapových výběrů. Současné počítače jsou dostatečně rychlé, abychom ve většině případů mohli zvolit počet opakování bootstrapových výběrů B mnohem větší než rozsah n. Větší počet opakování však často nepřináší další podstatné zlepšení našich výsledků. Chyba aproximace metodou Monte Carlo, kterou používáme pro pořízení bootstrapových statistik, by však měla být zanedbatelná vzhledem k chybě teoretického bootstrapového odhadu.

Zabývejme se dvěma základními okruhy problémů, totiž odhady rozptylu a odhady distribuční funkce (podrobněji viz [15, kap. 5.4]). Označme

$$\widehat{\text{var}}^{\,*} T_n^* = \frac{1}{B}\sum_{b=1}^{B}\left(T_{n,b}^* - \frac{1}{B}\sum_{k=1}^{B} T_{n,k}^*\right)^2 := v_{boot}^B. \tag{7}$$

Lze ukázat, že pro bootstrapový koeficient variace pro $s_{boot}^B = \sqrt{v_{boot}^B}$ přibližně platí

$$c = CV^*(s_{boot}^B) = \frac{\sqrt{Var^* s_{boot}^B}}{E^* s_{boot}^B} \approx \frac{1}{\sqrt{2B}},$$

když se zanedbá bootstrapový koeficient šikmosti. Potom pro požadovanou míru variability c je $B = \frac{1}{2}c^{-2}$, např. pro $c = 5\%$ je $B = 200$. Ukazuje se, že pro varianční odhady typu (7) se přesnost příliš nezlepší velkým počtem opakování, tj. zvětšováním B. Obecně se pro odhady momentů doporučuje volit B mezi 200-600, podle jiných doporučení však stačí jen 50-200.

Distribuční funkci H_n statistiky R_n odhadujeme jako

$$\widehat{H_n^*}(x) = \frac{1}{B}\sum_{b=1}^{B} I\{R_n(X_{1,b}^*, \ldots, X_{n,b}^*, F_n) \le x\} := H_{boot}^B(x). \tag{8}$$

Podle [15] je asymptotická chyba metody Monte Carlo

$$\rho_\infty(H_{boot}^B, H_n^*) = \epsilon_n + \sqrt{B^{-1}\log\log B},$$

kde $\epsilon_n = \rho_\infty(H_n, H_n^*)$ je chyba bootstrapové aproximace. Má-li být chyba Monte Carlo zanedbatelná vzhledem k ϵ_n, měli bychom volit B tak, aby $B^{-1}\log\log B = o(\epsilon_n^2)$. Pokud $\epsilon_n = O_P(n^{-1})$, je možno volit $B = n^2 \log\log n$. Pro $n = 30$ tak dostaneme přibližně $B \approx 1\,100$ opakování. Obecně pro odhady kvantilů a distribuční funkce se doporučuje $B = 1000$ jako minimální hodnota.

2.4.2 Redukce počtu operací. Existuje celá řada postupů jak zefektivnit a urychlit bootstrapové výpočty. Zmiňme např. *rovnovážný bootstrap,* podle kterého lze B bootstrapových výběrů pořídit následujícím postupem.

Nejprve se každé pozorování zkopíruje právě $B-$krát; dostaneme posloupnost délky nB. Nyní se provede náhodná permutace na prvky této posloupnosti. Prvky s pořadím $1,\ldots,n$ této zpermutované posloupnosti tvoří bootstrapový výběr $X^*_{1,1},\ldots,X^*_{n,1}$, prvky s pořadím $n+1,\ldots,2n$ bootstrapový výběr $X^*_{1,2},\ldots,X^*_{n,2}$, atd, prvky s pořadím $(n-1)B+1,\ldots,nB$ bootstrapový výběr $X^*_{1,B},\ldots,X^*_{n,B}$.

Další užívané techniky Monte Carlo pro bootstrap, např. *importance resampling*, jsou popsány v [7] a [15].

2.4.3 Odlehlá pozorování Odlehlá pozorování mohou ovlivnit bootstrapové výpočty. Je-li mezi hodnotami $X_1,\ldots,X_n$ jedno odlehlé pozorování, potom pravděpodobnost, že nebude obsaženo v bootstrapovém výběru, je $(1-\frac{1}{n})^n$, což pro velká n bude přibližně e^{-1}, tj. asi 37%. Ukazuje se ale, že histogram hodnot bootstrapových průměrů $\overline{X}^*_n$ není příliš citlivý na přítomnost odlehlých pozorování. Naopak histogram statistiky $\overline{X}^*_n - \overline{X}^*_n(k)$, kde

$$\overline{X}^*_n(k) = \frac{1}{n-2k}\sum_{i=k+1}^{n-k} X^*_{(i)}$$

je bootstrapová verze k-useknutého výběrového průměru, je velmi citlivý na odlehlá pozorování a přítomnost takových pozorování se projeví v tom, že histogram statistiky $\overline{X}^*_n(k)$ nebude unimodální. Statistiku $\overline{X}^*_n - \overline{X}^*_n(k)$ lze užít k detekci odlehlých pozorování. Teoretická zdůvodnění lze nalézt např. v článku [17].

2.5 Meze použitelnosti metody bootstrap

V předchozích odstavcích jsme se pokusili vysvětlit některé přednosti metody bootstrap, zejména schopnost redukovat vychýlení odhadů a zpřesňovat rozdělení statistik a tedy i přesnost pokrytí neznámých parametrů konfidenčními intervaly pro pivotální, resp. studentizované statistiky. Od r. 1979, kdy vyšel základní článek o bootstrapu [4], byla teoreticky zkoumána použitelnost a konzistence metody bootstrap v mnoha nejrůznějších statistických úlohách. Obecné teoretické výsledky, za kterých je bootstrap konzistentní, jsou uvedeny např. v knize [11].

Zmiňme se aspoň o několika příkladech, které ukazují, že bootstrap nelze používat automaticky.

- Pro rozdělení s nekonečným rozptylem neodhaduje bootstrap konzistentně rozdělení hladkých funkcí výběrového průměru.
- Bootstrap neodhaduje konzistentně rozdělení odhadů parametrů, které leží na hranici parametrické množiny.
- Bootstrap neodhaduje konzistentně rozdělení extremálních odhadů.

Řadu dalších příkladů lze nalézt např. v [15]. Pro většinu nich lze dokázat, že metoda je konzistentní pro bootstrapové výběry rozsahu m, pro které $\frac{m}{n}\to 0$ při $m,n\to\infty$.

- Jestliže X_n nejsou nezávislé, bootstrapové statistiky nedávají konzistentní odhady parametrů a rozdělení. V tomto případě je třeba metodu bootstrap modifikovat tak, aby bootstrapový výběr odrážel závislostní strukturu.

3 Bootstrap pro závislá pozorování

Existuje celá řada modifikací metody bootstrap pro závislá pozorování, které se liší tím, jakým způsobem využívají informaci o procesu, kterým jsou generována data.

3.1 Bootstrap závislý na modelu (parametrický)

Předpokládejme, že proces, kterým jsou generována data, je specifikován až na nějaké parametry. Nejčastěji používanou bootstrapovou technikou je tzv. *reziduální bootstrap*. Tato varianta se hodí v případě, že data jsou identifikována jako parametrický model typu AR nebo ARMA, případně jako dynamický regresní model. Předpokládejme např., že náhodné veličiny $X_1, \ldots, X_n$ se řídí autoregresním modelem AR(p)

$$X_t = \beta_1 X_{t-1} + \cdots + \beta_p X_{t-p} + Y_t,\ t = 1, \ldots, n, \tag{9}$$

kde Y_t jsou *iid* s nulovou střední hodnotou a konečným kladným rozptylem σ^2, $X_0, \ldots, X_{1-p}$ jsou počáteční pozorování a jsou splněny předpoklady pro stacionaritu posloupnosti $\{X_t\}$. Bootstrap v klasické podobě zde použít nelze, neboť $X_1, \ldots, X_n$ nejsou nezávislé. Nezávislé chyby Y_t většinou neznáme, ale můžeme je odhadnout. Použijeme-li konzistentní odhady pro neznámé parametry $\beta_1, \ldots, \beta_p$, potom odhadnutá rezidua se chovají přibližně jako nezávislé náhodné veličiny. Generování bootstrapového výběru potom probíhá podle následujícího algoritmu, který lze zobecnit i na ARMA modely.

- Nejprve se odhadnou rezidua $\widehat{Y}_t = X_t - \widehat{\beta}_1 X_{t-1} - \cdots - \widehat{\beta}_p X_{t-p}$, kde $\widehat{\beta}_1, \ldots, \widehat{\beta}_p$ jsou konzistentní odhady parametrů $\beta_1, \ldots, \beta_p$.
- Spočte se empirická distribuční funkce F_n centrovaných reziduí $\widehat{Y}_t - \overline{Y}_n$, kde $\overline{Y}_n$ je aritmetický průměr z hodnot $\widehat{Y}_1, \ldots, \widehat{Y}_n$. Centrování je nezbytné, pokud nemáme model s interceptem.
- Generují se náhodné veličiny $Y_1^*, \ldots Y_n^*$, které jsou (podmíněně při daných $X_1, \ldots, X_n$) *iid* a mají distribuční funkci F_n.
- Bootstrapový výběr, který kopíruje strukturu původních dat je generován předpisem

$$X_t^* = \widehat{\beta}_1 X_{t-1}^* + \cdots + \widehat{\beta}_p^* X_{t-p}^* + Y_t^*, \quad t = 1, \ldots, n,$$

kde se položí $X_{1-p}^* = 0, \ldots, X_0^* = 0$ nebo $X_{1-p}^* = X_{1-p}, \ldots, X_0^* = X_0$.

Bose [3] dokázal konzistenci této metody a její zpřesnění proti normální aproximaci pro rozdělení bootstrapových odhadů pro parametry $\beta_1, \ldots, \beta_p$ metodou nejmenších čtverců, Prášková [12] rozšířila tento výsledek na hladké

funkce průměrů jistých vektorů a z nich plynoucí studentizované odhady. Kreiss a Franke [9] dokázali konzistenci metody bootstrap pro třídu M-odhadů v modelech ARMA. Přehled dalších výsledků, zabývajících se použitím varianty reziduální bootstrap v modelech AR a ARMA lze nalézt např. v [10].

Další používanou metodou je tzv. *wild bootstrap*. Tato modifikace metody bootstrap byla původně odvozena pro regresní modely s heterogenními chybami, např. [19], [11]. Lze ji však aplikovat i na časové řady v situacích, kdy je identifikován parametrický model (např. ARMA), ve kterém ale nelze šum modelovat jako posloupnost nezávislých stejně rozdělených náhodných veličin. Sem patří i v současné době velmi populární modely s podmíněnou heteroskedasticitou (modely typu ARCH, GARCH), nebo autoregresní modely s náhodnými parametry, které mají podobnou strukturu jako modely ARCH, viz např. [13], [14], [6].

Uvažujme opět autoregresní model AR(p) definovaný v (9), ve kterém Y_t nejsou nezávislé a stejně rozdělené náhodné veličiny. Mějme pozorování $X_1, \dots, X_n$ a uvažujme odhady autoregresních parametrů metodou nejmenších čtverců, které jsou výpočetně velmi jednoduché. Vzhledem k obecné nestacionaritě však lze obtížně určit jejich asymptotické rozdělení. Odhad rozdělení metodou reziduální bootstrap není v tomto případě konzistentní (např. [13]).

V případě, že Y_t jsou nezávislé nebo slabě závislé, ale heterogenní, lze použít techniku wild bootstrap, který zachovává heteroskedasticitu. Možné jsou dvě varianty této metody.

První z nich generuje bootstrapový výběr na základě regrese. Spočtou se rezidua $r_t = X_t - \widehat{\beta}_1 X_{t-1} - \cdots - \widehat{\beta}_p X_{t-p}$, kde $\widehat{\beta}_1, \dots, \widehat{\beta}_p$ jsou odhady metodou nejmenších čtverců. Dále se generuje nový proces chyb

$$Y_t^w = r_t K_t, \quad t = 1, \dots, n,$$

kde K_t jsou *iid* s nulovou střední hodnotou a jednotkovým rozptylem, nezávislé na $X_0, \dots, X_n$. Potom se generuje bootstrapový výběr podle schématu

$$X_t^w = \widehat{\beta}_1 X_{t-1} + \cdots + \widehat{\beta}_p X_{t-p} + Y_t^w, \quad t = 1, \dots, n.$$

Bootstrapové hodnoty se tedy řídí regresním modelem s konstantními regresory $X_{t-1}, \dots, X_{t-p},\ t = 1, \dots, n$.

Ve druhé variantě se generuje proces chyb $Y_t^w = r_t K_t,\ t = 1, \dots, n$, stejně jako v regresní variantě, ale bootstrapový výběr se generuje podle autoregresního schématu

$$X_{1-p}^{*w} = 0, \dots, X_0^{*w} = 0,$$
$$X_t^{*w} = \widehat{\beta}_1 X_{t-1}^{*w} + \cdots + \widehat{\beta}_p X_{t-p}^{*w} + Y_t^w,\ t = 1, \dots, n,$$

takže přesně kopíruje strukturu závislosti v původním modelu. Zde opět $\widehat{\beta}_1, \dots, \widehat{\beta}_p$ jsou odhady parametrů $\beta_1, \dots, \beta_p$ v původním modelu AR(p).

Lze ukázat, že obě tyto varianty konzistentně odhadují rozdělení odhadů $\widehat{\beta}_1, \dots, \widehat{\beta}_p$ získaných metodou nejmenších čtverců za různých podmínek na hetroskedasticitu chyb Y_t, viz [13], [14], [6].

3.2 Bootstrap nezávislý na modelu (blokový)

Nechť je k dispozici n pozorování časové řady $X_1, \ldots, X_n$; předpokládáme, že jde o stacionární posloupnost, ale o závislostní struktuře nemáme žádné informace.

Bootstrapový výběr se v takovém případě dá sestrojit následovně. Vektor $(X_1, \ldots, X_n)$ se nejdříve rozdělí na bloky délky l. Pro $n = k \cdot l$ tak dostaneme $N = k$ nepřekrývajících se bloků

$$\mathbf{Y}_1 = (X_1, \ldots, X_l), \mathbf{Y}_2 = (X_{l+1}, \ldots, X_{2l}), \ldots, \mathbf{Y}_k = (X_{(k-1)l+1}, \ldots, X_n).$$

Jinou možností je vytvořit $N = n - l + 1$ klouzavých bloků

$$\mathbf{Y}_1 = (X_1, \ldots, X_l), \mathbf{Y}_2 = (X_2, \ldots, X_{l+1}), \ldots, \mathbf{Y}_{n-l+1} = (X_{n-l+1}, \ldots, X_n).$$

Dále se provádí nezávislý náhodný výběr s vracením z populace vektorů $\mathbf{Y}_1, \ldots, \mathbf{Y}_N$. Dostáváme tak postupně vektory $\mathbf{Y}_1^*, \mathbf{Y}_2^* \ldots$. Za bootstrapový výběr potom považujeme vektor náhodných veličin

$$(X_1^*, \ldots, X_n^*) = (\mathbf{Y}_1^*, \mathbf{Y}_2^*, \ldots, \mathbf{Y}_k^*).$$

Existují i další varianty, jak vytvářet bloky. Důležité je, že v bootstrapovém výběru je vždy l po sobě jdoucích pozorování, která mají stejnou strukturu jako původní data.

Z teoretického hlediska, pro dosažení konzistentních výsledků, musí být délka bloku dostatečně dlouhá, $l \to \infty$ pro $n \to \infty$. Existují teoretická odvození pro stanovení optimální délky bloku (např. [10]). V praxi však nelze podle těchto teoretických výsledků vždy postupovat, neboť teoretické stanovení délky bloku závisí na hodnotách autokovarianční funkce, kterou neznáme. Většinou se tedy délka bloku stanoví tak, že se řada pozorování nejdříve odhadne nějakým parametrickým modelem, ve kterém je teoretická autokovarianční funkce známá, a její odhad se potom dosadí do vzorců pro výpočet délky bloku. Existují i jiné algoritmy, založené na konkrétních pozorováních.

Více o blokovém bootstrapu se lze dočíst v Lahiri [10] nebo např. v práci [8]. Tam je možno nalézt také další prameny, které zde pro nedostatek místa neuvádíme.

Reference

[1] Babu G.J., Singh K. (1984). *On one term Edgeworth correction by Efron's bootstrap.* Sankhyā A **46**, 219 – 232.

[2] Bickel P.J., Freedman D.A. (1981). *Some asymptotic theory for the bootstrap.* Ann. Statist. **9**, 1196 – 1217.

[3] Bose A. (1988). *Edgeworth correction by bootstrap in autoregression.* Ann. Statist. **16**, 1709 – 1722.

[4] Efron B. (1979). *Bootstrap methods: another look at the jackknife.* Ann. Statist. **7**, 1 – 26.

[5] Fuller W.A. (1976). *Introduction to statistical time series*. Wiley, New York.

[6] Gonçalves S., Kilian L. (2004). *Bootstrapping autoregression with conditional heteroskedasticity of unknown form*. J. Econometrics **123**, 89 – 120.

[7] Hall P. (1992). *The bootstrap and the Edgeworth expansion*. Springer-Verlag, New York.

[8] Härdle W., Horowitz J., Kreiss J.-P., (2003). *Bootstrap methods for time series*. International Statistical Review **71**, 435 – 459.

[9] Kreiss J.P., Franke J. (1992). *Bootstrapping stationary autoregressive – moving average models*. J. Time Ser. Anal. **13**, 297 – 317.

[10] Lahiri (2003). *Resampling methods for dependent data*. Springer-Verlag, New York.

[11] Mammen E. (1992). *When does bootstrap work? Asymptotic results and simulations*. Springer-Verlag, Heidelberg.

[12] Prášková Z. (1995). *A contribution to bootstrapping autoregressive processes*. Kybernetika **31**, 359 – 373.

[13] Prášková Z. (2002). *Bootstrap in nonstationary autoregression*. Kybernetika **38**, 389 – 404.

[14] Prášková Z. (2003). *Wild bootstrap in RCA(1) model*. Kybernetika **39**, 1 – 12.

[15] Shao J., Tu D. (1995). *The jackknife and bootstrap*. Springer-Verlag, New York.

[16] Singh K. (1981). *On the asymptotic accuracy of Efron's bootstrap*. Ann. Statist. **9**, 1187 – 1195.

[17] Singh K., Xie M. (2003). *Bootlier-plot - bootstrap based outlier detection plot*. Sankhyā **65**, 532 – 559.

[18] Šindlář J. (2003). *Počítačové postupy a výběry z konečné populace*. Diplomová práce MFF UK, Praha.

[19] Wu C.F.J. (1986). *Jackknife, bootstrap and other resampling methods in regression analysis (with discussions)*. Ann. Statist. **14**, 1261 – 1350.

Poděkování: Práce vznikla za podpory výzkumného záměru MŠMT číslo MSM 113200008 a grantu GAČR č. 201/03/0945.

Adresa: Z. Prášková, Univerzita Karlova, Fakulta matematicko-fyzikální, Katedra pravděpodobnosti a matematické statistiky, Sokolovská 83, 186 75 Praha 8

E-mail: praskova@karlin.mff.cuni.cz