ROCZNIKI POLSKIEGO TOWARZYSTWA MATEMATYCZNEGO
Seria III: MATEMATYKA STOSOWANA VII (1976)

Recenzje

S. N. Abdelhamid i D. Anbar „Określenie optymalnej funkcji obserwacji w jednowymiarowych procedurach aproksymacji stochastycznej"*

W omawianych pracach zbadane zostały asymptotyczne własności jednowymiarowych procedur aproksymacji stochastycznej (AS), opisywanych podanym niżej — ogólnym — równaniem (1). Rozważania Abdelhamida obejmują przy tym zarówno zadanie Robbinsa–Monro (RM), jak i Kiefera–Wolfowitza (KW)(1); rozważania Anbara dotyczą tylko zadania RM.

Podstawowy rezultat obydwóch prac polega na sformułowaniu i rozwiązaniu problemu optymalnego wyboru funkcji obserwacji (funkcji h w (1)).

Niech

$$X_{n+1} = X_n - a_n c_n^{-1} h(Y_n), \qquad n = 1, 2, \ldots, \tag{1}$$

gdzie a_n, c_n są liczbami dodatnimi, X_1 jest dowolną zmienną losową, h jest funkcją borelowską: $\boldsymbol{R} \to \boldsymbol{R}$, Y_n zaś jest zmienną losową związaną z obserwacjami wartości funkcji f, dokonywanymi na n-tym etapie procedury (Y_n nazywana jest krótko *obserwacją losową*, a h — konsekwentnie — *funkcją obserwacji*).

W szczególności, (1) opisuje klasyczne procedury RM i KW, jeżeli za h przyjąć funkcję identycznościową, $h = I$ (por. (7) i odpowiednio (16) lub początkowe założenia Tw. 4). W przypadku $h = \operatorname{sign}$, (1) opisuje procedury Fabiana (F), [3].

Pierwsze części omawianych prac poświęcone są ustaleniu klasy funkcji h, dla których: (I) ciąg $\{X_n\}$ jest zbieżny z prawdopodobieństwem 1 do rozwiązania Θ oraz (II) zmienne losowe $n^{\beta/2}(X_n - \Theta)$ mają rozkład asymptotycznie normalny (przy stosownie obranym, ustalonym, β). Z kolei, w takiej klasie określa się funkcję obserwacji optymalną, tj. minimalizującą drugi moment asymptotycznego rozkładu zmiennych $n^{\beta/2}(X_n - \Theta)$.

Jest przy tym godne uwagi, że pomimo prostoty schematu iteracyjnego (1), X_n okazuje się asymptotycznie efektywnym estymatorem parametru Θ. Ostatnie stwierdzenie obowiązuje w tym sensie, że — przy odpowiednich założeniach — wa-

---

* S. N. Abdelhamid, *Transformation of observations in stochastic approximation*, Ann. Statist. 1.6 (1973); D. Anbar, *On optimal estimation methods using stochastic approximation procedures*, ibidem.

(1) Niech f: $\boldsymbol{R} \to \boldsymbol{R}$ będzie funkcją borelowską zmiennej rzeczywistej. Zadanie RM polega na iteracyjnym określeniu pierwiastka równania $f(\Theta) = 0$ w sytuacji, gdy dokładnej postaci funkcji f nie znamy, dla każdego zaś $x \in \boldsymbol{R}$ obserwować możemy nieobciążoną ocenę wartości $f(x)$. Odpowiednio, przez zadanie KW, realizowane w podobnej sytuacji, rozumie się iteracyjne określenie punktu Θ, w którym f osiąga swoje minimum (lub maksimum).

riancja rozkładu asymptotycznego odpowiada dolnej granicy nierówności Craméra-Rao dla estymatora nieobciążonego.

Dalej przyjmuje się, że wszystkie zmienne losowe określone są na przestrzeni probabilistycznej $(\Omega, \mathscr{F}, P)$ i wprowadza następujące oznaczenia: $\mathscr{X}_n = [X_1, \ldots, X_n]$, $M_n(X_n) = E_{\mathscr{X}_n}[Y_n]$. Relacje między zmiennymi losowymi (w tym zbieżność) rozumie się jako zachodzące z prawdopodobieństwem 1.

1. Zbieżność procedur typu (1).

TWIERDZENIE 1 (Abdelhamid). *Niech dana będzie procedura* (1). *Niech* $\Theta \in \boldsymbol{R}$, τ_n *będzie ciągiem liczb nieujemnych,* $\tilde{M}_n(X_n) = E_{\mathscr{X}_n}[h(Y_n)]$ *i* $\tilde{M}_n$ *— funkcje borelowskie. Niech, dalej, spełnione będą warunki:*

(2) *jeżeli* $\varepsilon > 0$, $|x-\Theta| > \varepsilon$ *i* $n > n_0(\varepsilon)$, *to*

$$(x-\Theta)\tilde{M}_n(x) > 0;$$

(3) *jeżeli* $0 < \delta_1 < \delta_2 < \infty$, *to*

$$\sum_{n=1}^{\infty} [a_n c_n^{-1} \inf_{\delta_1 \leqslant |x-\Theta| \leqslant \delta_2} |\tilde{M}_n(x)|] = \infty;$$

(4) *istnieją stałe dodatnie* H_0 *i* H_1 *takie, że*

$$|\tilde{M}_n(x)| \leqslant H_0 + H_1 \tau_n |x-\Theta| \quad \text{dla wszystkich } x \in \boldsymbol{R} \text{ i } n = 1, 2, \ldots;$$

(5a) *jeżeli* τ_n *jest ciągiem nieograniczonym, to*

$$\lim_{n\to\infty} a_n c_n^{-1} \tau_n = 0,$$

w przypadku przeciwnym

(5b)
$$\lim_{n\to\infty} a_n c_n^{-1} = 0;$$

(6)
$$\sum_{n=1}^{\infty} a_n^2 c_n^{-2} E_{\mathscr{X}_n}[h(Y_n) - \tilde{M}_n(X_n)]^2 < \infty.$$

Wtedy $X_n \to \Theta$. ■

Twierdzenie 1 jest naturalnym uogólnieniem znanego twierdzenia Burkholdera, sformułowanego przy nieco mocniejszych założeniach i dla $h = I$.

W dalszym ciągu obowiązywać będą następujące założenia:

Założenie 1. *Dla procedury* (1) *przyjmuje się*

(7)
$$Y_n = M_n(X_n) + V_n,$$

gdzie V_n *są zmiennymi losowymi o rozkładzie warunkowym* (względem $\mathscr{X}_n$) *danym dystrybuantą* G, *symetryczną względem zera i mającą gęstość* g. *Funkcje* M_n *są borelowskie.*

Założenie 2. h: $\boldsymbol{R} \to \boldsymbol{R}$ *jest nieparzystą funkcją borelowską, nieujemną na* $[0, \infty)$. *Ponadto całka* $\Psi(t) = \int h(t+v)g(v)dv = \int h(v)g(v-t)dv$ *istnieje dla wszystkich* $t \in \boldsymbol{R}$.

Ze względu na to, że przedmiotem naczelnego zainteresowania jest znalezienie optymalnej funkcji obserwacji (chciałoby się dodać — porównywalnej z funkcją $h = I$, przyjętą w oryginalnych procedurach RM i KW), w pracy Abdelhamida wyprowadzone zostały warunki gwarantujące zbieżność (1) z $h \neq I$, jeżeli założenia twierdzenia 1 są spełnione dla $h = I$. (Z tego samego powodu warunek (12) w twierdzeniu 2 sformułowano dla funkcji M_n, a nie bezpośrednio dla $\tilde{M}_n$).

2. Asymptotyczne rozkłady procedur typu (1).

TWIERDZENIE 2 (Abdelhamid). *Niech* α_0, a, c, β *będą liczbami dodatnimi,* γ *— liczbą nieujemną,* $\xi_0 \in \boldsymbol{R}$. *W* (1) *przyjmijmy*

(8) $$a_n = an^{-1}, \quad c_n = cn^{-\gamma}, \quad 0 \leqslant \gamma < \tfrac{1}{2}, \quad n = 1, 2, \ldots$$

Załóżmy $X_n \to \Theta$ (w tym celu wystarcza, żeby $\tilde{M}_n$ spełniała (2)–(5), ponieważ (11) $\Rightarrow$ (6)).

Niech

(9) *h będzie ciągła prawie wszędzie (względem G);*
Ψ' *istnieje w zerze i* $\Psi'(0) = H(h) > 0$.

(10) $$\beta = 1-2\gamma \quad \textit{oraz} \quad a > \frac{\beta}{2\alpha_0 H(h)}.$$

funkcja S^2, *gdzie*

(11) $$S^2(t) = \int [h(t+v)-\Psi(t)]^2 g(v)dv, \qquad t \in R,$$

jest ograniczona i ponadto ciągła w zerze.

Niech, dalej, α_n, ξ_n *będą* $\mathscr{X}_n$*-mierzalnymi zmiennymi losowymi, przy czym dla* $s = \beta(2\gamma)^{-1}$, *jeżeli* $\gamma \neq 0$, *i* $s = 0$, *jeżeli* $\gamma = 0$:

(12) $$c_n^{-1} M_n(X_n) = \alpha_n(X_n - \Theta) + \xi_n c_n^s,$$

(13) $$\alpha_n \to \alpha_0 \quad \textit{oraz} \quad \xi_n \to \xi_0 \text{(}^2\text{)}.$$

Wtedy asymptotyczny rozkład $n^{\beta/2}(X_n - \Theta)$ *jest rozkładem normalnym o parametrach*

(14) *wartość oczekiwana* $= -2ac^s H(h)\xi_0[2a\alpha_0 H(h)-\beta]^{-1}$,

(15) *wariancja* $= a^2c^{-2}S_0^2(h)\,[2a\alpha_0 H(h)-\beta]^{-1}$,

gdzie $S_0^2(h) = S^2(0)$. ■

Twierdzenie 2 jest konsekwencją ogólnego twierdzenia Fabiana [4], słusznego dla wielowymiarowych procedur AS.

Następne dwa twierdzenia odnoszą się *explicite* do zadań KW lub RM.

Założenie 3 (Procedura KW). f: $\boldsymbol{R} \to \boldsymbol{R}$ *jest funkcją borelowską,*

$$\sup_{-k < x-\Theta < -(1/k)} \bar{D}f(x) < 0, \quad \inf_{(1/k) < x-\Theta < k} \underline{D}f(x) > 0,$$

dla pewnego Θ *i każdego naturalnego k, przy czym* $\bar{D}f(x)$ *i* $\underline{D}f(x)$ *oznaczają pochodne — górną i dolną — funkcji f w x. Istnieją takie stałe A, B, że dla wszystkich* $x \in \boldsymbol{R}$ $|f(x+1)-f(x)| < A|x-\Theta|+B$. *W* (1) *przyjmuje się* $c_n \to 0$, $\sum a_n = \infty$, $\sum a_n^2 c_n^{-2} < \infty$, *oraz dla* Y_n

(16) $$M_n(X_n) = f(X_n + c_n) - f(X_n - c_n),$$

(17) $$E_{\mathscr{X}_n}[V_n^2] \leqslant \sigma^2 < \infty \quad \textit{dla wszystkich } n.$$

(²) W przypadku $s = 0$ należy przyjąć dodatkowo $\xi_0 = 0$.

U w a g a 1. Założenie 3 implikuje spełnienie warunków (2)–(6) dla $h = I$.

TWIERDZENIE 3 (Zadanie KW; Abdelhamid). *Niech f i Y_n spełniają założenie* 3 *z wyjątkiem (być może)* (17). *Niech f''' istnieje i będzie ciągła w otoczeniu Θ, natomiast $f''(\Theta) = M > 0$. Załóżmy, że spełnione są warunki* (2)–(5), (8) *z* $\gamma = 1/6$, (9), (11) *oraz* $a > [6MH(h)]^{-1}$.

Wówczas $X_n \to \Theta$ i prawdziwa jest teza twierdzenia 2, *przy czym* $\beta = 2/3$, $s = 2$, $\xi_0 = \frac{1}{3}f'''(\Theta)$, $\alpha_0 = 2M$. ■

Twierdzenie analogiczne do powyższego obowiązuje, jeżeli założenia czynione dla f''' zastąpić identycznymi, ale dla f'', nadto zaś przyjąć $\gamma = 1/4$ i $a > [8MH(h)]^{-1}$; wówczas $\beta = 1/2$, $s = 1$, $\xi_0 = 0$, $\alpha_0 = 2M$.

TWIERDZENIE 4 (Uogólnione zadanie RM; Abdelhamid). *Niech f_n będzie ciągiem funkcji borelowskich: $\boldsymbol{R} \to \boldsymbol{R}$, takich, że dla $M_n = f_n$ spełnione są warunki* (2)–(4) *z* $\tau_n = 1$, $c_n = 1$ *i* $f_n(\Theta) = 0$. *Niech $d > 0$ i ciąg*

$$D_n(x) = \begin{cases} (x-\Theta)^{-1} f_n(x), & x \neq \Theta, \\ d, & x = \Theta,\ n = 1, 2, \ldots, \end{cases} \tag{18}$$

będzie zbieżny w sposób ciągły do d w punkcie Θ (tzn. dla dowolnego ciągu $x_n \in \boldsymbol{R}$, $n = 1, 2, \ldots$, zbieżność $x_n \to \Theta$ implikuje $D_n(x_n) \to d$). *Dla procedury* (1) *załóżmy* (2)–(5), (9), (11) *oraz* $a_n = an^{-1}$ *i* $a > [2dH(h)]^{-1}$. *Wtedy $X_n \to \Theta$ i prawdziwa jest teza twierdzenia* 2, *przy czym* $\beta = 1$, $s = 0$, $\xi_0 = 0$, $\alpha_0 = d$. ■

U w a g a 2. Jeżeli $f_n = f$, $n = 1, 2, \ldots$ (klasyczne zadanie RM), to $d = f'(\Theta)$, por. (18)([3]); w przypadku ogólnym wypada przyjąć $f_n'(\Theta) \to d$.

3. Minimalizacja drugiego momentu rozkładu asymptotycznego. Niech ξ oznacza zmienną losową o rozkładzie normalnym i parametrach danych przez (14), (15). Stąd

$$E\xi^2 = \frac{4a^2H^2(h)\xi_0^2}{(2a\alpha_0 H(h) - \beta^2)}\, c^{2s} + \frac{a^2 S_0^2(h)}{(2a\alpha_0 H(h) - \beta)}\, c^{-2}.$$

Przy założeniu $\xi_0 \neq 0$ (por. np. Tw. 3) można dokonać minimalizacji $E\xi^2$ względem (a, c), by w konsekwencji otrzymać optymalne wartości tych stałych ($a_0 = a_0(h)$, $c_0 = c_0(h)$). Jeżeli $\xi_0 = 0$ ($E\xi = 0$), optymalne c nie istnieje, można natomiast określić optymalną wartość $a_0 = a_0(h)$; w przypadku $\xi_0 = 0$ należy rozróżnić dwie sytuacje: $s \neq 0$ (szczególne zadanie KW — por. np. komentarz do Tw. 3) i $s = 0$ (zadanie RM, por. Tw. 4).

Mając dla podanych trzech przypadków $E\xi^2(a_0, c_0, h)$ lub, odpowiednio, $\mathrm{Var}\,\xi(a_0, h)$ (por. np. środkowy człon równości (20)), dokonuje się ich minimalizacji względem h.

([3]) Warunek $d > 0$ ($d = \alpha_0 > 0$, por. Tw. 2) nie pozwala brać pod uwagę np. funkcji $f(x) = |x-\Theta| \cdot (x-\Theta)$. Z drugiej strony wiadomo, że przy parzystej g i danej f, procedura F ($h = \mathrm{sign}$, $c_n = 1$) daje $X_n \to \Theta$ (mimo że oryginalna procedura RM jest wówczas rozbieżna, por. [3], str. 128, i [2]).

Niech $\mathscr{H}$ oznacza rodzinę wszystkich borelowskich funkcji h, spełniających założenie 2, $S_0^2(h) = 1$ i takich, że dla $t = 0$ dopuszczalne jest różniczkowanie funkcji Ψ pod znakiem całki:

$$0 < H(h) = \Psi'(0) = \int h(v)\big(-g'(v)\big)dv.$$

W $\mathscr{H}$ poszukiwanie optymalnej funkcji h_0 równoznaczne jest z maksymalizacją $H(h)$ (por. np. (20)). Jeśliby z założenia $S_0^2(h) = 1$ zrezygnować, optymalna h_0 określona byłaby z dokładnością do stałej multiplikatywnej, ulegającej redukcji zarówno w wyrażeniach na $E\xi^2$ jak i w (1).

Twierdzenie 5 (Abdelhamid i Anbar). *Niech gęstość g ma pochodną prawie wszędzie (względem G),*

$$0 < \Gamma^2 = \int (g'(v)/g(v))^2 dG(v) < \infty. \tag{19}$$

Niech, dalej, $h_0 = -\frac{1}{\Gamma}\cdot\left(\frac{g'}{g}\right)$ prawie wszędzie (względem G) i $h_0 \in \mathscr{H}$. Wówczas w $\mathscr{H}$, h^* maksymalizuje $H(h)$ wtedy i tylko wtedy, gdy $h^* = h_0$ prawie wszędzie (względem G). ■

Przykładowo, dla zadania z twierdzenia 4 ($\xi_0 = 0$, $s = 0$),

$$a_0(h_0) = \frac{1}{dH(h)}\bigg|_{h=h_0} = \frac{1}{d\Gamma},$$

$$\operatorname{Var}\xi(a_0, h_0) = \frac{S_0^2(h)}{d^2H^2(h)}\bigg|_{h=h_0} = (d\Gamma)^{-2}. \tag{20}$$

U w a g a 3. h_0 zależy tylko od założonej gęstości g, bez względu na naturę zadania AS (RM czy KW).

Optymalna funkcja obserwacji prowadzi do uzyskania wariancji minimalnej nie tylko wśród procedur AS typu (1), ale w klasie wszystkich regularnych nieobciążonych estymatorów parametru Θ.

W przypadku zadania RM fakt ten rozumie się następująco: niech

$$\hat{Y}_n = F_n(\Theta)+V_n, \quad f_n(x) = F_n(\Theta)-F_n(x), \quad n = 1, 2, \ldots, \tag{21}$$

gdzie F_n są znane (ale nieznane jest Θ), funkcje f_n spełniają warunki twierdzenia 4, $f_n'(\Theta) \to d > 0$, V_n są niezależne i mają rozkład G, spełniający założenie 1 oraz (19). Wówczas, jeżeli $t_n = t_n(\hat{Y}_1, \ldots, \hat{Y}_n)$ jest regularnym asymptotycznie normalnym estymatorem zgodnym Θ (opartym na n pierwszych obserwacjach), to dolna granica wariancji asymptotycznego rozkładu $n^{1/2}(t_n-\Theta)$ — odpowiadająca asymptotycznej efektywności t_n — dana jest przez (20). Zwraca się uwagę, że w (21) przyjęto, iż obserwacji podlegają nieobciążone oceny wartości $F_n(\Theta)$.

Optymalną procedurę (1) można oczywiście zastosować do sytuacji opisanej warunkiem (21) (przy założeniu $Y_n = \hat{Y}_n - F_n(X_n)$) i tym sposobem uzyskać rekurencyjną asymptotycznie efektywną procedurę estymacji parametru Θ (rozwiązanie szczególnego zadania estymacji parametru rozkładu prawdopodobieństwa) ([4]).

W [1] wykazano asymptotyczną efektywność procedury optymalnej także w przypadku zadania KW, przy założeniu $f(x) = d(x-\Theta)^2$; ostatnie ograniczenie wynikać musi stąd, że dla zadania KW, $\beta = 1$ wtedy tylko, gdy f — parzysta (por. [5], str. 444).

4. Uwagi. Na podstawie twierdzenia 5 otrzymuje się, że procedury RM i KW są optymalne wtedy i tylko wtedy, gdy G jest rozkładem normalnym; procedury F optymalne są natomiast wtedy i tylko wtedy, gdy G jest dwustronnym rozkładem wykładniczym.

W omawianych pracach podane są jeszcze optymalne funkcje obserwacji dla rozkładów Studenta, logistycznego i Hubera ($g(v)$ — normalna dla $|v| < T$ i dwustronnie wykładnicza dla $|v| \geqslant T$). Ponadto Anbar przedstawił rozwiązanie optymalne dla zadania RM i G-normalnego ε-zniekształconego ($G(v) = (1-\varepsilon)\Phi(v) + \varepsilon H(v)$, gdzie Φ — rozkład normalny, $N(0, 1)$, H — nieznany rozkład symetryczny i ε — znana stała dodatnia).

W rozpatrywanych zadaniach, a_0 i c_0 (lub tylko a_0) zależą od nieznanych (!) wartości $f'(\Theta)$, $f''(\Theta)$ albo $f'''(\Theta)$. Dla zadania RM procedura estymacji optymalnego a_0 opracowana została w [4]; dla zadania KW procedura estymacji a_0 lub jednocześnie a_0 i c_0 podana jest w [5].

Konstrukcja procedury optymalnej wymaga natomiast znajomości gęstości g. (Ostatnio wszakże w [6] zaproponowano procedurę, która nie wymaga — w przypadku zadania RM — znajomości g i estymuje jednocześnie h_0 oraz a_0).

Wymienione procedury estymacji nie zmieniają asymptotycznych własności procedury optymalnej (1).

([4]) W [7] znaleźć można rekurencyjne asymptotycznie efektywne procedury rozwiązujące ogólne zadanie estymacji parametrów rozkładu; w przypadku zadania (21) rozwiązania przedstawione tutaj i w [7] pokrywają się.

Prace cytowane

[1] S. N. Abdelhamid, *Transformation of observations in stochastic approximation*, Praca doktorska, Michigan State Univ. 1971.

[2] Э. Д. Аведьян, *К одной модификации алгоритма Роббинса-Монро*, Авт. и Телемех. 4 (1967), str. 165-167.

[3] V. Fabian, *Stochastic approximation methods*, Czech. Math. J. 10 (1960), str. 123-159.

[4] —, *On asymptotic normality in stochastic approximation*, Ann. Math. Statist. 39 (1968), str. 457-465.

[5] —, *Stochastic approximation*, w *Optimizing methods in statistics*, J. S. Rustagi (ed.) Academic Press, New York 1971.

[6] —, *Asymptotically efficient stochastic approximation; the RM case*, Ann. Statist. 1.3 (1973), str. 486-495.

[7] М. Б. Невельсон, Р. З. Хасьминский, *Стохастическая аппроксимация и рекуррентное оценивание*, Изд. Наука, Москва 1972.

Jacek Koronacki
Warszawa