1 Wykład 10. Rozszerzenia i uogólnienia

Widzimy zatem, że ograniczoność ciągu aproksymacji przy luźnych wymaganiach dotyczących odwzorowania $\mathbf{f}$ (założenia (??) i (??)) wymaga już pewnej uporządkowanej struktury probabilistycznej zakłóceń $\{\boldsymbol{\xi}_n\}$ (założenie różnic martyngałowych). Można oczekiwać, że w bardziej skomplikowanych przypadkach procedur aproksymacji stochastycznej założenie to będzie także aktywne.

Uwaga 1 *Zauważmy także, że założeń (??) nie trzeba było używać do zagwarantowania zbieżności prawie na pewno, gdy wiadomo było, że odwzorowanie* $\mathbf{f}$ *spełniało warunek (??).*

Istnieje jednak sposób ominięcia tych wzmożonych wymagań dotyczących zakłóceń, a mających na celu uzyskanie ograniczoności procedur aproksymacji stochastycznej. Mianowicie, jeśli znalibyśmy zbiór V, w którym na pewno leży nieznany parametr $\boldsymbol{\theta}$, to można by rozważać procedurę następującą:

jeśli w $n-$tym kroku iteracyjnym wielkość

$$\mathbf{p}_n = \mathbf{x}_n - \mu_n \left(\mathbf{f}(\mathbf{x}_n) + \boldsymbol{\xi}_{n+1}(\mathbf{x}_n)\right)$$

leży wewnątrz zbioru V, to za $\mathbf{x}_{n+1}$ przyjmujemy $\mathbf{p}_n$, jeśli zaś $\mathbf{p}_n \notin V$ to za $\mathbf{x}_{n+1}$ przyjmujemy jakiś punkt zbioru V.

Pozostaje wybrać ten punkt. Mamy dużą dowolność i dobrze byłoby wybrać ten punkt właściwie. Okazuje się, że jeśli zbiór V jest *ograniczony, domknięty i wypukły* i jeśli za $\mathbf{x}_{n+1}$ przyjmiemy rzut (prostokątny) $\boldsymbol{\pi}_n$ punktu $\mathbf{p}_n = \mathbf{x}_n - \mu_n \left(\mathbf{f}(\mathbf{x}_n) + \boldsymbol{\xi}_{n+1}(\mathbf{x}_n)\right)$ na V, to dla dowolnego punktu $\mathbf{a}$ $\left|\mathbf{p}_n - \mathbf{a}\right| \geq \left|\boldsymbol{\pi}_n - \mathbf{a}\right|$. Innymi słowy, jeśli zamiast procedury (??) rozważać procedurę:

$$\mathbf{x}_{n+1} = \begin{cases} \mathbf{p}_n = \mathbf{x}_n - \mu_n \left(\mathbf{f}(\mathbf{x}_n) + \boldsymbol{\xi}_{n+1}(\mathbf{x}_n)\right), & \textit{gdy} \quad \mathbf{p}_n \in V \\ \boldsymbol{\pi}_n = \text{rzut } \mathbf{p}_n \,\text{na}\, V, & \textit{gdy} \quad \mathbf{p}_n \notin V \end{cases}, \tag{1}$$

to możemy do analizy zbieżności tych procedur stosować zwykłe, rozważane wcześniej, oszacowania. Fakt ten wynika z następujących lematów.

Założenie, że nieznana wartość $\boldsymbol{\theta}$ leży wewnątrz pewnego zbioru znanego i, że w związku z tym tam właśnie należy jej szukać jest równoważne założeniu, że istnieją ograniczenia na położenie punktu $\boldsymbol{\theta}$. Prowadzi to nas do procedur aproksymacji stochastycznej z ograniczniami. Jak się okazuje ta klasa procedur była i jest szeroko znana i badana. Szczególnie tzw. wersje Kiefera -Wolfowitza tych procedur okazały się ważne i doprowadziły do powstania nowego działu metod numerycznych a mianowicie optymalizacji stochastycznej.

2 Bardziej złożone procedury

Zauważmy, że dotychcząs zakłócenia obserwacji pochodziły jakby z innego źródła niż wartości estymatorów, czyli punkty $\{\mathbf{x}_n\}$. Nie można więc dotychczasowymi metodami badać omawianej poprzednio procedury wyliczającej zadany kwantyl nieznanego rozkładu. Przypomnijmy ten przykład : $\{\xi_n\}_{n\geq 1}$ był ciągiem niezależnych zmiennych losowych o rozkładzie *normalnym* (ogólnie o dystrybuancie F). Obserwacje w punkcie x były dane wzorem: $Z_i(x) = I(\xi_i < x) - .85$ (ogólnie $Z_i(x) = I(\xi_i < x) - \alpha$). Zauważmy, że mamy tu $EZ_i(x) = F(x) - .85$ (ogólnie $EZ_i(x) = F(x) - \alpha$), a nawet zachodzi silniejsza własność: $E\left(Z_i(x)|\xi_{i-1}, \ldots, \xi_1\right) = F(x) - .85$. Zapiszmy więc

$$Z_i(x) = F(x) - \alpha + \zeta_i(x),$$

gdzie ciąg zmiennych losowych $\{\zeta_i(x)\}_{n\geq 1}$ ma następującą własność $E\zeta_i(x) = 0$ a nawet ogólniej :

$$E\left(\zeta_i(x)|\xi_{i-1}, \ldots, \xi_1\right) = 0 \ p.n.$$

Przypomnijmy używaną procedurę :

$$x_i = x_{i-1} - \frac{1}{i} Z_i(x_{i-1});\ x_0 = xo;\ i \geq 1.$$

Mamy tu

$$E\left(Z_i(x_{i-1}) - F(x_{i-1}) + \alpha | x_{i-1}, \ldots, x_1\right) = 0$$

i

$$E\left(Z_i(x_{i-1}) - F(x_{i-1}) + \alpha | \xi_{i-1}, \ldots, \xi_1\right) = 0,$$

gdyż oczywiście $\sigma\left(x_i, \ldots, x_1\right) \subset \sigma\left(\xi_i \ldots, \xi_1\right),\ i \geq 1$. Jest to własność *różnicy martyngałowej* względem filtracji $\{\sigma\left(\xi_1, \ldots, \xi_n\right)\}_{n\geq 1}$.

W dalszym ciągu będziemy zakładali, że ciąg normalny $\{\mu_n\}_{n\geq 0}$ spełnia dodatkowo warunek:

$$\sum_{n\geq 1} \mu_n^2 < \infty. \tag{2}$$

Uwaga 2 *Zauważmy, że zamiast jednej funkcji* $\mathbf{f}$, *której zera szukamy, można używać ciągu funkcji* $\{\mathbf{f}_n\}_{n\geq 1}$ *takiego jednak, aby np.*

$$\exists \boldsymbol{\theta} \in \mathbb{R}^m, \exists \{\delta_n\}_{n\geq 1}\ \forall \mathbf{x} \in \mathbb{R}^m : (\mathbf{x} - \boldsymbol{\theta})' \, \mathbf{f}_n(\mathbf{x}) > \boldsymbol{\delta}_n \left|\mathbf{x} - \boldsymbol{\theta}\right|^2 ; \liminf_{n\to\infty} \delta_n > 0 \tag{3}$$

i

$$\exists \{\kappa_{1n}\}_{n\geq 1}, \{\kappa_{2n}\}_{n\geq 1}, \forall \mathbf{x} \in \mathbb{R}^m : |\mathbf{f}_n(\mathbf{x})| \leq \kappa_{1n} \left|\mathbf{x} - \boldsymbol{\theta}\right| + \kappa_{2n}, \tag{4}$$

$$\limsup_{n\to\infty} \left(\kappa_{1n} + \kappa_{2n}\right) < \infty. \tag{5}$$

Rozważmy procedurę:

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \mu_n \mathbf{F}_n(\mathbf{x}_n, \xi_n),\ n \geq 1. \tag{6}$$

Oznaczmy

$$\mathbf{G}_n(\mathbf{x}) = E\left(\mathbf{F}_n(\mathbf{x}, \xi_n) | \xi_{n-1}, \ldots, \xi_0\right).$$

Mamy twierdzenie:

Twierdzenie 3 *Załóżmy, że funkcje* $\{\mathbf{G}_n(\mathbf{x})\}_{n\geq 1}$ *spełniają warunki (3) i (4) w pewnym punkcie* θ*. Załóżmy także, że szumy*

$$\boldsymbol{\zeta}_n(\mathbf{x}) = \mathbf{F}_n(\mathbf{x}, \boldsymbol{\xi}_n) - \mathbf{G}_n(\mathbf{x}), n = 1, 2, \dots ,$$

spełniają warunek:

$$E\left(|\boldsymbol{\zeta}_n(\mathbf{x})|^2 \,|\boldsymbol{\xi}_{n-1}, \dots, \xi_1\right) \leq L_n(1 + |\mathbf{x}|^2), p.n. \sup L_n < 0 \; p.n. \tag{7}$$

Wówczas przy założeniu, że ciąg normalny $\{\mu_i\}_{i\geq 0}$ *spełnia założenia (2), procedura (6) zbiega prawie na pewno do* θ*.*

Procedury tego typu tzn. z funkcjami $\mathbf{f}_n$ zależnymi od numeru iteracji, a także być może z zaburzeniami zależnymi od numeru iteracji i dotychczas znalezionego estymatora, wykorzystuje się w tzw. identyfikacji dyskretnych procesów stochastycznych. Przedstawimy tu twierdzenie o zbieżności odpowiedniej procedury będące uogólnieniem Twierdzenia 3 i użyteczne właśnie do celów identyfikacji, a także w zagadnienieach tzw. optymalizacji stochastycznej, omówionych pokrótce niżej.

Aby to zrobić szybko, udowodnimy kilka użytecznych lematów liczbowych.

Lemat 4 *Załóżmy, że ciąg liczbowy* $\{d_n\}_{n\geq 1}$*spełnia zależność rekurencyjną:*

$$d_{n+1}^2 \leq \left[1 - 2\delta_n\mu_n + \mu_n^2\gamma_n\right]^+ d_n^2 + \mu_n g_n d_n + \mu_n h_n, \tag{8}$$

gdzie $\{g_n\}_{n\geq 1}$, $\{h_n\}_{n\geq 1}$, $\{\delta_n\}_{n\geq 1}$ *i* $\{\gamma_n\}_{n\geq 1}$ *są pewnymi ciągami liczbowymi, spełniającymi następujące założenia:*

$$\liminf_{n\to\infty} \delta_n > 0; \lim_{n\to\infty} \mu_n\gamma_n = 0.$$

Wówczas spełniona jest zależność rekurencyjna:

$$d_{n+1} \leq \lambda_n \max(d_n, \epsilon_n),$$

gdzie $\lambda_n = \sqrt{1 - \mu_n\delta_n + \mu_n^2\gamma_n}$*, zaś ciąg* $\{\epsilon_n\}_{n\geq 1}$ *jest ciągiem dodatnich pierwiastków równań:*

$$\epsilon_n^2\delta_n - \epsilon_n g_n - h_n = 0; n \geq 1;$$

w szczególności, jeśli $g_n \underset{n\to\infty}{\longrightarrow} 0$, $h_n \underset{n\to\infty}{\longrightarrow} 0$, *to* $d_n \underset{n\to\infty}{\longrightarrow} 0$.

Dowód. Był już kilka razy (bez takiego sformułowania) przedstawiany przy okazji dowodów twierdzeń o zbieżności procedur aproksymacji stochastycznej. ■

Lemat 5 *Załóżmy, że ciąg liczbowy* $\{d_n\}_{n\geq 1}$ *spełnia zależność rekurencyjną:*

$$d_{n+1}^2 \leq \left[1 - 2\delta_n\mu_n + \mu_n^2\gamma_n\right]^+ d_n^2 + \mu_n g_n d_n + \mu_n h_n, \tag{9}$$

gdzie $\{g_n\}_{n\geq 1}$, $\{h_n\}_{n\geq 1}$, $\{\delta_n\}_{n\geq 1}$ *i* $\{\gamma_n\}_{n\geq 1}$ *są pewnymi ciągami liczbowymi, spełniającymi następujące założenia:*

$$\delta_n = \delta_n^{'} + \delta_n^{''}; \liminf_{n\to\infty} \delta_n^{'} > 0; \lim_{n\to\infty} \mu_n \delta_n^{'} = 0, \tag{10}$$

$$\infty > \sup_{m>n} \left| \sum_{i=n}^{m} \mu_i \delta_i^{''} \right|; \lim_{n\to\infty} \mu_n \delta_n^{''} = 0; \tag{11}$$

$$0 = \lim_{n\to\infty} \mu_n \gamma_n. \tag{12}$$

Wówczas spełniona jest inna zależność rekurencyjna:

$$d_{n+1} \leq \lambda_n \max(d_n, \epsilon_n),$$

gdzie $\lambda_n = \sqrt{\frac{M_n}{M_{n+1}}\left(1 - \mu_n \delta_n^{'} \exp(2\mu_n \delta_n^{''}) + \mu_n^2 \gamma_n \exp(2\mu_n \delta_n^{''})\right)}$, $\epsilon_n = \tau_n / M_n$, τ_n *jest dodatnim pierwiastkiem równania:*

$$\tau_n^2 \delta_n^{'} \exp(2\mu_n \delta_n^{''}) - \tau_n g_n \exp(\mu_n \delta_n^{''}) \sqrt{M_{n+1}} - h_n M_{n+1} = 0,$$

zaś

$$M_n = \limsup_{m\to\infty} \exp\left(-2 \sum_{i=n}^{m} \mu_i \delta_i^{''}\right).$$

Twierdzenie 6 *Załóżmy, że funkcje* $\{\mathbf{G}_n(\mathbf{x})\}_{n\geq 1}$ *spełniają warunki (10) i (4) w pewnym punkcie* θ*. Załóżmy także, że szumy* $\boldsymbol{\zeta}_n(\mathbf{x}) = \mathbf{F}_n(\mathbf{x}, \boldsymbol{\xi}_n) - \mathbf{G}_n(\mathbf{x})$ *spełniają warunek (7) Wówczas przy założeniu, że ciąg normalny* $\{\mu_i\}_{i\geq 0}$ *spełnia założenia (2), procedura (6) zbiega prawie na pewno do* θ.

Dowód. Jest analogiczny do dowodu Twierdzenia 3, z tym, że wykorzystuje się Lemat 5. ■

Uwaga 7 *Zauważmy, że w analogiczny sposób można udowodnić inne, podobne twierdzenia dotyczące zbieżności, kombinując różne założenia dotyczące postaci funkcji* $\{\mathbf{F}_n\}_{n\geq 1}$ *i zakłóceń. W szczególności zamiast Twierdzenia 6, można rozważać twierdzenie analogiczne do Twierdzeń o zbieżności z poprzedniego wykladu i* **??**.

2.1 Wstęp do optymalizacji stochastycznej

Dotychczas omawiane procedury poszukiwania zer układu funkcji nazywane są procedurami *Robbinsa-Monro.* Procedury poszukiwania ekstremów układów funkcji w warunkach losowych nazywane są procedurami *Kiefera-Wolfowitza,* gdyż po raz pierwszy omawiane były w pracy Kiefera i Wolfowitza . Zajmijmy się przez chwilę jednowymiarowymi wersjami tego typu procedur.

Załóżmy, że dana jest funkcja $\psi : \mathbb{R} \to \mathbb{R}$, której minimum znajduje się w punkcie θ. Punkt ten chcielibyśmy znaleźć, ale nie możemy obserwować wartości funkcji ψ. Weźmy pod uwagę zbieżny do zera ciąg $\{c_n\}_{n\geq 1}$ liczb dodatnich i załóżmy, że funkcja ψ jest funkcją *różniczkowalną,* o pochodnej spełniającej globalny *warunek Lipschitza* . Zauważmy, że wartości funkcji

$$\hat{\psi}_n(x) = \frac{\psi(x+c_n) - \psi(x-c_n)}{2c_n}$$

zbiegają do $\psi'(x)$ w każdym punkcie $x \in \mathbb{R}$. Zakładamy, że wartości funkcji $\hat{\psi}$ nie mogą być bezpośrednio obserwowane, a jedynie można obserwować wartości funkcji

$$\Psi_n(x) = \psi(x) + \xi_n,$$

gdzie $\{\xi_n\}$ jest ciągiem zmiennych losowych o średniej zero i skończonych wariancjach. Punkt θ będziemy estymowali przy pomocy ciągu:

$$x_{n+1} = x_n - \mu_n \left(\frac{\Psi_{2n+1}(x_n + c_n) - \Psi_{2n}(x_n - c_n)}{2c_n} \right). \tag{13}$$

Załóżmy, że szereg

$$\sum_{n=1}^{\infty} \frac{\mu_n(\xi_{2n+1} - \xi_{2n})}{c_n} \tag{14}$$

zbiega z prawdopodobieństwem 1. Jeśli np. w najprostszej sytuacji zmienne losowe $\{\xi_n\}$ są *różnicami martyngałowymi* o wspólnie ograniczonych wariancjach, to wystarczy założyć że np.

$$\sum_{n=1}^{\infty} \left(\frac{\mu_n}{c_n} \right)^2 < \infty. \tag{15}$$

Rozwińmy funkcję $\psi(x + c_n)$ wokół dowolnego punktu x. Mamy

$$\psi(x + c_n) = \psi(x) + c_n \psi^{'}(x) + r_n(x),$$

gdzie $r_n(x)$ jest resztą spełniającą warunek $|r_n(x)| \leq c_n^2 L$, gdzie L jest globalną stałą Lipschitza, której istnienie postulowaliśmy. Zatem

$$\hat{\psi}_n(x) = \frac{\psi(x+c_n) - \psi(x-c_n)}{2c_n} = \psi^{'}(x) + R_n(x),$$

gdzie $|R_n(x)| \leq c_n L$.

Zauważmy, że z poprzednio omawianych twierdzeń wynika, że aby procedura (13) była zbieżna potrzeba, aby:

$$\forall \varepsilon > 0 : 1/\varepsilon \geq |x - \theta| \geq \varepsilon \Rightarrow \exists \delta(\varepsilon) : (x-\theta)\hat{\psi}_n(x) \geq \delta_n(\varepsilon)(x-\theta)^2, \tag{16}$$

$$\exists \kappa_1, \kappa_2, \forall n \geq 1, \left| \hat{\psi}_n(x) \right| \leq \kappa_{1n} |x - \theta| + \kappa_{2n}. \tag{17}$$

Zauważmy jednak, że w pierwszym przypadku mamy dla $1/\varepsilon \geq |x-\theta| \geq \varepsilon$:

$$\begin{aligned}(x-\theta)\hat{\psi}_n(x) &= (x-\theta)\psi^{'}(x) + (x-\theta)R_n(x) \geq \\ & (x-\theta)\psi^{'}(x) - |x-\theta|^2 c_n L/\varepsilon.\end{aligned}$$

Jeśli zatem gradient funkcji ψ spełnia warunek:

$$\forall\varepsilon > 0 : 1/\varepsilon \geq |x-\theta| > \varepsilon \Rightarrow \exists\delta^{'}(\varepsilon) : (x-\theta)\psi^{'}(x) \geq \delta^{'}(\varepsilon)(x-\theta)^2,$$

to warunek 16 jest spełniony ze stałą $\delta_n(\varepsilon) = \delta^{'}(\varepsilon) + \delta_n^{''}(\varepsilon)$, gdzie $\delta_n^{''}(\varepsilon) = -c_n L/\varepsilon$. Podobnie warunek istnienia globalnej stałej Lipschitza implikuje spełnienie warunku 17. A zatem postępując w standardowy sposób, jak w dowodach poprzednich dwu twierdzeń o zbieżności procedur aproksymacji stochastycznej, dochodzimy do oszacowania prawdziwego dla $1/\varepsilon \geq |x_n-\theta| \geq \varepsilon$:

$$d_{n+1}^2 \leq [1 - 2\mu_n\delta_n(\varepsilon) + \mu_n^2\kappa_{1n}]^+ d_n^2 + \mu_n d_n g_n(\varepsilon) + \mu_n h_n(\varepsilon), \tag{18}$$

gdzie $d_n = |x_n - \theta - G_n|$, $G_n = \sum_{i\geq n}\mu_i\frac{\xi_{2i+1}-\xi_{2i}}{c_i}$, zaś ciągi $\{g_n(\varepsilon)\}$ i $\{h_n(\varepsilon)\}$ zależą od ciągów $\{\mu_n\}$, $\{\delta_n(\varepsilon)\}$, $\{\kappa_{1n}\}$, $\{\kappa_{2n}\}$ i mają własność: $\forall\varepsilon > 0$: $g_n(\varepsilon) \underset{n\to\infty}{\longrightarrow} 0$; $h_n(\varepsilon) \underset{n\to\infty}{\longrightarrow} 0$ z prawdopodobieństwem 1. Można zatem rozumować jak w dowodzie Twierdzenia 6 (z ciągiem δ_n zależnym od ε), stosując Lemat 5, i dostać zbieżność ciągu $\{d_n\}$ do zera z prawdopodobieństwem 1. Aby jednak ten lemat można było zastosować, trzeba założyć, że warunek (11) (tj. $\sup_n \left|\sum_{i=0}^{n-1}\mu_i\delta_{i+1}^{''}\right| < \infty$ *p.n.* i $\mu_n\delta_{n+1} \to 0,\ n\to\infty$) jest spełniony. Pamiętając postać ciągu $\left\{\delta_n^{''}(\varepsilon)\right\}_{n\geq 1}$ nietrudno zauważyć, że warunek ten będzie spełniony wówczas, gdy

$$\sum_{n\geq 1}\mu_n c_n < \infty. \tag{19}$$

Naszkicowaliśmy zatem dowód twierdzenia:

Twierdzenie 8 . *Niech ciągi liczbowe $\{\mu_n\}_{n\geq 0}$ i $\{c_n\}_{n\geq 0}$ będą tak wybrane, aby spełnione były warunki (2, 15, 19). Załóżmy także, że zakłócenia $\{\xi_n\}$ są takie, że warunek (15) gwarantuje zbieżność szeregu (14). Załóżmy, że funkcja $\psi : \mathbb{R}\to\mathbb{R}$ jest różniczkowalna w każdym punkcie i jej pochodna spełnia globalny warunek Lipschitza, a także warunki (16, 17) z wybranym ciągiem $\{c_n\}$. Wówczas procedura (13) zbiega z prawdopodobieństwem 1 do θ.*

Uwaga 9 *Warunki (15) i (19) które mają spełniać ciągi współczynników $\{\mu_i\}_{i\geq 0}$ i $\{c_i\}_{i\geq 0}$ są znane i pojawiają się już we wspomnianej pracy Kiefera i Wolfowitza. Dowód Twierdzenia 8 jest oczywiście inny. Ciekawy, jak się wydaje jest fakt, że to klasyczne twierdzenie udało się dostać jako szczególny przypadek zastosowania Twierdzenia 6.*

Uwaga 10 *Należy w tym miejscu jeszcze raz przypomnieć, że procedura (13) była inspiracją dla wielu innych autorów do znajdowania rozszerzeń, uogólnień i udoskonaleń klasycznej procedury. W sumie procedura ta była zarodkiem, wokół którego powstał nowy dział metod numerycznych mianowicie optymalizacja stochastyczna. Istnieje olbrzymia literatura poświęcona tej dziedzinie.*

2.2 Szybkość zbieżności

Zauważmy, że dotychczas poznane twierdzenia pozwalają badać szybkość zbieżności w zachodzeniu *praw wielkich liczb.* Rozważmy mianowicie rekurencyjną postać PWL, tzn.

$$\overline{X}_{n+1} = (1-\mu_n)\overline{X}_n + \mu_n X_{n+1}, \tag{20}$$

gdzie $\{X_i\}_{i\geq 1}$ jest ciągiem zmiennych losowych o zerowej wartości oczekiwanej. Wiadomo, że jeśli ciąg $\{\mu_i\}_{i\geq 0}$ spełnia warunki:

$$\mu_0 = 0,\ \mu_n \in (0,1),\ \ \sum_{i\geq 0} \mu_i = \infty, \tag{21}$$

to ciąg $\left\{\overline{X}_n\right\}_{n\geq 1}$ można wyrazić formułą:

$$\overline{X}_n = \frac{\sum_{i=0}^{n-1} \alpha_i X_{i+1}}{\sum_{i=0}^{n-1} \alpha_i}, \tag{22}$$

gdzie $\alpha_0 = 1$, $\alpha_n = \mu_n / \prod_{i=1}^{n}(1-\mu_i)$, $n \geq 1$. Niech teraz $\{\beta_n\}_{n\geq 1}$ będzie pewnym ściśle rosnącym ciągiem liczbowym. Zauważmy, że jeśli oznaczymy : $Z_n = \beta_n \overline{X_n}$, to dostaniemy zależność rekurencyjną:

$$Z_{n+1} = (1+\gamma_n)(1-\mu_n)Z_n + \beta_n \mu_n X_{n+1}, \tag{23}$$

gdzie oznaczyliśmy dla symetrii wzorów $\gamma_n = (\beta_{n+1} - \beta_n)/\beta_n$, której zbieżność można badać znanymi metodami.

Posługując się podobną techniką można badać szybkość zbieżności procedur aproksymacji stochastycznej.

Więcej na temat szybkości zbieżności procedur aproksymacji stochastycznej można znaleźć np. w pracach: Fabiana i Rupperta. W szczególności można tam znaleźć warunki na funkcje $\mathbf{f}$, przy których dla każdego $\gamma < 1/2$ zbiega do zera ciąg $\{n^{\gamma}(\mathbf{x}_n - \boldsymbol{\theta})\}_{n\geq 1}$.

Istnieją także prace poświęcone zatrzymywaniu procedur aproksymacji stochastycznej, czyli mówiąc w skrócie następującemu zagadnieniu. Dla ustalenia uwagi rozważmy procedurę (6). Należy znaleźć taki moment zatrzymania τ, by z prawdopodobieństwem nie mniejszym niż $\delta > 0$ spełniony był warunek

$$|\mathbf{x}_\tau - \theta| < \varepsilon,$$

gdzie ε i δ to zadane z góry liczby. Niestety nie udało się do tej pory znaleźć zadawalającej reguły zatrzymania τ. Podejmowano jednak szereg prób. Opis ich można znaleźć np. w Farrel, Sielkien lub Yin.

Jeśli chodzi o związki procedur aproksymacji stochastycznej z *Centralnym Twierdzeniem Granicznym,* to mamy następujący wynik szczegółowy.

Rozważmy procedurę 1-wymiarowej aproksymacji stochastycznej :

$$X_{n+1} = X_n - \frac{a}{n+1}\left(f(X_n) + \xi_{n+1}\right), \tag{24}$$

z funkcją f spełniającą następujące warunki:

$$\begin{aligned}
\forall\epsilon &> 0: \sup_{|x-\theta|>\epsilon} f(x)(x-\theta) > 0, && (25)\\
\exists\kappa_1,\kappa_2 &: |f(x)| \leq \kappa_1|x-\theta| + \kappa_2, && (26)\\
\exists B &> 0: f(x) = B(x-\theta) + \delta(x), && (27)\\
\delta(x) &= o(x-\theta), \text{ gdy } x \to \theta.
\end{aligned}$$

Twierdzenie 11 *Niech będzie dana procedura aproksymacji stochastycznej (24) z funkcją f spełniającą warunki (25, 26, 27). Załóżmy dodatkowo, że $\{\xi_i\}_{i\geq 0}$ jest ciągiem niezależnych zmiennych losowych o zerowych wartościach oczekiwanych i wariancjach równych σ^2.*
Jeśli

$$aB > \frac{1}{2},$$

to ciąg zmiennych losowych

$$\sqrt{n}(X_n - \theta) \tag{28}$$

ma rozkład asymptotycznie normalny

$$N(0, \frac{a^2\sigma^2}{2aB-1}).$$

Dowód. Można znaleźć w monografi Chasminskiego i Newelsona. Nie jest on prosty i elementarny dlatego też nie przytaczamy go tu. ■