Kybernetika

O STOCHASTICKÉ APROXIMACI

VÁCLAV DUPAČ

ACADEMIA

PRAHA

1. ÚVOD

Mějme za úkol nalézt řešení θ rovnice $M(x) = 0$, kde M je reálná funkce reálné proměnné. Z numerických metod se obyčejně na prvním místě uvádí metoda Newtonova, popsaná vztahem

$$x_{t+1} = x_t - M(x_t)/M'(x_t)\,;$$

ta však má dobré vlastnosti jen máme-li k dispozici dobré počáteční přiblížení a je samozřejmě nepoužitelná tehdy, jsme-li schopni numericky počítat jen funkční hodnoty $M(x)$ ale nikoli hodnoty derivace $M'(x)$. V takovém případě se nabízí jednoduchá iterační metoda

$$x_{t+1} = x_t - a\,M(x_t)\,, \tag{1.1}$$

kde a je konstanta.

Předpokládáme-li, že graf funkce M leží mezi dvěma přímkami o kladných směrnicích, procházejícími bodem $(\theta, 0)$, tj. je-li

$$M(x) \gtreqqless 0 \quad \text{pro} \quad x \gtreqqless \theta\,, \tag{1.2}$$

$$K_1|x - \theta| \leqq |M(x)| \leqq K_2|x - \theta| \tag{1.3}$$

pro nějaká $0 < K_1 \leqq K_2 < +\infty$ a všechna $x \in R$, potom *aproximace x_t získané dle* (1.1) *konvergují k řešení θ* pro každé počáteční přiblížení x_1, pokud $0 < a < 2/K_2$. Rychlost konvergence je aspoň geometrická s kvocientem

$$q = \max\,(1 - aK_1, aK_2 - 1)\,. \tag{1.4}$$

Tvrzení se dokáže snadno: dle (1.1) je

$$x_{t+1} - \theta = x_t - \theta - aM(x_t) = (x_t - \theta)\left(1 - a\,\frac{M(x_t)}{x_t - \theta}\right),$$

dle (1.2), (1.3) je

$$1 - aK_2 \leqq 1 - a\,\frac{M(x)}{x - \theta} \leqq 1 - aK_1, \quad \text{tj.}$$

$$\left|1 - a\,\frac{M(x)}{x - \theta}\right| \leqq \max\,(1 - aK_1, aK_2 - 1) \quad \text{pro všechna} \quad x \neq \theta\,,$$

odtud $|x_{t+1} - \theta| \leqq q|x_t - \theta|$, kde $q < 1$.

Uvažujme nyní situaci, kdy funkční hodnoty $M(x)$ zjišťujeme jen s nějakou pozorovací (nebo experimentální) chybou $e(x)$, tj. kdy výsledkem pozorování (experimentu) v bodě x není $M(x)$ nýbrž $Y(x) = M(x) + e(x)$. O chybě $e(x)$ předpokládáme, že je náhodná bez systematické složky, tj. že má nulovou střední hodnotu.

Zkusme v této situaci opět hledat θ pomocí iterací (1.1), s $Y_t = M(X_t) + e(t + 1, X_t)$ na místě $M(X_t)$. (Velkými písmeny značíme náhodné veličiny; u chyb jsme vyznačili, že jde o různé veličiny pro různá t.) Snadno zjistíme, že nyní nelze volit a jako jedinou konstantu, ale že je třeba volit posloupnost kladných konstant a_t, která konverguje k 0 a to tak, že řada $\sum_{t=1}^{\infty} a_t = +\infty$ zatímco $\sum_{t=1}^{\infty} a_t^2 < +\infty$.

Kdyby totiž a bylo konstantní nezávisle na t, pak by díky náhodným chybám rozdíly $X_{t+1} - X_t = aY_t$ nekonvergovaly k 0 a tedy ani posloupnost X_t by nebyla konvergentní. Musí tedy být $a_t \to 0$; kdyby však konvergence k 0 byla tak rychlá, že $\sum_{t=1}^{\infty} a_t = A < +\infty$, pak by aproximace X_t nemusely dospět k θ, mohly by se neomezeně blížit k jiné hodnotě než θ. Jako příklad může sloužit situace, kdy $Y(x)$ nabývá pro všechna x jen hodnot ± 1 a kdy počáteční přiblížení X_1 je od řešení θ vzdáleno o více než A. Konečně požadavek $\sum_{t=1}^{\infty} a_t^2 < +\infty$ se uplatní v průběhu důkazu konvergence X_t k θ s pravděpodobností 1.

Platí: Jsou-li náhodné chyby v jednotlivých krocích nezávislé a jejich rozptyly omezené, splňuje-li funkce M předpoklady (1.2), (1.3) a konstanty a_t požadavek

$$a_t > 0\,, \quad \sum_{1}^{\infty} a_t = +\infty\,, \quad \sum_{1}^{\infty} a_t^2 < +\infty\,,$$

pak *aproximace* definované předpisem X_1 libovolné, $\mathsf{E}X_1^2 < +\infty$, $X_{t+1} = X_t - a_t Y_t$, $1 \leqq t < +\infty$, kde $Y_t = M(X_t) + e(t+1, X_t)$, *konvergují k θ s pravděpodobností* 1 *i podle kvadratického středu.*

Popsaná stochastická aproximační metoda se nazývá *Robbinsova-Monroova*; poprvé byla studována v práci jmenovaných autorů [26].

Vyšetření její konvergence není obtížné (ani za podstatně obecnějších předpokladů), využijeme-li konvergenční věty pro supermartingaly. Konvergenci stochastických aproximací dokážeme hned pro případ vícerozměrný, tedy pro úlohu řešit soustavu rovnic, neboť mezi jednorozměrným a vícerozměrným případem zde není zásadní rozdíl.

Nejprve si připravíme potřebné prostředky. Podrobnější výklad látky obsažené v odst. 2,3,4 lze nalézt v monografii Nevel'sona a Has'minského [22].

2. POUŽÍVANÉ POJMY A VÝSLEDKY

Na pravděpodobnostním prostoru (Ω, A, P) mějme posloupnost σ-algeber $\{\mathscr{F}_t\}_{t=1}^{\infty}$ takových, že $\mathscr{F}_1 \subset \mathscr{F}_2 \subset \ldots \subset \mathscr{A}$, a posloupnost $\{X_t\}_{t=1}^{\infty}$ náhodných veličin s konečnými středními hodnotami takových, že X_t je měřitelná vzhledem k $\mathscr{F}_t$. Posloupnost $\{X_t, \mathscr{F}_t\}_{t=1}^{\infty}$ tvoří *martingal*, jestliže $\mathsf{E}(X_{t+1} \mid \mathscr{F}_t) = X_t$ s pravděpodobností 1, resp. *supermartingal*, jestliže $\mathsf{E}(X_{t+1} \mid \mathscr{F}_t) \leqq X_t$ s pravděpodobností 1, pro $1 \leqq t < +\infty$. Je-li $\mathscr{F}_t = \mathscr{F}_t^X$, což je nejmenší σ-algebra vzhledem ke které jsou $X_1, X_2, \ldots, X_t$ měřitelné, pak se místo $\{X_t, \mathscr{F}_t\}$ píše jen $\{X_t\}$ a místo $\mathsf{E}(X_{t+1} \mid \mathscr{F}_t)$ se užívá zápisu $\mathsf{E}(X_{t+1} \mid X_1, X_2, \ldots, X_t)$.

Významná věta – např. [3], VII, § 4 – říká, že každý *nezáporný supermartingal* $\{X_t, \mathscr{F}_t\}$ *má s pravděpodobností* 1 *konečnou limitu*, $\lim_{t\to\infty} X_t = X$. Zde X je obecně náhodná veličina, v našich úvahách však bude X většinou konstanta.

Pro všechna celá kladná $s, t, s \leqq t$, všechna $x \in R^l$ a všechna $A \in \mathscr{B}^l$ (= borelovské množiny v R^l) nechť je definována funkce $P(s, t, x, A)$, která jakožto funkce x je $\mathscr{B}^l$-měřitelná a jakožto funkce A je pravděpodobnostní míra na $\mathscr{B}^l$. Taková funkce se nazývá *přechodová funkce*.

Nechť $P(s, t, x, A)$ je nějaká přechodová funkce. Posloupnost $\{X_t\}_{t=1}^{\infty}$ l-rozměrných náhodných vektorů se nazývá *markovská* s touto přechodovou funkcí, jestliže pro všechna $1 \leqq s < t$ a všechna $A \in \mathscr{B}^l$ platí

$$\mathsf{P}(X_t \in A \mid X_1, X_2, \ldots, X_s) = P(s, X_s, t, A) \tag{2.1}$$

s pravděpodobností 1. Podmíněná pravděpodobnost na levé straně (2.1) je pak také rovna podmíněné pravděpodobnosti $\mathsf{P}(X_t \in A \mid X_s)$; odtud plyne názorný význam přechodové funkce:

$$\mathsf{P}(X_t \in A \mid X_s = x) = P(s, x, t, A),$$

resp. obecněji

$$\mathsf{E}(g(X_t) \mid X_s = x) = \int_{R^l} g(y)\, P(s, x, t, \mathrm{d}y). \tag{2.2}$$

Mějme markovskou posloupnost $\{X_t\}_{t=1}^{\infty}$ s přechodovou funkcí $P(s, x, t, A)$. *Vytvořující operátor* L této posloupnosti se definuje vztahem

$$LV(t, x) = \int_{R^l} [V(t + 1, y) - V(t, x)]\, P(t, x, t + 1, \mathrm{d}y); \tag{2.3}$$

jeho definičním oborem jsou funkce $V(t, x)$, $1 \leqq t < +\infty$, $x \in R^l$, pro něž je integrál v (2.3) konečný. Dle (2.2) lze také psát

$$LV(t, x) = \mathsf{E}(V(t + 1, X_{t+1}) - V(t, X_t) \mid X_t = x); \tag{2.4}$$

$LV(t, x)$ je tedy očekávaný přírůstek za 1 krok funkce $V(t, x)$ podél trajektorie markovské posloupnosti $\{X_t\}$ vycházející v čase t z bodu x. Všimněme si ještě, že z (2.4) plyne platnost vztahu

$$\mathsf{E}(V(t + 1, X_{t+1}) \mid X_t) = V(t, X_t) + LV(t, X_t) \tag{2.5}$$

s pravděpodobností 1.

Je-li funkce V taková, že $V(t, x) \geqq 0$ a $LV(t, x) \leqq 0$ pro všechna $1 \leqq t < +\infty$, $x \in R^l$ a je-li $\mathsf{E}\, V(1, X_1) < +\infty$, pak je $\{V(t, X_t), \mathscr{F}_t^X\}$ *nezáporný supermartingal*. Dle markovské vlastnosti je totiž

$$\mathsf{E}(V(t + 1, X_{t+1}) \mid \mathscr{F}_t^X) = \mathsf{E}(V(t + 1, X_{t+1}) \mid X_t)$$

a z (2.5) a z $LV \leqq 0$ dále plyne

$$\mathsf{E}(V(t+1, X_{t+1}) \mid X_t) \leqq V(t, X_t)$$

s pravděpodobností 1. Požadavek konečných středních hodnot $\mathsf{E}\, V(t, X_t)$ je zde splněn vzhledem k předpokladu $\mathsf{E}\, V(1, X_1) < +\infty$ a vzhledem k tomu, že posloupnost $\mathsf{E}\, V(t, X_t)$ je nerostoucí. Předpoklad $\mathsf{E}\, V(1, X_1) < +\infty$ je triviálně splněn, volíme-li $X_1 = x_1$ konstantní.

Pro zjednodušení dalších zápisů zavedeme označení

$$V_{\sim}(x) = \inf_{1 \leqq t < +\infty} V(t, x), \quad V^{\sim}(x) = \sup_{1 \leqq t < +\infty} V(t, x).$$

Věta 2.1. Nechť $\{X_t\}_{t=1}^{\infty}$ je markovská posloupnost s vytvořujícím operátorem L. Nechť funkce $V(t, x)$ a $\varphi(t, x)$ jsou nezáporné pro všechna $1 \leqq t < +\infty$, $x \in R^l$ a takové, že

(2.6) $$V_{\sim}(x) \to +\infty \quad \text{pro} \quad |x| \to +\infty;$$

(2.7) $$\text{existuje} \quad \theta \in R^l \quad \text{tak, že} \quad \varphi^{\sim}(\theta) = 0 \quad \text{a}$$

$$\inf_{\varepsilon < |x-\theta| < 1/\varepsilon} \varphi_{\sim}(x) < 0 \quad \text{pro každé} \quad \varepsilon > 0,$$

(2.8) $LV(t, x) \leqq -\alpha_t\, \varphi(t, x)$ pro všechna t, x a pro nějakou posloupnost

$$\text{kladných čísel } \alpha_t \text{ takovou, že } \sum_{t=1}^{\infty} \alpha_t = +\infty.$$

Potom

(i) existuje náhodná veličina $R < +\infty$ tak, že $|X_t| \leqq R$ pro všechna $1 \leqq t < +\infty$ s pravděpodobností 1,

(ii) $\sum_{k=1}^{\infty} \alpha_k\, \varphi(k, X_k) < +\infty$ s pravděpodobností 1,

(iii) θ je hromadný bod posloupnosti $\{X_t\}$ s pravděpodobností 1.

Důkaz. Protože v tvrzení věty nevystupují střední hodnoty, nýbrž jen výroky platící s pravděpodobností 1, postačí, dokážeme-li větu pro libovolné pevné $X_1 = x_1$.

(i) Dle předchozího textu je $\{V(t, X_t), \mathscr{F}_t^X\}$ nezáporný supermartingal, existuje tedy konečná $\lim_{t\to\infty} V(t, X_t)$. Kdyby posloupnost $\{X_t\}$ byla neomezená na nějaké množině kladné pravděpodobnostní míry, pak by pro každé ω z této množiny existovala posloupnost indexů $\{\tau_n(\omega)\}$ tak, že $|X_{\tau_n}| \to +\infty$ a dle (2.6) také $V_{\sim}(X_{\tau_n}) \to +\infty$ a tedy i $V(\tau_n, X_{\tau_n}) \to +\infty$, což je spor s existencí konečné $\lim_{t\to\infty} V(t, X_t)$.

(ii) Utvoříme-li střední hodnoty na levé i pravé straně rovnosti (2.5), sečteme tyto rovnosti od 1 do t a změníme znaménka, dostaneme

$$-\sum_{k=1}^{t} \mathsf{E} L V(k, X_k) = -\mathsf{E}\, V(t+1, X_{t+1}) + V(1, x_1) \leqq V(1, x_1) ;$$

na druhé straně dle (2.8) je

$$\mathsf{E}\left(\sum_{k=1}^{t} \alpha_k\, \varphi(k, X_k)\right) \leqq -\sum_{k=1}^{t} \mathsf{E} L V(k, X_k) ,$$

tedy $\mathsf{E}\left(\sum_{k=1}^{t} \alpha_k\, \varphi(k, X_k)\right) \leqq V(1, x_1)$, $\forall_{1 \leqq t < +\infty}$ a tedy i $\mathsf{E}\left(\sum_{k=1}^{\infty} \alpha_k\, \varphi(k, X_k)\right) \leqq V(1, x_1)$, podle pravidla o monotónní konvergenci. Je tedy nutně $\sum_{k=1}^{\infty} \alpha_k\, \varphi(k, X_k) < +\infty$ s pravděpodobností 1.

(iii) Kdyby existovala množina kladné pravděpodobnostní míry tak, že v každém jejím bodě by bylo $\varphi(t, X_t) \geqq \delta = \delta(\omega) > 0$ pro všechna t, pak by platilo i $\sum_{t=1}^{\infty} \alpha_t \,.\, \varphi(t, X_t) \geqq \delta \sum_{t=1}^{\infty} \alpha_t = +\infty$, což je spor s dokázaným již tvrzením (ii). Existuje tedy pro P-skoro všechna ω posloupnost indexů $\{\tau_n(\omega)\}$ tak, že $\varphi(\tau_n, X_{\tau_n}) \to 0$; vzhledem k (2.7) to však znamená, že $\{X_{\tau_n}\}$ může mít jen dva hromadné body: θ a nevlastní bod ∞. Případ ∞ ovšem nemůže nastat vzhledem k tvrzení (i). □

Místo vlastnosti (2.8) můžeme na funkci V klást snadněji ověřitelný požadavek

$$LV(t, x) \leqq g_t(1 + V(t, x)) - \alpha_t\, \varphi(t, x) \tag{2.9}$$

pro nějakou posloupnost kladných čísel g_t takovou, že $\sum_{t=1}^{\infty} g_t < +\infty$ (vše ostatní beze změny). Položíme-li totiž $W(t, x) = (1 + V(t, x)) \prod_{k=t}^{\infty} (1 + g_k)$, pak po úpravách dostaneme $L W(t, x) = (L V(t, x) - g_t(1 + V(t, x)) \prod_{k=t+1}^{\infty} (1 + g_k) \leqq -\alpha_t\, \varphi(t, x)$, tj. W má vlastnost (2.8). Protože W má zřejmě i ostatní vlastnosti požadované ve větě 2.1, platí i všechna tvrzení věty 2.1.

Připojíme-li ještě další předpoklady o funkci V, dostaneme již konvergenční větu.

Věta 2.2. Nechť $\{X_t\}_{t=1}^{\infty}$ je markovská posloupnost s vytvořujícím operátorem L. Nechť funkce $V(t, x)$ a $\varphi(t, x)$ jsou nezáporné a takové, že pro nějaké $\theta \in R^l$ platí

$$V^{\sim}(\theta) = 0 , \quad \lim_{x \to \theta} V^{\sim}(x) = 0 , \quad \inf_{|x-\theta| > \varepsilon} V_{\smile}(x) > 0 , \quad \forall_{\varepsilon > 0} , \tag{2.10}$$

$$\lim_{|x| \to \infty} V_{\smile}(x) = +\infty ,$$

a dále (2.7) a (2.9), kde $\alpha_t > 0, g_t > 0, \sum_{t=1}^{\infty} \alpha_t = +\infty, \sum_{t=1}^{\infty} g_t < +\infty$. Potom $\lim_{t \to \infty} X_t = \theta$ s pravděpodobností 1.

Poznámka. Všimněte si, že podmínka (2.7) připouští i funkce $\varphi(t, x)$ pro které je $\lim_{|x| \to \infty} \varphi(t, x) = 0$.

Důkaz. Opět stačí uvažovat $X_1 = x_1$ konstantní. Položíme-li $W(t, x) = (1 + V(t, x)) \prod_{k=t}^{\infty} (1 + g_k)$, pak podle předchozí úvahy je $LW(t, x) \leqq -\alpha_t \varphi(t, x) \leqq 0$, tedy $\{W(t, X_t), \mathscr{F}_t^X\}$ je nezáporný supermartingal a existuje konečná $\lim_{t \to \infty} W(t, X_t)$ a vzhledem k tomu jak je $W(t, x)$ definováno, také konečná $\lim_{t \to \infty} V(t, X_t)$. Podle tvrzení (iii) věty 2.1 je θ s pravděpodobností 1 hromadným bodem posloupnosti $\{X_t\}$ a protože předpokládáme $V^{\sim}(x) \to 0$ pro $x \to \theta$, je také 0 hromadným bodem posloupnosti $\{V(t, X_t)\}$. To však znamená, že $\lim_{t \to \infty} V(t, X_t) = 0$, tedy i $\lim_{t \to \infty} V_{\sim}(X_t) = 0$, a protože předpokládáme $\inf_{|x - \theta| > \varepsilon} V_{\sim}(x) > 0$, znamená to konečně, že $\lim_{t \to \infty} X_t = \theta$. □

Poznámka. Z důkazu věty vyplývá, že střední hodnoty $\mathsf{E}\, W(t, X_t)$ a tedy i $\mathsf{E}\, V(t, X_t)$ jsou omezené.

3. KONVERGENČNÍ VĚTY PRO ROBBINSOVU-MONROOVU METODU

Mějme funkci $M : R^l \to R^l$, měřitelnou vzhledem k $\mathscr{B}^l$. Na pravděpodobnostním prostoru mějme posloupnost σ-algeber $\mathscr{F}_1 \subset \mathscr{F}_2 \subset \ldots \subset \mathscr{A}$ a posloupnost l-rozměrných náhodných funkcí $\{e(t, x), x \in R^l\}_{t=2}^{\infty}$ měřitelných vzhledem k $\mathscr{B}^l \times \mathscr{F}_t$, takových, že $e(t, x)$, $x \in R^l$, je nezávislé na $\mathscr{F}_{t-1}$ a $\mathsf{E}\, e(t, x) = 0$, $\forall_{t,x}$.

Funkci M nazýváme regresní funkcí. Předpokládáme, že v čase t ji můžeme v libovolně zvoleném bodě x nevychýleně odhadnout pozorováním Y.

Robbinsova-Monroova (dále jen RM) *stochastická aproximační metoda* je popsána předpisem: Zvolme libovolně X_1 ($\mathscr{F}_1$-měřitelné); dále definujeme rekurentně

(3.1) $$X_{t+1} = X_t + a_t Y_t, \quad 1 \leqq t < +\infty, \quad \text{kde}$$
$$Y_t = M(X_t) + e(t + 1, X_t)$$

a a_t jsou kladné konstanty.

Je intuitivně zřejmé (a lze formálně dokázat), že takto definovaná posloupnost $\{X_t\}$ je markovská; její vytvořující operátor je

(3.2) $$LV(t, x) = \mathsf{E}\, V(t + 1, x + a_t(M(x) + e(t + 1, x))) - V(t, x).$$

Všude v dalším značí K s indexem či bez indexu kladné konstanty; přitom tentýž symbol může mít v různých větách různou hodnotu. Symbol ∇ značí gradient.

Věta 3.1. Nechť existuje funkce $V: R^l \to R$, dvakrát spojitě diferencovatelná, s omezenými druhými derivacemi, taková, že pro nějaké $\theta \in R^l$ je

$$(3.3) \qquad V(\theta) = 0\,, \quad V(x) > 0 \quad \forall_{x \neq \theta}, \quad \sup_{\varepsilon < |x-\theta| < 1/\varepsilon} (M(x), \nabla V(x)) < 0 \quad \forall_{\varepsilon > 0}\,,$$

$$\lim_{|x| \to \infty} V(x) = +\infty\,,$$

a dále

$$(3.4) \qquad |M(x)|^2 + \mathsf{E}|e(t, x)|^2 \leqq K(1 + V(x) + |(M(x), \nabla V(x))|)\,,$$

$$(3.5) \qquad \sum_{t=1}^{\infty} a_t = +\infty\,, \quad \sum_{t=1}^{\infty} a_t^2 < +\infty\,.$$

Potom platí $\lim_{t \to \infty} X_t = \theta$ s pravděpodobností 1.

Důkaz. S užitím Taylorova vzorce dostáváme

$$LV(x) = \mathsf{E}\, V(x + a_t(M(x) + e(t + 1, x))) - V(x) \leqq$$

$$\leqq a_t(M(x), \nabla V(x)) + K_1 a_t^2(|M(x)|^2 + \mathsf{E}|e(t + 1, x)|^2) \leqq$$

$$\leqq K_2 a_t^2(1 + V(x)) - a_t(1 - K_2 a_t)\,|(M(x), \nabla V(x))|\,.$$

Tvrzení plyne nyní z věty 2.2, kde položíme

$$g_t = K_2 a_t^2,\ \alpha_t = a_t(1 - K_2 a_t),\ \varphi(t, x) = |(M(x), \nabla V(x))|\,.$$ □

Zvolíme-li speciálně $V(x) = (C(x - \theta), (x - \theta))$, kde C je kladně definitní matice, můžeme větu formulovat takto.

Věta 3.2. Nechť existuje kladně definitní matice C taková, že

$$(3.6) \qquad \sup_{\varepsilon < |x-\theta| < 1/\varepsilon} (M(x), C(x - \theta)) < 0 \quad \forall_{\varepsilon > 0}\,,$$

dále nechť

$$(3.7) \qquad |M(x)|^2 + \mathsf{E}|e(t, x)|^2 \leqq K_1(1 + |x|^2)\,,$$

a nechť je splněno (3.5).
Potom $\lim_{t \to \infty} X_t = \theta$ s pravděpodobností 1.

Důkaz je týž jako u předchozí věty; pro LV nyní dostáváme

$$(3.8) \qquad LV(x) \leqq K_2 a_t^2(1 + (V(x)) - 2a_t|(M(x), C(x - \theta))|\,.$$

V (3.7) je ovšem možno psát $1 + |x - \theta|^2$ místo $1 + |x|^2$. □

Nahradíme-li požadavek (3.6) přísnějším, potom dostaneme i konvergenci X_t k θ podle kvadratického středu. Přitom (jde-li *jen* o konvergenci podle kvadratického středu) lze vynechat předpoklad $\sum_{t=1}^{\infty} a_t^2 < +\infty$.

Věta 3.3. Nechť existuje kladně definitní matice C taková, že

$$(M(x), C(x - \theta)) \leqq -K(C(x - \theta), x - \theta) \quad \forall_{x \in R^l} ; \tag{3.9}$$

dále nechť je splněno (3.7) a

$$a_t \to 0 , \quad \sum_{t=1}^{\infty} a_t = +\infty . \tag{3.10}$$

Potom $\lim_{t \to \infty} \mathsf{E}|X_t - \theta|^2 = 0$, pokud $\mathsf{E}|X_1|^2 < +\infty$.

Důkaz. Nerovnost (3.8) nabude nyní tvaru

$$L V(x) \leqq K_2 a_t^2 - K_3 a_t V(x) , \tag{3.11}$$

kde $K_3 = 2K - \varepsilon$ s libovolným $\varepsilon > 0$ a (3.11) platí pro $t \geqq t_0(\varepsilon)$. Ze vztahu (2.5) plyne

$$\mathsf{E}\, L V(X_t) = \mathsf{E}\, V(X_{t+1}) - \mathsf{E}\, V(X_t) ; \tag{3.12}$$

spojením (3.11) a (3.12) pak dostáváme

$$\mathsf{E}\, V(X_{t+1}) \leqq (1 - K_3 a_t)\, \mathsf{E}\, V(X_t) + K_2 a_t^2 . \tag{3.13}$$

Odtud

$$\mathsf{E}\, V(X_t) \leqq \prod_{u=t_0}^{t-1} (1 - K_3 a_u)\, \mathsf{E}\, V(X_{t_0}) + K_2 \sum_{k=t_0}^{t-1} a_k^2 \prod_{u=k+1}^{t-1} (1 - K_3 a_u) . \tag{3.14}$$

Chování výrazů typu pravé strany (3.14) je známo z diferenčních rovnic; v našem případě jde pravá strana (3.14) k 0. Je tedy $\mathsf{E}\, V(X_t) \to 0$ a odtud i $\mathsf{E}|X_t - \theta|^2 \to 0$. □

4. RYCHLOST KONVERGENCE RM METODY

Konstanty a_t v RM metodě se zpravidla volí ve tvaru

$$a_t = a/t^{\alpha} , \quad a > 0 , \quad 0 < \alpha \leqq 1 \tag{4.1}$$

(resp. $\frac{1}{2} < \alpha \leqq 1$, žádá-li se konvergence řady $\sum_{t=1}^{\infty} a_t^2$). Vyšetříme pro tuto volbu a_t rychlost konvergence X_t k θ s pravděpodobností 1 i podle kvadratického středu.

Věta 4.1. Za předpokladů věty 3.3 a při volbě konstant (4.1) platí pro $t \to \infty$

(i) $\mathsf{E}|X_t - \theta|^2 = O(t^{-\alpha})$, je-li buďto $0 < \alpha < 1$ nebo $\alpha = 1$ a $2Ka > 1$,

(ii) $X_t - \theta = O(t^{-\alpha+\frac{1}{2}+\delta})$ s pravděpodobností 1 pro libovolné $\delta > 0$, je-li buďto $\frac{1}{2} < \alpha < 1$ nebo $\alpha = 1$ a $2Ka > 1$.

Důkaz. (i) Klademe opět $V(x) = (C(x - \theta), x - \theta)$. Dle (3.13) je

$$\mathsf{E}\, V(X_{t+1}) \leqq \left(1 - \frac{(2K - \varepsilon)\, a}{t^\alpha}\right) \mathsf{E}\, V(X_t) + \frac{K_4}{t^2} \quad \text{pro} \quad t \geqq t_0(\varepsilon)\,. \tag{4.2}$$

Opět se lze odvolat na známé vlastnosti řešení diferenčních nerovnic. Tzv. *Chungova lemma* [13] říká: Nechť $\{b_t\}$ je posloupnost nezáporných čísel taková, že

$$b_{t+1} \leqq \left(1 - \frac{c}{t^\alpha}\right) b_t + \frac{c_1}{t^{\alpha+\beta}}$$

pro všechna $t \geqq t_0$, kde $c > 0$, $c_1 > 0$, $\beta > 0$ a buďto $0 < \alpha < 1$ nebo $\alpha = 1$ a $c > \beta$. Potom $b_t = O(t^{-\beta})$.

Užitím této lemmy na (4.2) dostáváme $\mathsf{E}\, V(X_t) = O(t^{-\alpha})$ a tedy i $\mathsf{E}|X_t - \theta|^2 = O(t^{-\alpha})$.

(ii) Nyní položme $V_1(t, x) = t^\gamma V(x) + t^{-\eta}$, kde $V(x) = (C(x - \theta), x - \theta)$, $0 < \gamma < 2\alpha - 1$, $0 < \eta < 2\alpha - 1 - \gamma$. S užitím (3.11) dostáváme

$$\begin{aligned} L V_1(t, x) &= (t + 1)^\gamma L V(x) + ((t + 1)^\gamma - t^\gamma) V(x) + (t + 1)^{-\eta} - t^{-\eta} \leqq \\ &\leqq -(2K - \varepsilon)\, a t^{-\alpha+\gamma} V(x) + K_5 t^{-2\alpha+\gamma} + \gamma t^{-1+\gamma} V(x) - \tfrac{1}{2}\eta t^{-1-\eta} \leqq \\ &\leqq -K_6 t^{-\alpha+\gamma} V(x)\,, \quad \text{je-li} \quad 0 < \alpha < 1 \quad \text{nebo} \quad \alpha = 1 \quad \text{a} \quad 2Ka > 1\,, \end{aligned}$$

neboť součet členů neobsahujících $V(x)$ je záporný (je totiž $-2\alpha + \gamma < -1 - \eta$). Je tedy $V_1(t, X_t)$ supermartingal a existuje konečná $\lim_{t\to\infty} V_1(t, X_t) = \lim_{t\to\infty} t^\gamma V(X_t)$; odtud $t^\gamma |X_t - \theta|^2 = O(1)$, neboli $|X_t - \theta| = O(t^{-\gamma/2})$. Vyjádříme-li γ jako $2\alpha - 1 - 2\delta$, $\delta > 0$, dostáváme tvrzení věty. □

Případ "$\alpha = 1$ a současně $2Ka \leqq 1$" jsme z věty vyjmuli, protože je z praktického hlediska nepříznivý. Nicméně podrobnějším vyšetřením (4.2) bychom zjistili, že za předpokladů věty je

$$\mathsf{E}|X_t - \theta|^2 = \begin{cases} O(t^{-1} \log t) & \text{pro} \quad \alpha = 1\,, \quad 2Ka = 1\,, \\ O(t^{-2Ka}) & \text{pro} \quad \alpha = 1\,, \quad 2Ka < 1\,; \end{cases}$$

pro konvergenci s pravděpodobností 1 pak platí

$$X_t - \theta = O(t^{-Ka+\delta}) \quad \text{pro} \quad \alpha = 1\,, \quad 2Ka \leqq 1\,,$$

avšak tentokrát jen za dodatečných předpokladů (o omezenosti vyšších momentů $e(t, x)$, [22], nebo o existenci druhé derivace funkce M, [20]).

O rychlosti konvergence $X_t \to \theta$ vypovídá i následující věta, obsahující ve srovnání s předchozími větami předpoklad o stejnoměrné omezenosti náhodných složek $e(t, x)$.

Věta 4.2. Předpokládejme, že

$$(M(x), x - \theta) \leqq -K|x - \theta|^2\,, \quad |M(x)| \leqq K_1(1 + |x|)\,, \quad \forall_x\,;$$

$$|e(t, x)| \leqq K_2,\ \forall_{t,x}\,; \quad a_t = a/t\,, \quad a > 0\,.$$

Potom pro každé $\varepsilon > 0$ existuje $K_3 = K_3(\varepsilon)$ tak, že $\mathsf{P}(|X_t - \theta| \geqq \varepsilon) \leqq \exp(-K_3 t)$, $\forall_t$.

Důkaz (obecnějšího tvrzení) je podán v práci [25].

5. KONVERGENČNÍ VĚTY PRO MODIFIKACE RM METODY

Konvergenční věty, o nichž byla řeč v předchozích dvou odstavcích, obsahovaly předpoklady typu $|M(x)| \leqq K(1 + |x|)$, $\mathsf{E}|e(t, x)|^2 \leqq K_1(1 + |x|^2)$, omezující růst regresní funkce resp. rozptylů náhodných složek pro $|x| \to \infty$. Splnění těchto předpokladů jen v nějaké omezené oblasti je mnohem slabší požadavek (totiž požadavek omezenosti funkcí M a $\mathsf{E}|e(t, .)|^2$ v té oblasti). Přitom v reálných situacích jsme zpravidla schopni udat omezenou oblast U, v které už neznámé θ s jistotou leží, a modifikovat aproximační metodu tak, aby tuto oblast neopustila. Formalizace této myšlenky je obsažena v následující větě, dokázané v [22].

Nechť $\Phi : R^l \to R^l$ je $\mathscr{B}^l$-měřitelná funkce taková, že

$$\Phi(x) = x \quad \text{pro} \quad x \in U \quad \text{a} \quad \Phi(x) \in U \quad \text{pro} \quad x \in R^l\,. \tag{5.1}$$

Zmíněná modifikace RM posloupnosti, označme ji $\{\hat{X}_t\}$, se pak definuje předpisem

$$\hat{X}_{t+1} = \Phi(\hat{X}_t + a_t(M(\hat{X}_t) + e(t + 1, \hat{X}_t)))\,, \tag{5.2}$$

kde $M(x)$, $e(t, x)$, a_t mají stejný význam jako dříve. Procedura může být i znáhodněná, tj. v roli Φ lze připustit i náhodnou funkci $\Phi(x, \omega)$, $x \in R^l$, nezávislou na $\{e(t, x), x \in R^l\}_{t=2}^{\infty}$. Označme $\hat{L}$ vytvořující operátor markovské posloupnosti $\{\hat{X}_t\}$.

Věta 5.1. Nechť kořen θ rovnice $M(x) = 0$ leží v omezené oblasti U; nechť existuje funkce $V(x) : R^l \to R$, dvakrát spojitě diferencovatelná, s omezenými druhými derivacemi taková, že

$$V(\theta) = 0\,, \quad V(x) > 0\,, \quad \forall_{x \neq \theta}, \quad \sup_{\substack{|x-\theta| > \varepsilon \\ x \in \bar{U}}} (M(x), \nabla V(x)) < 0\,, \quad \forall_{\varepsilon > 0},$$

kde $\bar{U}$ značí uzávěr U. Nechť funkce $M(x)$ a $\mathsf{E}|e(t, x)|^2$ jsou stejnoměrně omezené pro $2 \leqq t < +\infty$, $x \in \bar{U}$; nechť $a_t > 0$, $\sum_{t=1}^{\infty} a_t = +\infty$, $\sum_{t=1}^{\infty} a_t^2 < +\infty$. Konečně nechť

$$\hat{L}V(x) \leqq LV(x)\,, \quad \forall_{x\in U}\,. \tag{5.3}$$

Potom pro posloupnost $\{\hat{X}_t\}$ definovanou předpisem (5.2) platí

$$\lim_{t\to\infty} \hat{X}_t = \theta \quad \text{s pravděpodobností} \quad 1\,.$$

Poznámka. Podmínka (5.3) je splněna zejména tehdy, platí-li $V(\Phi(x)) \leqq V(x)$ pro všechna $x \in R^l$, (případně s pravděpodobností 1, je-li $\Phi(x)$ náhodná funkce). V jednorozměrném případě se věta velmi zjednoduší.

Věta 5.2. Nechť kořen θ rovnice $M(x) = 0$ leží v omezeném intervalu $U = (u_1, u_2)$; nechť $\sup_{\substack{|x-\theta|>\varepsilon \\ x\in\langle u_1,u_2\rangle}} M(x)(x-\theta) < 0$, $\forall_{\varepsilon>0}$. Nechť funkce $M(x)$ a $\mathsf{E}\, e^2(t, x)$ jsou stejnoměrně omezené pro $2 \leqq t < +\infty$, $x \in \langle u_1, u_2\rangle$; nechť $a_t > 0$, $\sum_{t=1}^{\infty} a_t = +\infty$, $\sum_{t=1}^{\infty} a_t^2 < +\infty$. Potom

$$\lim_{t\to\infty} \hat{X}_t = \theta \text{ s pravděpodobností } 1.$$

Důkaz. Věta je důsledkem předchozí věty, kde volíme funkci V (s omezenou V'') klesající pro $x < \theta$, rovnou 0 pro $x = \theta$, rostoucí pro $x > \theta$ a splňující $V(u_1) = V(u_2)$. Potom je $V(x') < V(x)$, $\forall_{x'\in U, x\notin U}$, a tedy $V(\Phi(x)) \leqq V(x)$, $\forall_{x\in R}$. □

Předpokladu $|M(x)| \leqq K(1 + |x|)$ omezujícího růst regresní funkce se lze zbavit i tehdy, není-li znám interval (u_1, u_2) obsahující θ. Stačí omezit shora veličiny $(a/t)\, Y_t$, o které opravujeme aproximace X_t v jednotlivých krocích. Pro případ omezených rozptylů pozorování Y_t platí následující věta, dokázaná v práci [14]. Je v ní použito symboliky

$$[y]_c^d = \max(c, \min(y, d))\,, \quad \text{pro} \quad c < d\,.$$

Věta 5.3. Nechť M je omezená v každém omezeném intervalu, nechť

$$\sup_{|x-\theta|>\varepsilon} M(x)(x-\theta) < 0\,, \quad \forall_{\varepsilon>0}\,; \quad \mathsf{E}\, e^2(t, x) \leqq K\,, \quad \forall_{t,x}\,.$$

Posloupnost $\{X_t\}$ nechť je definována předpisem

$$X_{t+1} = X_t + [a_t Y_t]_{-1}^{1}\,,$$

kde $Y_t = M(X_t) + e(t+1, X_t)$, $a_t = a/t$, $a > 0$. Potom $\lim_{t\to\infty} X_t = \theta$ s pravděpodobností 1. Je-li navíc $|M(x)| \geqq K_1|x-\theta|$ v nějakém okolí bodu θ a je-li $\mathsf{E}\, X_1^2 <$

$< +\infty$, pak také $X_t \to \theta$ pode kvadratického středu; pro $2aK_1 > 1$ platí $\mathsf{E}(X_t - \theta)^2 = O(t^{-1})$.

Následující modifikace RM metody rovněž nevyžaduje předpoklad $|M(x)| \leqq \leqq K(1 + |x|)$. Zdá se být výhodná tam, kde experimentátor má důvod se domnívat, že neznámé θ leží v intervalu (u_1, u_2), ale chce být zabezpečen i proti možnosti, že $\theta \notin (u_1, u_2)$.

Věta 5.4. Nechť M je omezená v každém omezeném intervalu, nechť

$$\sup_{\varepsilon < |x-\theta| < 1/\varepsilon} M(x)(x - \theta) < 0\,, \ \forall_{\varepsilon > 0}\,; \quad \mathsf{E}\, e^2(t, x) \leqq K_1(1 + x^2),\ \forall_{t,x}\,;$$

$$a_t > 0\,, \quad \sum_{t=1}^{\infty} a_t = +\infty\,, \quad \sum_{t=1}^{\infty} a_t^2 < +\infty\,.$$

Posloupnost $\{X_t\}$ nechť je definována předpisem

$$X_{t+1} = \begin{cases} u_1\,, & \text{je-li } X_t > u_2 \text{ a současně } X_t + a_t Y_t < u_1\,,\\ u_2\,, & \text{je-li } X_t < u_1 \text{ a současně } X_t + a_t Y_t > u_2\,,\\ X_t + a_t Y_t & \text{ve všech ostatních případech,} \end{cases}$$

kde $Y_t = M(X_t) + e(t + 1, X_t)$. Potom $\lim_{t\to\infty} X_t = \theta$ s pravděpodobností 1.

Poznámka. Věta 5.4 je uvedena v [27] jako jeden z příkladů na tam dokázanou *konvergenční větu pro nezáporné "skoro-supermartingaly"*:

(i) Jsou-li Z_t, β_t, U_t, V_t, $1 \leqq t < +\infty$, nezáporné náhodné veličiny, měřitelné vzhledem k $\mathscr{F}_t$ (kde $\mathscr{F}_1 \subset \mathscr{F}_2 \subset \ldots$ je neklesající posloupnost σ-algeber) a takové, že

$$\mathsf{E}(Z_{t+1} \mid \mathscr{F}_t) \leqq (1 + \beta_t)\, Z_t + U_t - V_t\,,$$

potom $\lim_{t\to\infty} Z_t$ existuje a je konečná a $\sum_{t=1}^{\infty} V_t$ konverguje skoro jistě na množině

$$\Big[\sum_{t=1}^{\infty} \beta_t < +\infty\,, \quad \sum_{t=1}^{\infty} U_t < +\infty\Big]\,.$$

(ii) Zavedeme-li pro $b > 0$ $\beta_t(b) = \beta_t I_{[Z_t \leqq b]}$, $U_t(b) = U_t I_{[Z_t \leqq b]}$, $V_t(b) = V_t I_{[Z_t \leqq b]}$, potom za týchž předpokladů platí, že $\lim_{t\to\infty} Z_t$ existuje skoro jistě na

$$B = \bigcap_b \Big[\sum_{t=1}^{\infty} \beta_t(b) < +\infty\,, \quad \sum_{t=1}^{\infty} U_t(b) < +\infty\Big]$$

a $\sum_{t=1}^{\infty} V_t < +\infty$ skoro jistě na $B \cap [\lim_{t\to\infty} Z_t < +\infty]$.

Důkaz věty 5.4 probíhá nyní takto: Snadno se ověří nerovnost

$$\mathsf{E}((X_{t+1} - \theta)^2 \mid \mathscr{F}_t) \leqq (X_t - \theta)^2 + 2a_t M(X_t)(X_t - \theta) +$$
$$+ a_t^2(M^2(X_t) + \mathsf{E}\, e^2(t+1, X_t))\left(1 + \frac{u_1^2 + u_2^2}{(u_2 - u_1)^2}\right).$$

Ze skoro-supermartingalové věty (ii) pro $Z_t = (X_t - \theta)^2$ pak vyplývá, že $\lim_{t\to\infty} X_t = \theta$ skoro jistě na množině $[\liminf_{t\to\infty} |X_t| < +\infty]$. Zbývá dokázat, že $\mathsf{P}(\liminf_{t\to\infty} |X_t| = +\infty) = 0$. K tomu cíli se položí $Z_t = [(X_t - \max(\theta, u_2))^+]^2$ a vypočte, že

$$\mathsf{E}(Z_{t+1} \mid \mathscr{F}_t) \leqq (1 + 2Ka_t^2)\, Z_t + K_2 a_t^2 .$$

Z věty (i) pak plyne $\mathsf{P}(\lim_{t\to\infty} Z_t = +\infty) = \mathsf{P}(\lim_{t\to\infty} X_t = +\infty) = 0$. Symetrickou úvahou dostaneme, že $\mathsf{P}(\lim_{t\to\infty} X_t = -\infty) = 0$. Poslední možnost je, že $\limsup_{t\to\infty} X_t = +\infty$ a $\liminf_{t\to\infty} X_t = -\infty$, v takovém případě je však $X_t \in \langle u_1, u_2 \rangle$ nekonečně mnohokrát a tedy $\liminf_{t\to\infty} |X_t| < +\infty$. □

6. ADAPTIVNÍ RM METODA

Modifikací jiného druhu je tzv. adaptivní RM metoda, studovaná poprvé Venterem v práci [29]. Je motivována výsledkem, který uvedeme teprve později a který říká, že RM posloupnost $\{X_t\}$ s konstantami $a_t = a/t$ je za jistých předpokladů asymptoticky normální, přičemž asymptotický rozptyl se minimalizuje volbou $a = 1/|M'(\theta)|$. Derivaci $M'(\theta)$ sice zpravidla neznáme, ale můžeme ji odhadovat v průběhu aproximačního postupu:

V t-tém kroku provedeme dvě pozorování Y_t' a Y_t'' funkce M v bodech $X_t + c_t$ a $X_t - c_t$, utvoříme podíl $(Y_t' - Y_t'')/2c_t$ a odhadneme $M'(\theta)$ průměrem těchto podílů,

$$W_t = \frac{1}{t-1} \sum_{i=1}^{t-1} \frac{Y_i' - Y_i''}{2c_i} ;$$

průměr W_t lze popsat také rekurentně,

$$(6.1) \qquad W_1 = 0\,, \quad W_{t+1} = W_t + \frac{1}{t}\left(\frac{Y_t' - Y_t''}{2c_t} - W_t\right), \quad 1 \leqq t < +\infty\,.$$

Předpokládáme, že jsou známa čísla $0 < r_1 < r_2 < +\infty$ taková, že

$$r_1 \leqq M'(\theta) \leqq r_2 \,. \tag{6.2}$$

Této apriorní informaci přizpůsobíme odhad W_t derivace $M'(\theta)$, totiž nahradíme jej odhadem

$$[W_t] = [W_t]_{r_1}^{r_2} \,(= \max(r_1, \min(W_t, r_2))) \tag{6.3}$$

a dosadíme do aproximačního schématu:

$$X_{t+1} = X_t - \frac{1}{t[W_t]} Y_t \,. \tag{6.4}$$

Protože pozorování funkce M v bodě X_t neprovádíme, nahrazujeme je průměrem pozorování v bodech $X_t \pm c_t$, tj. Y_t v (6.4) značí $Y_t = \frac{1}{2}(Y_t' + Y_t'')$.

Poznamenejme ještě, že u Venterovy metody je zvykem uvažovat funkci M rostoucí v bodě θ (nikoli klesající, jak tomu bylo dosud); proto je v (6.4) znaménko minus, nikoli plus.

Věta 6.1. Nechť pro všechna $x \in R$ a $2 \leqq t < +\infty$ je $M(x) \lesseqqgtr 0$ pro $x \lesseqqgtr \theta$,

$$M^2(x) + \mathsf{E}\, e^2(t, x) \leqq K_1(1 + x^2) \tag{6.5}$$

$$\text{existuje} \quad M''(x) \quad \text{spojitá a omezená,} \tag{6.6}$$

$$c_t = \frac{c}{t^\gamma} \quad \text{pro nějaké} \quad 0 < \gamma < \tfrac{1}{2} \,. \tag{6.7}$$

Potom pro posloupnosti $\{X_t\}$ a $\{W_t\}$ definované vztahy (6.1), (6.3), (6.4) platí

$$\lim_{t\to\infty} X_t = \theta \,, \quad \lim_{t\to\infty} W_t = M'(\theta) \quad \text{s pravděpodobností} \quad 1 \,.$$

Důkaz. Nechť L je vytvořující operátor posloupnosti $\{X_t, W_t\}$; aplikujme jej na funkci $V(x, w) = (x - \theta)^2$:

$$L V(x, w) = \mathsf{E}\left(x - \theta - \frac{1}{t[w]} \cdot \frac{M(x + c_t) + M(x - c_t)}{2} + \right.$$
$$\left. + \frac{e_1(t + 1, x)}{t[w]}\right)^2 - (x - \theta)^2 \,,$$

kde

$$e_1(t + 1, x) = \tfrac{1}{2}(e(t + 1, x + c_t) + e(t + 1, x - c_t)) \,. \tag{6.8}$$

Z předpokladu (6.6) plyne

$$\left|\frac{M(x + c_t) + M(x - c_t)}{2} - M(x)\right| < \frac{K_2}{t^{2\gamma}};$$

odtud a z ostatních předpokladů dostáváme

$$LV(x, w) \leqq - \frac{2}{t[w]} M(x)(x - \theta) + \frac{2K_2}{t^{1+2\gamma}} |x - \theta| +$$

$$+ \frac{2}{t^2[w]^2}\left(M^2(x) + \frac{K_2^2}{t^{4\gamma}}\right) + \frac{1}{t^2[w]^2} \mathsf{E}\, e_1^2(t + 1, x) \leqq - \frac{2}{tr_2} M(x)(x - \theta) +$$

$$+ K_3\left(\frac{1}{t^{1+2\gamma}} + \frac{1}{t^2}\right)(1 + V(x, w)).$$

Podle (malé obměny) věty 2.2 máme $\lim_{t\to\infty} X_t = \theta$ s pravděpodobností 1. Dále,

$$W_{t+1} = \frac{1}{t}\sum_{i=1}^{t} \frac{M(X_i + c_i) - M(X_i - c_i)}{2c_i} + \frac{1}{t}\sum_{i=1}^{t} \varepsilon_i' - \frac{1}{t}\sum_{i=1}^{t} \varepsilon_i'', \tag{6.9}$$

kde

$$\varepsilon_i' = \frac{e(i + 1, X_i + c_i)}{2c_i}, \quad \varepsilon_i'' = \frac{e(i + 1, X_i - c_i)}{2c_i}.$$

Z dokázaného již tvrzení $X_t \to \theta$ s pravděpodobností 1 snadno plyne $(M(X_t + c_t) - M(X_t - c_t))/2c_t \to M'(\theta)$ a tedy také první člen na pravé straně (6.9) konverguje k $M'(\theta)$ (vše s pravděpodobností 1). Konvergence druhého a třetího členu k nule vyplývá ze zákona velkých čísel, neboť $\mathsf{E}(\varepsilon_{t+1}' \mid \varepsilon_1', \varepsilon_1'', \ldots, \varepsilon_t', \varepsilon_t'') = 0$, stejně tak pro ε_{t+1}'' a řady $\sum_{t=1}^{\infty} \frac{\mathsf{E}\varepsilon_t'^2}{t^2}$ i $\sum_{t=1}^{\infty} \frac{\mathsf{E}\varepsilon_t''^2}{t^2}$ konvergují; podle poznámky na konci odst. 2 je totiž posloupnost $\mathsf{E}(X_t - \theta)^2$ omezená, tedy vzhledem k (6.5) i posloupnost $\mathsf{E}\, e^2(t + 1, X_t \pm c_t)$ je omezená, tj. obě posloupnosti $\mathsf{E}\varepsilon_t'^2$ i $\mathsf{E}\varepsilon_t''^2$ jsou řádu $O(t^{2\gamma})$. □

7. ASYMPTOTICKÁ NORMALITA

Původně byla asymptotická normalita RM posloupnosti dokazována metodou momentů (Chung [13]), tj. ověřením, že momenty všech řádů vhodně normovaných veličin X_t konvergují k odpovídajícím momentům normovaného normálního rozdělení. Tento přístup však předpokládá existenci a omezenost momentů všech řádů veličin $e(t, x)$, což není pro samotnou asymptotickou normalitu nutnou podmínkou.

Uvedeme proto novější výsledek Fabianův [8], který je dostatečný obecný, takže z něj plyne asymptotická normalita RM metody, adaptivní RM metody i některých dalších aproximačních metod, o nichž bude řeč později.

Věta 7.1. Nechť $\mathscr{F}_1 \subset \mathscr{F}_2 \subset \ldots (\subset \mathscr{A})$ je posloupnost σ-algeber, nechť U_t, V_t, T_t jsou l-rozměrné náhodné vektory, Γ_t, Φ_t $l \times l$-rozměrné náhodné matice; nechť Γ_t, Φ_t, V_{t-1} jsou $\mathscr{F}_t$-měřitelné, $\Sigma, \Gamma, \Phi, P \in R^{l \times l}$, Γ kladně definitní, P ortogonální, $P^{\mathrm{T}}\Gamma P = \Lambda$ diagonální. Dále nechť

$$\Gamma_t \to \Gamma\,, \quad \Phi_t \to \Phi\,, \quad T_t \to T \text{ s pravděpodobnosti } 1\,,$$

$$\mathsf{E}(V_t \mid \mathscr{F}_t) = 0\,, \quad \|\mathsf{E}(V_t V_t^{\mathrm{T}} \mid \mathscr{F}_t)\| < K\,,$$

$$\lim_{t\to\infty} \mathsf{E}(V_t V_t^{\mathrm{T}} \mid \mathscr{F}_t) = \Sigma \text{ s pravděpodobnosti } 1\,,$$

$$\lim_{t\to\infty} \mathsf{E}(|V_t|^2 I_{[|V_t|^2 \geqq rt]}) = 0\,, \quad \forall_{r>0}\,.$$

Je-li

$$U_{t+1} = (I - t^{-1}\Gamma_t)\, U_t + t^{-(1+\beta)/2}\Phi_t V_t + t^{-1-\beta/2} T_t \tag{7.1}$$

pro všechna $1 \leqq t < +\infty$ a nějaké $0 \leqq \beta < 2\lambda$, kde $\lambda = \min\limits_{1\leqq i\leqq l} \Lambda^{(ii)}$, potom asymptotické rozdělení veličiny $t^{\beta/2}U_t$ je normální se střední hodnotou $(\Gamma - (\beta/2)I)^{-1}\, T$ a kovarianční maticí PQP^{T}, kde

$$Q^{(ij)} = \frac{(P^{\mathrm{T}}\Phi\Sigma\Phi^{\mathrm{T}}P)^{(ij)}}{\Lambda^{(ii)} + \Lambda^{(jj)} - \beta}\,.$$

Důkaz viz [8].

Věta 7.2. Nechť RM posloupnost $\{X_t\}$ definóvaná předpisem (3.1) s $a_t = a/t$ konverguje k θ s pravděpodobností 1. Dále nechť

$$M(x) = B(x - \theta) + \delta(x)\,, \quad \text{kde} \quad \delta(x) = o(|x - \theta|) \quad \text{pro} \quad x \to \theta \tag{7.2}$$

a matice B je záporně definitní,

$$\|\mathsf{E}(e(t, x)\, e^{\mathrm{T}}(t, x))\| < K \quad \text{pro všechna} \quad 2 \leqq t < +\infty \tag{7.3}$$

a všechna x z nějakého okolí θ,

$$\lim_{t\to\infty, x\to\theta} \mathsf{E}(e(t, x)\, e^{\mathrm{T}}(t, x)) = \Sigma\,, \tag{7.4}$$

$$\lim_{t\to\infty, x\to\theta} \mathsf{E}(|e(t, x)|^2 I_{[|e(t,x)|^2 \geqq rt]}) = 0\,, \quad \forall_{r>0}\,. \tag{7.5}$$

Nechť P je ortogonální matice taková. že $-P^{\mathrm{T}}BP = \Lambda_*$ je diagonální; nechť $2a\lambda_* > 1$, kde $\lambda_* = \min\limits_i \Lambda_*^{(i,i)}$.

Potom asymptotické rozdělení veličiny $t^{1/2}(X_t - \theta)$ je normální se střední hodnotou 0 a kovarianční maticí $aPQP^{\mathrm{T}}$, kde

$$Q^{(ij)} = \frac{a(P^{\mathrm{T}}\Sigma P)^{(ij)}}{a\Lambda_*^{(ii)} + a\Lambda_*^{(jj)} - 1}\,. \tag{7.6}$$

Důkaz. Zbytek $\delta(x)$ v (7.2) lze psát ve tvaru $\delta(x) = D(x)(x - \theta)$, kde matice $D(x) \to 0$ pro $x \to \theta$. Důkaz pak plyne z věty 7.1, položíme-li v ní $U_t = X_t - \theta$, $V_t = e(t + 1, X_t)$, $T_t = 0$, $-\Gamma_t = a(B + D(X_t))$, $\Phi_t = I$, $\beta = 1$. □

Připojme k větě 7.2 následující dvě poznámky:

(i) Předpoklad, že B je záporně definitní a současně $2a\lambda_* > 1$ (neboli že B je záporně definitní a všechna charakteristická čísla B jsou $< -1/(2a)$) lze nahradit slabším předpokladem – viz [22] –, že všechna charakteristická čísla matice B mají reálné části $< -1/(2a)$. Kovarianční matice asymptotického rozdělení má pak tvar

$$a \int_0^\infty e^{Av} \Sigma e^{A^T v} \, dv, \tag{7.7}$$

kde $A = B + (1/2a) I$. (V případě záporně definitní B lze integrál (7.7) vypočítat a uvést na tvar (7.6).)

(ii) Jestliže náhodné veličiny $e(t, x) = e(t)$ nezávisí na x a jsou stejně rozdělené (pro všechna t), pak je (7.3)–(7.5) splněno, jakmile $\mathsf{E}|e(2)|^2 < +\infty$ (a je $\Sigma = \mathsf{E}(e(2)\, e^T(2))$).

Vyslovme ještě větu 7.2 pro případ jednorozměrný (její znění je pak jednodušší).

Věta 7.3. Nechť jednorozměrná RM posloupnost $\{X_t\}$ definovaná předpisem (3.1) s $a_t = a/t$ konverguje k θ s pravděpodobností 1. Nechť funkce M má v bodě θ zápornou vlastní derivaci, nechť $2a|M'(\theta)| > 1$. Dále nechť $\mathsf{E}e^2(t, x) < K$ pro všechna t a všechna x z nějakého okolí θ, $\lim_{t\to\infty, x\to\theta} \mathsf{E}e^2(t, x) = \sigma^2 < +\infty$,

$$\lim_{t\to\infty, x\to\theta} \mathsf{E}(e^2(t, x)\, I_{[e^2(t,x) \geq rt]}) = 0 \,, \quad \forall_{r>0} \,.$$

Potom $t^{1/2}(X_t - \theta)$ je asymptoticky normální s parametry

$$\left(0, \frac{a^2\sigma^2}{2a|M'(\theta)| - 1}\right). \tag{7.8}$$

Věta 7.2, resp. 7.3 o asymptotické normalitě zůstane v platnosti, nahradíme-li předpis (3.1) kterýmkoli z modifikovaných RM předpisů, které jsme uvažovali v odst. 5. Jestliže totiž některá modifikovaná metoda $\{\hat{X}_t\}$ konverguje k θ s pravděpodobností 1, pak příslušnou modifikaci je nutno provést jen konečně mnohokrát, a tedy od nějakého $T = T(\varepsilon)$ je už s pravděpodobností $> 1 - \varepsilon$ posloupnost $\{\hat{X}_t\}$ shodná s nemodifikovanou RM posloupností $\{X_t\}$ s počáteční hodnotou $\hat{X}_T$.

Ze vzorce (7.8) vidíme, že asymptotický rozptyl je minimalizován volbou $a = 1/|M'(\theta)|$ a jeho minimální hodnota je $\sigma^2/(M'(\theta))^2$. Ukážeme si nyní, že téhož minimálního asymptotického rozptylu lze dosáhnout adaptivní RM metodou, zavedenou v odst. 6.

Věta 7.4. Nechť jsou splněny předpoklady věty 6.1, přičemž v (6.7) nechť $\frac{1}{4} < \gamma < \frac{1}{2}$. Dále nechť

$$\lim_{t\to\infty, x\to\theta} \mathsf{E}e^2(t, x) = \sigma^2\,, \quad \lim_{t\to\infty, x\to\theta} \mathsf{E}\big(e^2(t, x)\, I_{[e^2(t,x)\geqq rt]}\big) = 0\,,$$

$$\mathsf{E}\big(e(t, x_1)\, e(t, x_2)\big) = 0\,, \quad \forall_{t, x_1 \neq x_2}\,.$$

Posloupnosti $\{X_t\}$ a $\{W_t\}$ nechť jsou definovány vztahy (6.1), (6.3), (6.4). Potom asymptotické rozdělení veličiny $t^{1/2}(X_t - \theta)$ je normální s parametry $(0, \sigma^2/2(M'(\theta))^2)$.

Poznámka. Zdánlivě je asymptotický rozptyl dvakrát menší než jeho minimální možná hodnota. To je však vysvětleno tím, že adaptivní metoda potřebuje dvojnásobný počet pozorování než obyčejná RM metoda. Je tedy třeba srovnávat přesnost t-té aproximace adaptivní metody s přesností $(2t)$-té aproximace obyčejné RM metody; oba srovnávané asymptotické rozptyly se pak skutečně shodují.

Důkaz věty 7.4. Dle (6.4) a (6.9) je

$$X_{t+1} - \theta = X_t - \theta - \frac{1}{t[W_t]}\left(\frac{M(X_t + c_t) + M(X_t - c_t)}{2} + e_1(t + 1, X_t)\right),$$

dále $\frac{1}{2}(M(X_t + c_t) + M(X_t - c_t)) = m_t(X_t - \theta) + O(c_t^2)$, kde $m_t = M'(\theta + \vartheta(X_t - \theta)) \to M'(\theta)$; celkem dostáváme

$$X_{t+1} - \theta = \left(1 - \frac{m_t}{t[W_t]}\right)(X_t - \theta) - \frac{e_1(t + 1, X_t)}{t[W_t]} + O\left(\frac{c_t^2}{t}\right). \tag{7.9}$$

Ve větě 7.1 položme $U_t = X_t - \theta$, $V_t = e_1(t + 1, X_t)$, $\Gamma_t = m_t/[W_t]$, $\Phi_t = -1/[W_t]$; $T_t = O(t^{1/2}c_t^2)$, $\beta = 1$. Vztah (7.9) se pak přepíše jako

$$U_{t+1} = (1 - t^{-1}\Gamma_t)\, U_t + t^{-1}\Phi_t V_t + t^{-3/2} T_t\,,$$

přičemž $\Gamma_t \to \Gamma = 1$, $\Phi_t \to \Phi = -1/M'(\theta)$, $T_t \to T = 0$. Položíme-li ještě $P = 1$, $\Sigma = \frac{1}{2}\sigma^2$, $\Lambda = 1$, dostáváme dle věty 7.1, že $t^{1/2}(X_t - \theta)$ je asymptoticky normální s parametry $(0, \sigma^2/2(M'(\theta))^2)$. □

V práci [8] je ukázáno, že konvergenční věta i věta o asymptotické normalitě pro adaptivní metodu zůstávají v platnosti i tehdy, neznáme-li dolní a horní odhad r_1, r_2 pro $M'(\theta)$ a nahrazujeme je všude posloupnostmi

$$r_{1t} = K_1 t^{-\alpha}\,, \quad r_{2t} = K_2 \log(t + 1)\,, \quad K_1 < K_2\,, \quad 0 < \alpha < \tfrac{1}{2}\,.$$

8. ZÁKON ITEROVANÉHO LOGARITMU

Hledejme pro jednorozměrnou RM metodu $\{X_t\}$ takovou nenáhodnou posloupnost $\{\varphi_t\}$, $\varphi_t \nearrow +\infty$, pro kterou s pravděpodobností 1 platí:

$$|X_t - \theta| > (1 + \varepsilon)\,\varphi_t \quad \text{jen pro konečně mnoho } t\text{, ale}$$

$$|X_t - \theta| > (1 - \varepsilon)\,\varphi_t \quad \text{pro nekonečně mnoho } t\,, \quad \forall_{\varepsilon > 0}\,.$$

Za jistých předpokladů je touto posloupností

$$\varphi_t = K_1 \left(\frac{2 \log \log t}{t} \right)^{1/2}.$$

Věta 8.1. Nechť jsou splněny předpoklady věty 7.3 s tím, že předpoklad $\mathsf{E}e^2(t, x) < < K$ je zesílen na

$$\mathsf{E}|e(t, x)|^{2+\delta} < K \quad \text{pro nějaké} \quad \delta > 0\,, \tag{8.1}$$

pro všechna t a všechna x z nějakého okolí θ; dále nechť $\mathsf{E}X_1^2 < +\infty$. Potom s pravděpodobností 1 platí

$$\limsup_{t \to \infty} \left(\frac{t}{2 \log \log t} \right)^{1/2} |X_t - \theta| = \frac{a\sigma}{(2a|M'(\theta)| - 1)^{1/2}}. \tag{8.2}$$

Důkaz je podán v práci [10]; zde jej jen naznačíme: Lze psát (podobně jako v důkaze věty 7.2)

$$X_{t+1} - \theta = \left(1 - \frac{\gamma_t}{t}\right)(X_t - \theta) + \frac{a}{t}\, e(t + 1, X_t)\,, \tag{8.3}$$

kde $-\gamma_t = a(M'(\theta) + D(X_t))$, $D(x) \to 0$ pro $x \to \theta$. Iterováním vztahu (8.3) dostáváme

$$X_{t+1} - \theta = \beta_{0t}(X_1 - \theta) + \sum_{k=1}^{t} \frac{a}{k}\, \beta_{kt} e(k + 1, X_k)\,, \tag{8.4}$$

kde

$$\beta_{kt} = \prod_{j=k+1}^{t} \left(1 - \frac{\gamma_j}{j}\right), \quad 0 \leqq k \leqq t - 1\,; \quad \beta_{tt} = 1\,.$$

Zřejmě $1 - (\gamma_j/j) > 0$ pro $j \geqq j_0(\omega)$ s pravděpodobností 1; položíme-li

$$\beta_t = \prod_{j=j_0}^{t} \left(1 - \frac{\gamma_j}{j}\right),$$

lze psát $\beta_{kt} = \beta_t/\beta_k$ pro $t \geqq k \geqq j_0$. Vyšetřením logaritmu β_t zjistíme, že $\beta_t \cong \cong t^{-a|M'(\theta)|}\tau_t$, kde $\tau_t = \tau_t(\omega)$ je posloupnost s pomalou variací (tj. $\tau_t t^{\delta} \nearrow$, $\tau_t t^{-\delta} \searrow$ $\forall \delta > 0$, $t \geqq t_0(\delta)$). Přepíšeme tedy (8.4) ve tvaru

$$X_{t+1} - \theta = t^{-a|M'(\theta)|}\tau_t(\delta_0 + \sum_{k=1}^{t} \delta_k e(k + 1, X_k))\,, \tag{8.5}$$

kde δ_k jsou $\mathscr{F}_k$-měřitelné a $\delta_k \cong a k^{a|M'(\theta)|-1}\tau_k^{-1}$ pro $k \to +\infty$. Nyní se využije (i) faktu, že $\delta_k e(k + 1, X_k)$ v (8.5) jsou martingalové diference, (ii) vzorce

$$\sum_{k=1}^{t} k^r \tau_k \cong \frac{t^{r+1}}{r + 1}\, \tau_t \quad \text{pro} \quad t \to \infty\,,$$

který platí pro každou polsoupnost $\{\tau_k\}$ s pomalou variací a pro $r > -1$. Tento vzorec umožňuje vypočítat

$$s_k^2 = \sum_{k=1}^{t} \mathsf{E}(\delta_k^2 e^2(k+1, X_k) \,|\, \mathscr{F}_k) \cong \frac{a^2\sigma^2}{2a|M'(\theta)| - 1} \cdot \frac{t^{2a|M'(\theta)|-1}}{\tau_t^2}$$

a užít zákona iterovaného logaritmu pro martingaly, [28], z něhož plyne

$$\limsup_{t\to\infty} (2s_t^2 \log\log t)^{-1/2} \sum_{k=1}^{t} \delta_k e(k+1, X_k) = 1\,,$$

a po dosazení do (8.5) i tvrzení věty. □

Předpoklad (8.1) lze nahradit slabším, avšak méně přehledným požadavkem existence $\mathscr{F}_t$-měřitelných k_t takových, že $k_t \to 0$ s pravděpodobností 1 a

$$\sum_{t=2}^{\infty} \frac{\log\log t}{t k_t^2} \mathsf{E}(e^2(t+1, X_t)\, I_{[e^2(t+1,X_t) > t k_t^2/\log\log t]} \,|\, \mathscr{F}_t) < +\infty\,.$$

9. ZTRÁTY PŘI RM METODĚ A PŘI ADAPTIVNÍ METODĚ

Při užití stochastické aproximace často sledujeme dvojí cíl: přiblížit se co nejrychleji hledané hodnotě θ, ale tak, že i jednotlivé experimenty se provádějí na úrovni co nejbližší hodnotě θ. Je-li např. θ optimální dávka nějakého léku, jde nejen o to ji nalézt, ale i o to, co nejlépe léčit pacienty už v průběhu hledání.

Předpokládejme, že provedení pokusu na úrovni x znamená ztrátu (úměrnou) $(x - \theta)^2$. Jde pak tedy při stochastické aproximaci také o to, aby celková ztráta za t kroků metody,

$$C_t = \sum_{k=1}^{t} (X_k - \theta)^2\,, \tag{9.1}$$

byla minimální.

Vyšetřujeme jen jednorozměrný případ RM metody (3.1) a předpokládáme, že

$$\sup_{\varepsilon < |x-\theta| < 1/\varepsilon} M(x)(x - \theta) < 0\,, \quad \forall_{\varepsilon > 0}\,,$$

$$|M(x)| \leqq K(1 + |x|)\,, \quad \forall x\,,$$

existuje vlastní derivace $M'(\theta) < 0$ (označme $|M'(\theta)| = m$), $e(t, x) = e(t)$ nezávisí na x a jsou stejně rozdělené pro všechna t, $\mathsf{E}\, e(t) = 0$, $\mathsf{E}\, e^2(t) = \sigma^2 < +\infty$; $a_t = a/t$, $2am > 1$. Jak vyplývá z vět v předchozích odstavcích, platí pak $\lim X_t = \theta$ s pravděpodobností 1, veličiny $t^{1/2}(X_t - \theta)$ jsou asymptoticky normální s parametry $(0, a^2\sigma^2/(2am - 1))$ a $\limsup_{t\to\infty} (t/2 \log\log t)^{1/2} |X_t - \theta| = a\sigma/(2am - 1)^{1/2}$; navíc platí:

Věta 9.1. $C_t \cong (a^2\sigma^2/(2am - 1)) \log t$ pro $t \to \infty$, s pravděpodobností 1.

Důkaz plyne podobně jako zákon iterovaného logaritmu z reprezentace (8.5), $X_{t+1} - \theta \cong t^{-am}\tau_t \sum_{k=1}^{t} \delta_k e(k+1)$. Věta 9.1 i další výsledky v tomto odstavci jsou převzaty z práce Lai-Robbins [18]. □

Při adaptivní metodě, kterou jsme zavedli v odst. 6, se v i-tém kroku provádějí dvě pozorování Y_i' a Y_i'' na úrovních $X_i' = X_i + c_i$ a $X_i'' = X_i - c_i$; pro $t = 2s$ je pak celková ztráta

$$C_t = \sum_{i=1}^{s} (X_i' - \theta)^2 + \sum_{i=1}^{s} (X_i'' - \theta)^2 . \tag{9.2}$$

Učiňme tytéž předpoklady jako ve větě 6.1, s $\frac{1}{4} < \gamma < \frac{1}{2}$. Dále nechť $e(t+1, X_t') = e'(t+1)$ a $e(t+1, X_t'') = e''(t+1)$ nezávisí na x a $e'(2), e''(2), e'(3), e''(3), \ldots$, jsou nezávislé, stejně rozdělené se střední hodnotou 0 a rozptylem $\sigma^2 < +\infty$.

Věta 9.2. Pro C_t definované vztahem (9.2) platí s pravděpodobností 1

$$C_t = \frac{4^\gamma a^2}{1 - 2\gamma} t^{1-2\gamma} \quad \text{pro} \quad t \to \infty .$$

Nahradíme-li v (6.7) $c_s = c/s^\gamma$ volbou $c_s = c/(s \log s)^{1/2}$ a definici W_t nahradíme definicí

$$W_t = \frac{\sum_{i=1}^{t} c_i(Y_i' - Y_i'')}{2 \sum_{i=1}^{t} c_i^2} ,$$

potom s pravděpodobností 1

$$C_t = \frac{\sigma^2}{m^2} \log t .$$

(Přitom konvergence $X_t \to \theta$, $W_t \to m$ s pravděpodobností 1 zůstávají v platnosti, rovněž asymptotická normalita $t^{1/2}(X_t - \theta)$ s minimálním rozptylem.)

V [18] je také ukázáno, že není třeba dvou experimentů na každém kroku k odhadu derivace $M'(\theta)$; stačí její odhad metodou nejmenších čtverců z prvých t pozorování:

Věta 9.3. Nechť M je borelovsky měřitelná,

$$\inf_{\varepsilon < |x - \theta| < 1/\varepsilon} M(x)(x - \theta) > 0 , \quad \forall_{\varepsilon > 0} ,$$

$$|M(x)| < K(1 + |x|) , \quad \forall_{x \in R} ,$$

M je spojitě diferencovatelná v nějakém okolí θ, $M'(\theta) = m > 0$.

Definujme adaptivní RM metodu předpisem: X_1 libovolné, $X_{t+1} = X_t -$

$-(1/t[W_t])\,Y_t$, $1 \leqq t < +\infty$, kde $Y_t = M(X_t) + e(t+1)$; $X_1, e(2), e(3), \ldots$ nezávislé náhodné veličiny, $e(t)$, $t \geqq 2$, stejně rozdělené, $\mathsf{E}\,e(t) = 0$, $\mathsf{E}\,e^2(t) = \sigma^2 > 0$,

$$W_t = \frac{\sum_{i=1}^{t}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{t}(X_i - \bar{X})^2}\,, \quad \bar{X} = \frac{1}{t}\sum_{i=1}^{t} X_i\,, \quad \bar{Y} = \frac{1}{t}\sum_{i=1}^{t} Y_i\,,$$

$[W_t] = [W_t]_{r_1}^{r_2}$ pro nějaká $0 < r_1 < m < r_2 < +\infty$.

Potom platí

$$\lim_{t\to\infty} X_t = \theta\,, \quad \lim_{t\to\infty}[W_t] = m \quad \text{s pravděpodobností 1}\,;$$

$$t^{1/2}(X_t - \theta) \quad \text{je asymptoticky normální s parametry} \quad (0, \sigma^2/m^2)\,;$$

$$\limsup_{t\to\infty}\left(\frac{t}{2\log\log t}\right)^{1/2}|X_t - \theta| = \frac{\sigma}{m} \quad \text{s pravděpodobností 1}\,;$$

$$C_t = \sum_{i=1}^{t}(X_i - \theta)^2 \cong \frac{\sigma^2}{m^2}\log t \quad \text{pro} \quad t \to \infty\,, \quad \text{s pravděpodobností 1}\,.$$

Je-li navíc splněno

$$M(x) = m(x - \theta) + O(|x - \theta|^{1+\eta}) \quad \text{pro} \quad x \to \theta \quad \text{a nějaké} \quad \eta > 0\,,$$

potom také

$$(\log t)^{1/2}([W_t] - m) \text{ je asymptoticky normální s parametry } (0, m^2)\,.$$

10. ASYMPTOTICKÁ VYDATNOST. ROBUSTNOST

Uvažujme jednorozměrnou RM posloupnost $\{X_t\}$ definovanou v (3.1); předpokládejme, že

$$M(x) \gtreqless 0 \quad \text{pro} \quad x \lesseqgtr \theta\,, \tag{10.1}$$

$$K|x - \theta| \leqq |M(x)| \leqq K_1|x - \theta|\,, \quad \forall_x\,, \tag{10.2}$$

$$\text{existuje vlastní} \quad M'(\theta) < 0 \quad (\text{označme } |M'(\theta)| = m)\,; \tag{10.3}$$

$$\mathsf{E}e^2(t, x) < K_2\,, \quad \forall_{t,x}\,, \tag{10.4}$$

$$\lim_{t\to\infty, x\to\theta} \mathsf{E}e^2(t, x) = \sigma^2\,, \tag{10.5}$$

$$\lim_{R\to\infty}\ \lim_{\varepsilon\to 0}\ \sup_{|x-\theta|<\varepsilon} \mathsf{E}(e^2(t, x)\, I_{[|e(t,x)|>R]}) = 0\,; \tag{10.6}$$

$$a_t = a/t\,, \quad 2aK > 1\,. \tag{10.7}$$

Víme již, že za těchto předpokladů (které jsou o něco silnější než je třeba) je

$t^{1/2}(X_t - \theta)$ asymptoticky normální, volba $a = 1/m$ vede k minimálnímu asymptotickému rozptylu σ^2/m^2, a i když m neznáme, lze téhož výsledku dosáhnout adaptivní metodou.

Vyšetříme, zda lze asymptotický rozptyl ještě zmenšit vhodnou transformací pozorovaných veličiny, tj. zavedením $g(Y)$ místo Y v aproximačním schématu.

Budeme uvažovat jen takové transformace g, pro které jak transformovaná regresní funkce

$$M_g(x) = \mathsf{E}g(M(x) + e(t, x))$$

tak i náhodné odchylky od ní

$$e_g(t, x) = g(M(x) + e(t, x)) - M_g(x)$$

splňují opět předpoklady (10.1)–(10.6), obecně s jinými konstantami K_g, K_{g1}, K_{g2}, $m_g = |M'_g(\theta)|$, σ_g^2, avšak s týmtéž θ. Potom tedy RM metoda aplikovaná na transformovaný problém,

$$X_{t+1} = X_t + \frac{a}{t} g(Y_t), \quad \text{kde} \quad Y_t = M(X_t) + e(t+1, X_t), \tag{10.8}$$

s $2aK_g > 1$, má opět vlastnost $\lim_{t \to \infty} X_t = \theta$ s pravděpodobností 1 a $t^{1/2}(X_t - \theta)$ je asymptoticky normální s parametry $(0, a^2\sigma_g^2/(2am_g - 1))$; volba $a = 1/m_g$ vede na minimální asymptotický rozptyl σ_g^2/m_g^2 (při daném g); označme jej S_g^2.

Omezíme se na případ, kdy (původní) $e(t, x) = e(t)$ nezávisí na x a jsou stejně rozdělené pro všechna t, s hustotou f, která nechť má kladnou konečnou Fisherovu informaci,

$$0 < I(f) = \int_{-\infty}^{\infty} \left(\frac{f'(y)}{f(y)}\right)^2 f(y)\,\mathrm{d}y < +\infty\,.$$

Učiníme-li formální předpoklady

$$\frac{\mathrm{d}}{\mathrm{d}x} \int_{\infty-}^{\infty} g(y) f(y + x)\,\mathrm{d}y\,\Big|_{x=0} = \int_{-\infty}^{\infty} g(y) f'(y)\,\mathrm{d}y\,, \tag{10.9}$$

$$\int_{-\infty}^{\infty} g(y) f'(y)\,\mathrm{d}y < 0\,, \tag{10.10}$$

$$\int_{-\infty}^{\infty} g^2(y + x) f(y)\,\mathrm{d}y \quad \text{je spojitý v bodě} \quad x = 0 \tag{10.11}$$

(třídu funkcí g, které splňují tyto i výše uvedené předpoklady, označme $\mathscr{G}$), potom dostáváme

$$m_g = -m \int_{-\infty}^{\infty} g(y) f'(y)\,\mathrm{d}y\,,$$

$$\sigma_g^2 = \int_{-\infty}^{\infty} g^2(y) f(y)\,\mathrm{d}y$$

a minimální asymptotický rozptyl lze psát jako

$$S_g^2 = \frac{\int_{-\infty}^{\infty} g^2 f \, \mathrm{d}y}{m^2 \left(\int_{-\infty}^{\infty} g f' \, \mathrm{d}y \right)^2}.$$

Nyní jde o to volit g tak, aby realizovalo $\min_{g \in \mathscr{G}} S_g^2$, označme je g_0. Dle Schwarzovy nerovnosti je

$$\left(\int_{-\infty}^{\infty} g f' \, \mathrm{d}y \right)^2 \leqq \int_{-\infty}^{\infty} g^2 f \, \mathrm{d}y \, . \int_{-\infty}^{\infty} \left(\frac{f'}{f} \right)^2 f \, \mathrm{d}y \,,$$

tj $S_g^2 \geqq 1/m^2 \, I(f)$ a rovnost nastane právě pro $g_0 = c \, . \, (f'/f)$, pokud $c \, . \, (f'/f) \in \mathscr{G}$. Z (10.10) vyplývá, že $c < 0$; bez újmy na obecnosti nechť $c = -1$.

Věta 10.1. Za předpokladu, že $g_0 = -f'/f \in \mathscr{G}$, je asymptotický rozptyl veličin $t^{1/2}(X_t - \theta)$, kontruovaných dle (10.8) s $g = g_0$, roven $1/m^2 \, I(f)$.

Tento výsledek, který dokázali nezávisle Anbar a Abdelhamid, viz [1], lze nazvat *asymptotickou vydatností* transformované RM metody, v tomto smyslu: Je-li regrese lineární, $M(x) = -m(x - \theta)$, pak $1/m^2 \, I(f)$ je právě Cramérova-Raova dolní hranice pro rozptyly regulárních nevychýlených odhadů θ.

Příklady.

(i) Je-li f hustota normovaného normálního rozdělení, $f(y) = [1/\sqrt{(2\pi)}] \, \mathrm{e}^{-y^2/2}$, pak $I(f) = 1$, $g_0(y) = -f'(y)/f(y) = y$, $g_0 \in \mathscr{G}$ triviálně. Tedy v případě normálního rozdělení je RM metoda *bez* transformace optimální; $S_{g_0}^2 = 1/m^2$, (resp. $= \sigma^2/m^2$, je-li f normální s rozptylem $\sigma^2 \neq 1$).

(ii) Je-li f hustota dvojně exponenciálního rozdělení, $f(y) = \frac{1}{2} \mathrm{e}^{-|y|}$, pak $I(f) = 1$, $g_0(y) = -f'(y)/f(y) = \operatorname{sign} y$, (10.9)–(10.11) je splněno neboť $\int_{-\infty}^{\infty} |g_0(y) f'(y + x)| \, \mathrm{d}y = 1$ $\forall_x$, $\int_{-\infty}^{\infty} g_0(y) f'(y) \, \mathrm{d}y = -1$, $\int_{-\infty}^{\infty} g_0^2(y + x) f(y) \, \mathrm{d}y = 1$, $\forall_x$. Je tedy $g_0 \in \mathscr{G}$ a $g_0(y) = \operatorname{sign} y$ je optimální transformace v aproximačním schématu (10.8).

(iii) Je-li f hustota Huberova rozdělení,

$$f(y) = \begin{cases} C\mathrm{e}^{-y^2/2} & \text{pro } |y| \leqq K \,, \\ C\mathrm{e}^{-K|y| + K^2/2} & \text{pro } |y| > K \,, \end{cases}$$

potom

$$g_0(y) = -\frac{f'(y)}{f(y)} = \begin{cases} y & \text{pro } |y| \leqq K \,, \\ K \operatorname{sign} y & \text{pro } |y| > K \,, \end{cases} \tag{10.12}$$

a $g_0 \in \mathscr{G}$, jak se snadno zjistí.

Nechť nyní o hustotě základního rozdělení f víme jen, že je tvaru

$$f(y) = (1 - \varepsilon) \, \varphi(y) + \varepsilon \, h(y)$$

kde $0 < \varepsilon < 1$ je pevné, φ je hustota normálního rozdělení a h hustota nějakého symetrického rozdělení; přitom nechť $I(f) < +\infty$. Třídu takovýchto hustot f (normálních s ε-příměsí) označme $\mathscr{F}$. Označme f_0 hustotu Huberova rozdělení, jehož parametr K je řešením rovnice

$$(1 - \varepsilon)^{-1} = 2\,\Phi(K) - 1 + 2\,\varphi(K)/K\,, \tag{10.13}$$

kde Φ je normální distribuční funkce; označme $g_0(y) = -f_0'(y)/f_0(y)$ k němu příslušnou transformaci (10.12). Snadno se ověří, že $f_0 \in \mathscr{F}$.

Označíme-li

$$S_{g,f}^2 = \frac{\displaystyle\int_{-\infty}^{\infty} g^2 f\,\mathrm{d}y}{m^2\left(\displaystyle\int_{-\infty}^{\infty} g f'\,\mathrm{d}y\right)^2}$$

minimální asymptotický rozptyl RM posloupnosti při transformaci $g \in \mathscr{G}$ a při základním rozdělení $f \in \mathscr{F}$, pak zjistíme, že dvojice f_0, g_0 má povahu sedlového bodu, tj. že

$$S_{g_0,f}^2 \leqq S_{g_0,f_0}^2 \leqq S_{g,f_0}^2 \quad \forall f \in \mathscr{F}\,, \quad g \in \mathscr{G}\,.$$

Pravá větev této nerovnosti je vlastně tvrzení věty 10.1, o platnosti levé větve se lze přesvědčit dosazením do $S_{g_0,f}^2$ za $f = (1 - \varepsilon)\,\varphi + \varepsilon h$, za g_0 dle (10.12) a úpravami.

Je tedy g_0 dané vztahem (10.12) a (10.13) minimaxově optimální transformací vzhledem ke třídě $\mathscr{F}$ normálních rozdělení s ε-příměsí. Transformovaná RM metoda je v tomto smyslu *robustní*.

11. RM METODA V PŘÍPADĚ, ŽE I NEZÁVISLE PROMĚNNÁ JE ZATÍŽENA CHYBOU

Dosud jsme předpokládali, že pouze funkční hodnota $M(x)$ je pozorována s experimentální chybou, zatímco úroveň x experimentu lze nastavit přesně. Nyní se budeme zabývat situací, kdy ani to není možné a kdy tedy i hodnota nezávisle proměnné x je zatížena chybou. Omezíme se na jednorozměrný případ.

RM posloupnost $\{X_t\}$ je pak popsána vztahem

$$\begin{aligned} X_1 \text{ libovolné}\,, \quad X_{t+1} = X_t + a_t(M(X_t + d(t+1, X_t)) + \\ + e(t+1, X_t + d(t+1, X_t))),\ 1 \leqq t < +\infty \end{aligned} \tag{11.1}$$

kde náhodná veličina X_1 a měřitelné náhodné funkce $d(2, x)$, $e(2, x)$ $d(3, x)$, $e(3, x), \ldots$ jsou vesměs nezávislé, s $\mathsf{E}d(t, x) = 0$, $\mathsf{E}e(t, x) = 0$, $\forall_{t,x}$.

Uvažujme nejprve situaci, kdy experimentátor nemůže ovlivnit (zmenšit) chybu v nastavení hodnoty x; vyjádřeme to předpokladem, že

$$\text{rozdělení } d(t, x) \text{ nezávisí na } t\,. \tag{11.2}$$

Jak lze očekávat, platí konvergence $X_t \to \theta$ jen za značně speciálních předpokladů:

Věta 11.1. Nechť funkce M je antisymetrická vzhledem k θ, klesající a taková, že

$$|M(x_2) - M(x_1)| \leqq K + K_1|x_2 - x_1|, \quad \forall_{x_1,x_2}.$$

Posloupnost $\{X_t\}$ nechť je definována pomocí (11.1), (11.2), kde dále $\sum_{t=1}^{\infty} a_t = +\infty$, $\sum_{t=1}^{\infty} a_t^2 < +\infty$, pro každé x je rozdělení veličiny $d(2, x)$ symetrické, $\mathsf{E}X_1^2 < +\infty$, $\mathsf{E}\, d^2(2, x) \leqq K_2$, $\mathsf{E}e^2(t, x) \leqq K_3$, $\forall_{t,x}$.

Potom $\lim_{t\to\infty} (X_t - \theta_t) = 0$ s pravděpodobností 1 i podle kvadratického středu.

Důkaz je snadný: RM procedura $\{X_t\}$ řeší nyní ve skutečnosti úlohu $M^*(x) = 0$, kde $M^*(x) = \mathsf{E}\, M\, (x + d\, (2, x))$. Hořejší podmínky pak zaručují, že obě úlohy, $M(x) = 0$ a $M^*(x) = 0$, mají totéž řešení θ a aproximace se k němu blíží. □

Dále uvažujme odlišnou situaci, kdy přesnost v nastavení hodnot x lze v průběhu aproximací zvyšovat avšak pouze za cenu nepřímo úměrnou dosaženému rozptylu $\mathsf{E}d^2(t, x)$. Vyšetříme asymptotické chování střední kvadratické chyby $\mathsf{E}(X_t - \theta)^2$, a to nikoli jako funkce počtu pozorování t, nýbrž jako funkce celkového nákladu T na prvních t experimentů.

Věta 11.2. Nechť pro funkci M platí $M(\theta) = 0$,

$$-K_1 \leqq \frac{M(x_2) - M(x_1)}{x_2 - x_1} \leqq -K, \quad \forall_{x_1,x_2};$$

posloupnost $\{X_t\}$ nechť je definována pomocí (11.1), kde dále $a_t = a/t$, $2aK > 1$, $\mathsf{E}X_1^2 < \infty$, $\mathsf{E}d^2(t, x) \leqq K_2/t^{\beta}$, $\forall_{t,x}$, $\beta > 0$ je volitelné, $\mathsf{E}e^2(t, x) \leqq K_3$, $\forall_{t,x}$. Nechť $T = K_4 \sum_{j=1}^{t} j^{\beta}$ (celkový náklad na t experimentů), nechť $B_T = \mathsf{E}(X_t - \theta)^2$ (střední kvadratická chyba jako funkce celkového nákladu).

Potom $X_t \to \theta$ podle kvadratického středu,

$$B_T = \begin{cases} O(T^{-\beta/(1+\beta)}) & \text{pro} \quad 0 < \beta \leqq 1, \\ O(T^{-1/(1+\beta)}) & \text{pro} \quad \beta \geqq 1, \end{cases}$$

a volba $\beta = 1$ (která zaručuje, že $B_T = O(T^{-1/2})$) je optimální v tomto smyslu: je-li $\beta \neq 1$, pak existují $M(x)$, $d(t, x)$ a $e(t, x)$, vyhovující všem předpokladům věty a takové, že $B_T \geqq K_5 T^{-1/2+\varepsilon}$.

Důkazy vět 11.1 a 11.2 viz **Dupač-Král** [7].

Výsledek obsažený ve větě 11.2 potvrzuje naši intuitivní představu, že je zbytečné platit za vysokou přesnost v nastavení x v počáteční fázi aproximačního procesu; výsledek navíc ukazuje, jak optimálně rozvrhnout náklady.

12. DYNAMICKÁ RM APROXIMAČNÍ METODA

V některých situacích je nereálné předpokládat, že regresní funkce M zůstává neměnná v čase; naopak, musíme připustit, že se i v průběhu aproximačního procesu mění. Omezíme se na případ, kdy touto změnou je pouze posunutí.

Věta 12.1. Nechť v čase t platí regresní funkce $M(t, x) = M^0(x - \theta_t)$, $1 \leqq t < +\infty$, pozorovatelná jen s experimentální chybou, $M(t, x) + e(t, x)$, kde

$$(12.1) \qquad M^0(x) \gtreqless 0 \quad \text{pro} \quad x \gtreqless 0\,, \qquad K|x| \leqq |M^0(x)| \leqq K_1|x|\,, \quad \forall_x\,,$$

$$(12.2) \qquad \theta_t = b + ct \quad (\text{avšak s neznámými koeficienty } b, c)\,,$$

$\{e(t, x), x \in R\}_{t=2}^{\infty}$ jsou nezávislé měřitelné náhodné funkce s $\mathsf{E}e(t, x) = 0$, $\mathsf{E}e^2(t, x) \leqq K_2$, $\forall_{t,x}$.

Modifikujme RM posloupnost takto:

$$(12.3) \qquad X_1 \text{ libovolné}\,, \quad \mathsf{E}X_1^2 < +\infty\,; \quad X_{t+1} = X_t^* + a_t Y_t^*\,, \quad 1 \leqq t < +\infty\,,$$

kde $X_t^* = (1 + 1/t)\, X_t$, $Y_t^* = M(t + 1, X_t^*) + e(t + 1, X_t^*)$ a posloupnost kladných konstant a_t splňuje

$$\lim_{t\to\infty} t a_t = +\infty\,, \quad \sum_{t=1}^{\infty} a_t^2 < +\infty\,.$$

Potom $\lim_{t\to\infty}(X_t - \theta_t) = 0$ s pravděpodobností 1 i podle kvadratického středu. Tvrzení zůstává v platnosti, nahradíme-li (12.2) obecnějším předpokladem

$$(12.4) \qquad \theta_{t+1} - \left(1 + \frac{1}{t}\right)\theta_t = o(a_t)\,.$$

D ů k a z je podán v práci [4]. Zde dokážeme jen poněkud speciálnější větu 12.2.

Poznámka. Modifikace (12.3) se nazývá dynamická RM aproximační metoda. Má následující smysl: V čase $t + 1$ chceme určit přiblížení X_{t+1} k θ_{t+1}; vyjdeme z předchozího přiblížení X_t k θ_t, provedeme korekci na trend, $X_t^* = (1 + 1/t)\, X_t$, potom odhadneme hodnotu okamžité regresní funkce $M(t + 1, .)$ v bodě X_t^* pomocí pozorování Y_t^* a konečně provedeme další korekci, přičtením $a_t Y_t^*$.

Věta 12.2. Za předpokladů věty 12.1, kde volíme speciálně $a_t = a/t^\alpha$, $a > 0$, $\frac{1}{2} < \alpha < 1$, a předpoklad (12.4) zesílíme na

$$(12.5) \qquad \theta_{t+1} - \left(1 + \frac{1}{t}\right)\theta_t = O(t^{-\omega}) \quad \text{pro nějaké} \quad \omega > \alpha\,,$$

platí

$$\mathsf{E}(X_t - \theta_t)^2 = \begin{cases} O(t^{-\alpha}) & \text{pro} \quad \alpha \leqq \frac{2}{3}\omega\,, \\ O(t^{-2(\omega-\alpha)}) & \text{pro} \quad \alpha > \frac{2}{3}\omega\,. \end{cases}$$

Předpoklad (12.5) je splněn zejména v případě lineárního trendu (12.2) a to s $\omega = 1$.

Důkaz. Dle (12.1) je $M^0(x) = x\mu$, kde μ je funkce x taková, že $K \leqq -\mu \leqq K_1, \forall_x$; odtud $M(t, x) = (x - \theta_t)\,\mu$ a dle (12.5)

$$M(t + 1, X_t^*) = (1 + t^{-1})(X_t - \theta_t)\,\mu + O(t^{-\omega})\,\mu\,,$$

kde $O(t^{-\omega})$ je nenáhodná veličina. Dále je $\mathsf{E}(Y_t^{*2} \mid \mathscr{F}_t) \leqq K_3 + K_4(X_t - \theta_t)^2$. S pomocí těchto vztahů zjistíme, že

$$\mathsf{E}((X_{t+1} - \theta_{t+1})^2 \mid \mathscr{F}_t) \leqq \left(1 - \frac{K_5}{t^\alpha}\right)(X_t - \theta_t)^2 + \frac{K_6}{t^\omega}\left|X_t - \theta_t\right| + \frac{K_7}{t^{2\alpha}}\,;$$

užitím nerovnosti $\mathsf{E}|Z| \leqq \varepsilon + \varepsilon^{-1}\mathsf{E}Z^2$, kde položíme $\varepsilon = 1/(\delta t^{\omega-\alpha})$ pro nějaké malé $\delta > 0$, dostaneme pak

$$(12.6) \qquad \mathsf{E}(X_{t+1} - \theta_{t+1})^2 \leqq \begin{cases} \left(1 - \dfrac{K_8}{t^\alpha}\right)\mathsf{E}(X_t - \theta_t)^2 + \dfrac{K_9}{t^{2\alpha}} & \text{pro} \quad \alpha \leqq \tfrac{2}{3}\omega\,, \\[2ex] \left(1 - \dfrac{K_8}{t^\alpha}\right)\mathsf{E}(X_t - \theta_t)^2 + \dfrac{K_{10}}{t^{2\omega-\alpha}} & \text{pro} \quad \alpha > \tfrac{2}{3}\omega\,. \end{cases}$$

Tvrzení věty pak plyne z Chungovy lemmy (citované v důkazu věty 4.1). □

Výsledky o dynamické aproximační metodě lze v mnoha směrech zobecnit – na vícerozměrný případ, při slabších požadavcích na funkci M, pro obecnější trend. Následující věta je malou obměnou věty z práce [6].

Věta 12.3. Nechť funkce $M^0 : R^l \to R^l$ a náhodné funkce $\{e^0(t, x),\ x \in R^l\}_{t=1}^{\infty}$, splňují všeobecné předpoklady vyslovené v úvodu odst. 3 a dále (3.6), (3.7), (7.2), (7.4) a (7.6), vše s $\theta = 0$. Okamžitá regresní funkce a její pozorovací chyby nechť jsou definovány jako

$$M(t, x) = M^0(x - \theta_t)\,, \quad e(t, x) = e^0(t, x - \theta_t)\,,$$

kde θ_t splňují rekurentní vztah

$$\theta_{t+1} = (I + Q_t)\,\theta_t + q_t\,, \quad 1 \leqq t < +\infty\,,$$

v němž Q_t je známá matice a q_t obecně neznámý vektor, přičemž

$$\lim_{t\to\infty} t^\alpha Q_t = 0\,, \quad \lim_{t\to\infty} t^{3\alpha/2} q_t = q_\infty \quad \text{pro nějaké} \ \tfrac{1}{2} < \alpha < 1 \quad \text{a} \quad q_\infty \in R^l\,.$$

Definujme dynamickou RM posloupnost takto:

$$X_1 \ \text{libovolné}\,, \quad X_{t+1} = X_t^* + \frac{a}{t^\alpha}\,Y_t^*\,,$$

kde

$$X_t^* = (I + Q_t)\,X_t\,, \quad Y_t^* = M(t + 1, X_t^*) + e(t + 1, X_t^*)\,, \quad a > 0\,.$$

Potom $\lim_{t\to\infty}(X_t - \theta_t) = 0$ s pravděpodobností 1 a asymptotické rozdělení $t^{\alpha/2}(X_t - \theta_t)$ je l-rozměrné normální s parametry $(a^{-1}B^{-1}q_\infty,\ a\int_0^\infty e^{Bv}\,\Sigma\,e^{B^Tv}\,dv)$.

13. RM METODA VE SPOJITÉM ČASE

Na rekurentní vzorec (3.1) popisující metodu RM lze pohlížet také jako na stochastickou diferenční rovnici. Představme si nyní, že jednotlivé kroky aproximační metody následují za sebou v časových intervalech délky h a že také rozptyly náhodných chyb jsou úměrné h. Blíží-li se pak h k nule, je intuitivně zřejmé, že stochastická diferenční rovnice přejde ve stochastickou diferenciální rovnici. Obyčejně se volí *Itôova stochastická diferenciální rovnice*, neboť její matematická teorie je přímočará (viz např. [21]).

RM metodě v diskrétním čase

$$X_{t+1} = X_t + a_t(M(X_t) + e(t + 1, X_t))$$

bude tedy ve spojitém čase odpovídat stochastická diferenciální rovnice

$$\mathrm{d}X(t) = a(t)\,(M(X(t))\,\mathrm{d}t + \sigma(t, X(t))\,\mathrm{d}W(t)) , \tag{13.1}$$

kde $W(t)$, $t \geqq 0$, je Wienerův proces (tj. proces s nezávislými přírůstky, se spojitými trajektoriemi, takový, že $W(0) = 0$ a pro $t > 0$ má $W(t)$ normální rozdělení s $\mathsf{E}\,W(t) = 0$, $\mathsf{E}\,W^2(t) = t$);

$M(x)$, $x \in R$, je lipschitzovsky spojitá a taková, že $M(x) \gtreqqless 0$ pro $x \lesseqqgtr \theta$ (pro nějaké θ);

$\sigma(t, x)$, $t \geqq 0$, $x \in R$, je spojitá funkce dvou proměnných splňující

$$|\sigma(t, x) - \sigma(t, y)| \leqq K|x - y| , \quad \sigma^2(t, x) \leqq K_1(1 + x^2) \text{ pro všechna } t, x, y;$$

$a(t)$, $t \geqq 0$, je spojitá funkce taková, že

$$a(t) > 0 , \quad \int_0^\infty a(t)\,\mathrm{d}t = +\infty , \quad \int_0^\infty a^2(t)\,\mathrm{d}t < +\infty .$$

Platí, že za těchto předpokladů má stochastická diferenciální rovnice (13.1) jediné spojité řešení vyhovující počáteční podmínce $X(0) = x_0$ a že toto řešení konverguje s pravděpodobností 1 k θ pro $t \to \infty$, pro každé $x_0 \in R$.

I ve vícerozměrném případě platí pro RM metodu se spojitým časem konvergenční věty obdobné větám pro RM metodu s diskrétním časem, viz [22].

Všimněme si ještě dalších asymptotických vlastností řešení $X(t)$, viz Pflug [23].

Nechť nejprve v (13.1) je regresní funkce M lineární,

$$M(x) = -m\,.\,(x - \theta) , \quad \text{kde} \quad m > 0 ,$$

a nechť σ nezávisí na x, tj.

$$\sigma(t, x) = \sigma(t) .$$

Řešení rovnice (13.1) s počáteční podmínkou $X(0) = x_0$ lze pak nalézt explicitně,

$$X(t) = \Phi(t)\,(1 + V(t)) ,$$

kde

$$\mathrm{d}\Phi(t) = -m\, a(t)\, \Phi(t)\, \mathrm{d}t\,, \quad \Phi(0) = x_0\,, \tag{13.2}$$

$$\mathrm{d}V(t) = \frac{a(t)\,\sigma(t)}{\Phi(t)}\, \mathrm{d}W(t)\,, \quad V(0) = 0\,. \tag{13.3}$$

Odtud je patrné, že $X(t)$, $t \geqq 0$, je proces s nezávislými přírůstky, a že rozdělení $X(t)$ je normální s $\mathsf{E}\, X(t) = \Phi(t)$, $\operatorname{var} X(t) = \Phi^2(t)\, S(t)$, kde

$$S(t) = \int_0^t \frac{a^2(u)\,\sigma^2(u)}{\Phi^2(u)}\, \mathrm{d}u\,. \tag{13.4}$$

Je-li

$$0 < \liminf_{u\to\infty} \sigma^2(u) < \limsup_{u\to\infty} \sigma^2(u) < +\infty\,,$$

pak platí i zákon iterovaného logaritmu,

$$\limsup_{t\to\infty} \frac{|X(t)|}{\Phi(t)\,(2S(t)\log\log S(t))^{1/2}} = 1 \tag{13.5}$$

s pravděpodobností 1.

Je-li v (13.1) stále $M(x) = -m\,.\,(x - \theta)$, avšak připustíme obecné $\sigma(t, x)$, pak se hořejší výsledky zachovávají asymptoticky; důkaz je však již netriviální – [23].

Věta 13.1. Nechť $X(t)$ je řešení rovnice

$$\mathrm{d}X(t) = a(t)\,(-m(X(t) - \theta)\,\mathrm{d}t + \sigma(t, X(t))\,\mathrm{d}W(t))$$

s počáteční podmínkou $X(0) = x_0$. Předpokládáme $m > 0$, $0 < K_1 \leqq \sigma(t, x) \leqq$ $\leqq K_2 < +\infty$, píšeme $\sigma(t)$ místo $\sigma(t, \theta)$; $\Phi(t)$ a $S(t)$ definujeme opět vztahy (13.2) a (13.4); nechť je splněno

$$\frac{\Phi^4(t)}{a^2(t)} \int_0^t \frac{a^2(u)}{\Phi^2(u)}\, \mathrm{d}u \to 0 \quad \text{pro} \quad t \to \infty\,.$$

Potom

$$\mathsf{E}\, X(t) = \Phi(t), \ \forall_t, \ \frac{\operatorname{var} X(t)}{\Phi^2(t)\, S(t)} \to 1 \quad \text{pro} \quad t \to \infty\,,$$

$X(t)$ je asymptoticky normální s parametry $\Phi(t)$ a $\Phi^2(t)\, S(t)$ nezávisle na počáteční hodnotě x_0 a platí zákon iterovaného logaritmu (13.5).

Konečně následující věta umožňuje přenést asymptotické vlastnosti z případu lineární regrese M na obecný případ (stačí v ní položit za funkci $M_1(x) = M'(\theta)\,.$ $.\,(x - \theta)$.

Věta 13.2. Mějme rovnice

$$\mathrm{d}X(t) = a(t)\,(M(X(t))\,\mathrm{d}t + \sigma(t, X(t))\,\mathrm{d}W(t))$$
$$\mathrm{d}Y(t) = a(t)\,(M_1(Y(t))\,\mathrm{d}t + \sigma(t, Y(t))\,\mathrm{d}W(t))\,.$$

Nechť

$$M(x)(x - \theta) \leqq -K(x - \theta)^2, \quad M_1(x)(x - \theta) \leqq -K(x - \theta)^2$$

pro všechna x a nějaké $K > 0$;

$$0 < K_1 \leqq \sigma(t, x) \leqq K_2 < +\infty, \quad \forall_{t,x}.$$

Funkce $\Phi_j(t)$, $t \geqq 0$, $0 \leqq j < +\infty$, definujme rekurentně:

$$\Phi_0(t) = 1,$$

$$\mathrm{d}\Phi_j(t) = -2jKa(t)\,\Phi_j(t)\,\mathrm{d}t + a^2(t)\,\Phi_{j-1}(t)\,\mathrm{d}t.$$

Jestliže pro nějaké přirozené k platí

$$\int_0^\infty \Phi_k(t)\,\mathrm{d}t < +\infty \quad \text{a současně}$$

$$M(x) - M_1(x) = O(|x - \theta|^k) \quad \text{pro} \quad x \to \theta,$$

potom náhodné procesy $X(t)$ a $Y(t)$, $t \geqq 0$, mají totéž asymptotické rozdělení.

14. KIEFEROVA - WOLFOWITZOVA APROXIMAČNÍ METODA

Nechť M je reálná funkce l reálných proměnných nabývající své maximální hodnoty v jediném bodě θ. Opět předpokládáme, že v čase t můžeme funkci M v libovolně zvoleném bodě x nevychýleně odhadnout pozorovaním Y. Je-li M hladká a je-li θ její jediný stacionární bod, pak nalézt jej znamená nalézt řešení soustavy rovnic $\nabla M(x) = 0$; tuto soustavu pak můžeme – jak se na první pohled zdá – řešit RM aproximační metodou.

V této úvaze je však chyba, neboť ve skutečnosti máme možnost nevychýleně odhadovat jen funkci M, nikoli její gradient ∇M. Nahrazuje se proto gradient ∇M přibližným gradientem $\nabla_c M$, což je l-rozměrný vektor, jehož i-tá souřadnice je dána podílem přírůstků

$$\frac{M(x + c\varepsilon_i) - M(x - c\varepsilon_i)}{2c},$$

kde ε_i je i-tý jednotkový vektor v R^l, tj. vektor o souřadnicích $\delta_{ij} = \begin{cases} 1 & \text{pro} \quad j = i, \\ 0 & \text{pro} \quad j \neq i \end{cases}$ $1 \leqq j \leqq l$. Přitom se $c = c_t$ se volí jako posloupnost kladných čísel konvergující k 0 pro $t \to \infty$. Vektor náhodných chyb s nimiž v čase $t + 1$ pozorujeme (poloviční) diference $\frac{1}{2}[M(x + c_t\varepsilon_i) - M(x - c_t\varepsilon_i)]$, $1 \leqq i \leqq l$, označme $e(t + 1, x)$. O posloupnosti $\{e(t, x), x \in R^l\}_{t=2}^\infty$ učiňme tytéž předpoklady jako v úvodu odst. 3 (tj. předpoklad nezávislosti a nulových středních hodnot).

Kieferova - Wolfowitzova (dále jen KW) *stochastická aproximační metoda* je pak popsána předpisem: Zvolme libovolně X_1 ($\mathscr{F}_1$-měřitelné), dále definujme rekurentně

$$X_{t+1} = X_t + a_t \nabla_{c_t} M(X_t) + \frac{a_t}{c_t} e(t+1, X_t), \quad 1 \leqq t < \infty, \tag{14.1}$$

kde $\{a_t\}$, $\{c_t\}$ jsou omezené posloupnosti kladných konstant.

Pro KW metodu platí obměny konvergenčních vět z odst. 3; také jejich důkazy jsou zcela obdobné. Vyslovme zejména analogie vět 3.2, 3.3., viz [22].

Věta 14.1. Nechť funkce M má spojité parciální derivace 2. řádu, které splňují Lipschitzovu podmínku. Nechť existuje kladně definitní matice C tak, že

$$\sup_{|x-\theta|>\varepsilon} (\nabla M(x), C(x-\theta)) < 0, \quad \forall_{\varepsilon>0}; \tag{14.2}$$

dále nechť

$$|\nabla M(x)|^2 + \mathsf{E}|e(t,x)|^2 \leqq K_1(1 + |x|^2), \quad \forall_x, \tag{14.3}$$

$$\sum_{t=1}^{\infty} a_t = +\infty, \quad \sum_{t=1}^{\infty} a_t c_t^2 < +\infty, \quad \sum_{t=1}^{\infty} a_t^2 c_t^{-2} < +\infty. \tag{14.4}$$

Potom platí $\lim_{t\to\infty} X_t = \theta$ s pravděpodobností 1.

Důkaz. Podobně jako v důkaze věty 3.1, resp. 3.2, s $V(x) = (C(x-\theta), x-\theta)$, je

$$LV(x) = \mathsf{E}\, V(x + a_t \nabla_{c_t} M(x) + a_t c_t^{-1} e(t+1, x)) - V(x) \leqq$$
$$\leqq 2a_t(\nabla_{c_t} M(x), C(x-\theta)) + K_2 a_t^2 |\nabla_{c_t} M(x)|^2 + K_2 a_t^2 c_t^{-2}\, \mathsf{E}|e(t+1,x)|^2 \leqq$$
$$\leqq -2a_t|(\nabla M(x), C(x-\theta))| + K_3 a_t^2 c_t^{-2}(1 + V(x)),$$

kde jsme použili toho, že $|\nabla_{c_t} M(x) - \nabla M(x)| \leqq K_0 c_t^2$. Tvrzení pak plyne z věty 2.2. □

Věta 14.2. V předchozí větě nahraďme předpoklad (14.2) silnějším

$$(\nabla M(x), C(x-\theta)) \leqq -K|x-\theta|^2, \quad \forall x, \tag{14.5}$$

a předpoklad (14.4) slabším

$$\sum_{t=1}^{\infty} a_t = +\infty, \quad a_t = o(c_t^2); \tag{14.6}$$

ostatní předpoklady beze změny.

$$\text{Potom} \quad \lim_{t\to\infty} \mathsf{E}|X_t - \theta|^2 = 0, \quad \text{pokud} \quad \mathsf{E}|X_1|^2 < +\infty.$$

Pro KW metodu lze s příslušnými změnami odvodit prakticky všechna tvrzení, která se dokazují pro RM metodu (včetně asymptotické normality a zákona iterovaného logaritmu). My si však všimneme jen jednoho problému, který je specifický pro KW metodu (resp. byl snadno řešitelný pro RM metodu), totiž optimální volby konstant a_t, c_t.

Za kritérium optimality zvolíme rychlost konvergence $X_t \to \theta$ podle kvadratického středu. Omezíme se přitom na konstanty tvaru

$$a_t = a/t^\alpha \,, \quad a > 0 \,, \quad 0 < \alpha \leqq 1 \,, \tag{14.7}$$
$$c_t = c/t^\gamma \quad c > 0 \,, \quad 0 < \gamma < \tfrac{1}{2}\alpha \,;$$

v případě $\alpha = 1$ budeme navíc požadovat, aby

$$2aK > 1 - 2\gamma \,, \tag{14.8}$$

kde K je konstanta z nerovnosti (14.5).

Následující věta je obměnou výsledku z práce [5].

Věta 14.3. Za předpokladů jako ve větě 14.2 a za předpokladu existence spojitých třetích parciálních derivací funkce M v okolí θ a při volbě konstant (14.7), (14.8) platí

$$\mathsf{E}|X_t - \theta|^2 = \begin{cases} O(t^{-4\gamma}) & \text{pro} \quad \gamma < \frac{1}{6}\alpha \,, \\ O(t^{-\alpha+2\gamma}) & \text{pro} \quad \gamma \geqq \frac{1}{6}\alpha \,. \end{cases}$$

Zejména tedy platí

$$\mathsf{E}|X_t - \theta|^2 = O(t^{-2/3}) \quad \text{pro} \quad a_t = a/t \,, \quad c_t = c/t^{1/6} \,, \quad 3aK > 1 \,,$$

a tato volba konstant je optimální v následujícím smyslu: je-li v (14.7) buďto $\alpha \neq 1$ nebo $\gamma \neq 1/6$, potom existuje KW posloupnost vyhovující všem předpokladům věty 14.3 a taková, že $t^{2/3}\, \mathsf{E}|X_t - \theta|^2 \to +\infty$.

15. RM METODA BEZ PŘEDPOKLADU NEZÁVISLOSTI POZOROVACÍCH CHYB

Předpoklad nezávislosti náhodných funkcí $e(t, x)$ v RM metodě je v mnoha situacích nerealistický a proto se studuje i chování stochastických aproximací bez tohoto předpokladu. Postupuje se obyčejně tak, že se funkce $e(t, x)$ považují nejprve za *nenáhodné*, a vyšetří se, jaké podmínky I je třeba na ně klást, aby odpovídající posloupnost aproximací konvergovala k žádanému řešení. Potom se hledají pravděpodobnostní podmínky II, které by zaručily, že skoro všechny realizace *náhodných* funkcí $e(t, x)$ splňují podmínky I a že tedy posloupnost stochastických aproximací konverguje s pravděpodobností 1.

Ljungova práce [19] obsahuje řadu obecných i detailních výsledků tohoto druhu; uveďme zde jeden z nich.

Nechť $\{e_t\}_{t=2}^{\infty}$ je posloupnost m-rozměrných vektorů (nenáhodných), nechť pro $t = 1, 2, \ldots$ je $Q(t; x, e)$ měřitelné zobrazení $R^n \times R^m \to R^n$. Nechť ke každému $x \in R^n$, $\varrho \in R^+$, $e \in R^m$ existuje $K = K(x, \varrho, e)$ tak, že

$$|Q(t; x_1, e) - Q(t, x_2, e)| \leqq K|x_1 - x_2|$$

pro všechna $t = 1, 2, \ldots$ a všechna x_1, x_2 z ϱ-okolí bodu x. Dále nechť

$$\lim_{t\to\infty} \frac{1}{t} \sum_{k=1}^{t} Q(k, x, e(k)) = M(x), \quad \text{pro všechna} \quad x \in R^n,$$

$$\lim_{t\to\infty} \frac{1}{t} \sum_{k=1}^{t} K(x, \varrho, e(k))$$

existuje a je konečná pro všechna $x \in R^n$ a dostatečně malá $\varrho > 0$.
Rovnice $M(x) = 0$ nechť má jediné řešení θ; $x(\tau) \equiv \theta$ nechť je globální asymptoticky stabilní řešení soustavy obyčejných diferenciálních rovnic $\mathrm{d}x/\mathrm{d}\tau = M(x)$, (tj. řešení této soustavy při libovolných počátečních podmínkách konvergují k θ pro $\tau \to \infty$).

Definujme posloupnost $\{X_t\}$ předpisem

$$(15.1) \qquad X_1 \ \text{libovolné}, \quad X_{t+1} = X_t + \frac{1}{t+1} Q(t, X_t, e(t+1)), \quad t \geqq 1.$$

Věta 15. 1. Za hořejších předpokladů platí buďto

$$\lim_{t\to\infty} |X_t| = +\infty \quad \text{nebo} \quad \lim_{t\to\infty} X_t = \theta.$$

Poznámka. Možnost $\lim\limits_{t\to\infty} |X_t| = +\infty$ lze vyloučit, platí-li: $\theta \in D_2 \subset D_1$, D_2 kompaktní, D_1 otevřená omezená, $U(x) \geqq 0$ dvakrát spojitě diferencovatelná,

$$\sup_{x\in D_1 - D_2} (\nabla U(x), M(x)) < 0, \quad \sup_{x\in D_2} U(x) < \inf_{x\in D_1{}^c} U(x)$$

a modifikujeme-li předpis (15.1) takto:

$$X_{t+1} = \left[X_t + \frac{1}{t+1} Q(t, X_t, e(t+1)) \right]_{D_1, D_2},$$

kde

$$[z]_{D_1, D_2} = \begin{cases} z, & \text{pro} \quad z \in D_1 \\ \text{některý bod v } D_2, & \text{pro} \quad z \in D_1^c. \end{cases}$$

Uveďme ještě příbuzný výsledek Györfiho [11]. Máme řešit soustavu rovnic $Ax = b$, kde matice $A \in R^{n\times n}$ je kladně definitní, $b \in R^n$; řešení označme θ $(= A^{-1}b)$. Matici A soustavy, ani vektor pravých stran b neznáme; máme však v t-tém kroku $(t = 2, 3, \ldots)$ iteračního postupu k dispozici matici $A(t) = A + e^A(t)$ a vektor $b(t) = b + e^b(t)$ tak, že

$$(15.2) \qquad \lim_{t\to\infty} \frac{1}{t} \sum_{k=2}^{t} e^A(k) = 0, \quad \lim_{t\to\infty} \frac{1}{t} \sum_{k=2}^{t} e^b(k) = 0$$

a tak, že

$$(15.3) \qquad \lim_{t\to\infty} \frac{1}{t} \sum_{k=2}^{t} \|A(k)\|^2$$

existuje a je konečná (kde $\|.\|$ značí euklidovskou normu).

Definujme posloupnost $\{X_t\}$ předpisem

$$X_1 \text{ libovolné}, \quad X_{t+1} = X_t - \frac{1}{t+1}\left(A(t+1)X_t - b(t+1)\right). \tag{15.4}$$

Věta 15.2. Za předpokladů (15.2) a (15.3) platí

$$\lim_{t\to\infty} X_t = \mathit{0}.$$

Důkaz jen naznačíme. Větu zřejmě stačí dokázat pro $b = 0 = \mathit{0}$. Opakovaným užitím vztahu (15.4) dostaneme vyjádření

$$X_t = \sum_{k=1}^{t} C(k, t)\, b(k), \tag{15.5}$$

kde

$$C(k, t) = \begin{cases} \frac{1}{k} I, & \text{pro } k = t \\ \frac{1}{k}\left(I - \frac{1}{k+1} A(k+1)\right) \dots \left(I - \frac{1}{t} A(t)\right), & \text{pro } 1 \leqq k < t \end{cases}$$

a kde klademe $b(1) = X_1$, Dále se naleznou horní odhady pro $\|C(k, t)\|$:

$$\|C(k,t)\| \leqq K \cdot \frac{1}{k}\left(\frac{k}{t}\right)^{\delta} \tag{15.6}$$

pro všechna dostatečně malá $\delta > 0$ a dostatečně veká $k, t,\ k \leqq t$. Vzorec (15.5) se přepíše na vyjádření X_t pomocí $\bar{b}(t) = (1/t) \sum_{k=1}^{t} b(k)$, řekněme

$$X_t = \sum_{k=1}^{t} C_1(k, t)\, \bar{b}(k);$$

odhad (15.6) se odpovídajícím způsobem upraví pro $\|C_1(k, t)\|$, čímž se ukáže, že $\lim_{t\to\infty} \|C_1(k, t)\| = 0$ pro každé pevné k, a že $\sup_t \sum_{k=1}^{t} \|C_1(k, t)\| < +\infty$. To umožňuje použít Toeplitzovu větu, podle níž $\bar{b}(t) \to 0$ implikuje $\sum_{k=1}^{t} C_1(k, t)\, \bar{b}(k) \to 0$, neboli $X_t \to 0$. □

Příklad situace, v níž je použitelná věta 15.2: Hledejme nejlepší lineární predikci $\eta^* = (\theta, \xi)$ náhodné veličiny η na základě n-rozměrného náhodného vektoru ξ, tj. řešme úlohu

$$\mathsf{E}(\eta - (x, \xi))^2 = \min_{x \in R^n}! \tag{15.7}$$

K dispozici mějme jedinou realizaci stacionární ergodické náhodné posloupnosti $\{\xi_t, \eta_t\}_{t=1}^{\infty}$ (jejíž stacionární rozdělení je rozdělení dvojice ξ, η). Řešení θ úlohy (15.7) je – jak známo – řešením soustavy normálních rovnic

$$\mathsf{E}(\xi\xi^{\mathrm{T}})\, x = \mathsf{E}(\eta\xi),$$

předpokládáme-li, že kovarianční matice vektoru ξ je kladně definitní a že $\mathsf{E}\eta^2 < +\infty$. Označíme-li

$$\mathsf{E}(\xi\xi^{\mathrm{T}}) = A\,, \qquad \mathsf{E}(\eta\xi) = b\,,$$
$$\xi_t\xi^{\mathrm{T}} = A(t)\,, \qquad \eta_t\xi_t = b(t)\,,$$

pak jsme v situaci popsané větou 15.2; podmínka (15.3) nyní zní

$$\text{(15.8)} \qquad \lim_{t\to\infty} \frac{1}{t} \sum_{k=1}^{t} |\xi_k|^4 < +\infty \quad \text{s pravděpodobností 1}$$

neboť $\|\xi_t\xi_t^{\mathrm{T}}\| = |\xi_t|^2$; tato podmínka je však vzhledem k předpokládané ergodičnosti rovnocenná s podmínkou $\mathsf{E}|\xi|^4 < +\infty$, tj. s podmínkou existence konečných momentů 4. řádu náhodného vektoru ξ. Posloupnost $\{X_t\}$ definovaná předpisem

$$\text{(15.9)} \qquad X_{t+1} = X_t = \frac{1}{t+1}\left(\xi_{t+1}\xi_{t+1}^{\mathrm{T}}X_t - \eta_{t+1}\xi_{t+1}\right)$$

konverguje pak k θ s pravděpodobností 1.

(15.9) lze psát ve tvaru

$$X_{t+1} = X_t - \frac{1}{t+1}\left\{(X_t, \xi_{t+1}) - \eta_{t+1}\right\}\xi_{t+1}$$

z něhož je vidět, že k výpočtu jeho pravé strany není třeba maticového násobení.

Poznámka. V [11] je věta 15.2 formulována obecněji – místo n-rozměrných vektorů se uvažují prvky reálného Hilbertova prostoru, místo matic omezené lineární operátory; konvergencí se rozumí konvergence v normě, resp. v normě operátorů.

16. STRUČNĚ O TOM, CO NEBYLO PROBRÁNO

V celém článku jsme se zabývali v podstatě jen Robbinsovou-Monroovou aproximační metodou, s výjimkou odst. 14, který byl věnován Kieferově-Wolfowitzově metodě. Ve specifice KW metody bylo možno pokračovat. Např. Fabian [9] navrhl založit odhad derivace $M'(x)$ na více než dvou pozorováních v každém kroku; při vhodném rozvržení experimentu a pro dostatečně hladkou funkci M lze pro KW metodu dosáhnout téměř téže rychlosti konvergence jako pro RM metodu.

Řada prací, např. Kushner-Sanvicente [17], byla věnována modifikaci KW metody pro nalezení maxima s omezeními; v zásadě lze převést na stochastický případ všechny metody používané v případě deterministickém: metodu penalizační funkce, metodu Lagrangeových multiplikátorů, metodu přípustných směrů.

Existují i stochastické aproximační metody, které nelze redukovat na RM nebo KW metodu:

Krasulina [15] navrhla stochastickou aproximační metodu pro nalezení maximálního vlastního čísla a k němu příslušného vlastního vektoru symetrické matice A (s vesměs jednoduchými vlastními čísly), máme-li k dispozici jen posloupnost $A(t) = A + e^A(t)$, kde $e^A(t)$ jsou nezávislé stejně rozdělené náhodné matice, $\mathsf{E}\, e^A(1) = 0$, $\mathsf{E}\|e^A(1)\|^2 < +\infty$.

Révész [24] popsal stochastickou aproximační metodu neparametrického odhadu regresní funkce. Předpokládá, že (X_t, Y_t), $1 \leqq t < +\infty$, jsou nezávislé stejně rozdělené 2-rozměrné náhodné vektory takové, že $0 \leqq X_1 \leqq 1$ a $\mathsf{E}(Y_1 \mid X_1 = x) = r(x)$. Odhady $r_t(x)$ neznámé regresní funkce $r(x)$ se konstruují podle předpisu

$$r_{t+1}(x) = r_t(x) + \frac{1}{(t+1)\, a_{t+1}}\, k\left(\frac{x - X_{t+1}}{a_{t+1}}\right)(Y_{t+1} - r_t(x))\,, \quad 0 \leqq x \leqq 1\,,$$

$$\text{kde} \quad a_t = t^{-\alpha}\,, \quad \tfrac{1}{2} < \alpha < 1\,, \quad k(x) = \begin{cases} 1 & \text{pro } |x| \leqq \frac{1}{2}\,, \\ 0 & \text{jinak}\,. \end{cases}$$

Za jistých předpokladů o hustotě rozdělení veličiny X a hladkosti funkce $r(x)$ se dokazuje konvergence $\sup |r_t(x) - r(x)| \to 0$ s pravděpodobností 1, její řád a asymptotická normalita. ^{x}

Existují i stochastické aproximační metody pro nalezení jiného význačného bodu funkce M než je nulový bod či bod maxima – např. pro nalezení inflexního bodu, [2]. Také funkce M nemusí být regresní funkcí, nýbrž např. kvantilovou funkcí, tj. $M(x)$ není střední hodnotou pozorování Y_x v bodě x, nýbrž α-kvantilem rozdělení veličiny Y_x, viz [12].

V celém článku jsme přijali ještě jedno omezení, totiž omezení na případ jediného kořene rovnice $M(x) = 0$, resp. jediného bodu maxima funkce M. Případ více kořenů (resp. případ víceextremální) lze vyšetřovat týmiž metodami; výsledky lze najít např. v [22], viz též následující poznámku a citaci.

Pokud jde o teoretické přístupy k vyšetřování stochastických aproximačních metod, věnovali jsme jen malou zmínku (v odst. 15) modernímu přístupu, využívajícímu kvalitativních vlastností řešení obyčejných diferenciálních rovnic. Kromě zmíněné již práce [19] byl tento přístup rozpracován zejména v monografii Kushner-Clark [16].

Konečně poznamenejme, že stochastickým aproximacím jsou blízce příbuzné tzv. rekursivní odhady parametrů, zejména rekursivní metoda nejmenších čtverců. Dělicí čára mezi oběma tématy není nijak ostrá, srv. [19], [22].

LITERATURA

[1] D. Anbar: On optimal estimation methods using stochastic approximation procedures. Ann. Statist. *1* (1973), 1175–1184.

[2] D. L. Burkholder: On class of stochastic approximation procedures. Ann. Math. Statist. *27* (1956), 1044–1059.

[3] J. L. Doob: Stochastic Processes. J. Wiley, New York 1953.

[4] V. Dupač: A dynamic stochastic approximation. Ann. Math. Statist. *36* (1965), 1695–1702.

[5] V. Dupač: O Kiefer-Wolfowitzově aproximační metodě. Časopis Pěst. Matem. *82* (1957), 47—75.
[6] V. Dupač: On the dynamic stochastic approximation. Banach Center Publications, vol. 6, 109—110. Warszawa 1980.
[7] V. Dupač, F. Král: Robbins-Monro procedure with both variables subject to experimental error. Ann. Math. Statist. *43* (1972), 1089—1095.
[8] V. Fabian: On asymptotic normality in stochastic approximation. Ann. Math. Statist. *39* (1968), 1327—1332.
[9] V. Fabian: Stochastic approximation of minima with improved asymptotic speed. Ann. Math. Statist. *38* (1967), 191—200.
[10] В. Ф. Гапошкин, Т.П. Красулина: О законе повторного логарифма в процессах стохастической аппроксимации. Теор. вероятн. и ее примен. *19* (1974), 879—886.
[11] L. Györfi: Stochastic approximation from ergodic sample for linear regression. Z. Wahrscheinlich. Verw. Geb. *54* (1980), 47—55.
[12] D. L. Hanson, R. P. Russo: A new stochastic approximation procedure using quantile curves. Z. Wahrscheinlich. Verw. Geb. (v tisku).
[13] K. L. Chung: On a stochastic approximation method. Ann. Math. Statist. *25* (1954), 463—483
[14] J. Komlós, P. Révész: A modification of the Robbins-Monro process. Stud. Sci. Math. Hung. *8* (1973), 329—340.
[15] Т. П. Красулина: Метод стохастической аппроксимации для определения наибольшего собственного числа математического ожидания случайных матриц. Автоматика и телемеханика 1970, 2, 50—56.
[16] H. J. Kushner, D. S. Clark: Stochastic Approximation Methods for Constrained and Unconstrained Systems. Springer-Verlag, New York 1978.
[17] H. J. Kushner, E. Sanvicente: Penalty function methods for constrained stochastic approximation. J. Math. Anal. and Applications *46* (1974), 499—512.
[18] T. L. Lai, H. Robbins: Adaptive design and stochastic approximation. Ann. Statist. *7* (1979), 1196—1221.
[19] L. Ljung: Analysis of recursive stochastic algorithms. IEEE Trans. Autom. Control *AC-22* (1977), 551—575.
[20] P. Major, P. Révész: A limit theorem for the Robbins-Monro approximation. Z. Wahrscheinlich. Verw. Geb. *27* (1973), 79—86.
[21] P. Mandl: Elements of stochastic analysis. Kybernetika *14* (1978), příloha.
[22] М. Б. Невельсон, Р. З. Хасьминский: Стохастическая аппроксимация и рекуррентное оценивание. Наука, Москва. 1972.
[23] G. Pflug: Stetige stochastische Approximation. Metrika *26* (1979), 139—150.
[24] P. Révész: How to apply the method of stochastic approximation in the nonparametric estimation of a regression function. Math. Operationsforsch. Statist., Ser. Statistics *8* (1977), 119—126.
[25] P. Révész: Robbins-Monro procedure in a Hilbert space and its application in the theory of learning processes I. Stud. Sci. Math. Hung. *8* (1973), 391—398.
[26] H. Robbins, S. Monro: A stochastic approximation method. Ann. Math. Statist. *22* (1951), 400—407.
[27] H. Robbins, D. Siegmund: A convergence theorem for non negative almost supermartingales and some applications. In: Optimizing Methods in Statistics (J. S. Rustagi, ed.). Academic Press, New York 1971, 233—257.
[28] W. Stout: A martingale analogue of Kolmogorov's law of the iterated logarithm. Z. Wahrscheinlich. verw. Geb. *15* (1970), 279—290.
[29] J. H. Venter: An extension of the Robbins-Monro procedure. Ann. Math. Statist. *38* (1967), 181—190.