ROCZNIKI POLSKIEGO TOWARZYSTWA MATEMATYCZNEGO
SERIA III: MATEMATYKA STOSOWANA III (1974)

R. ZIELIŃSKI (Warszawa)

# H. Robbins, D. Siegmund: Twierdzenie o zbieżności nieujemnych prawie supermartyngałów i pewne jego zastosowania*

W pracy przedstawiono pewne ogólne twierdzenie o zbieżności nieujemnych prawie supermartyngałów (tzn. ciągów $z_n$ spełniających podany niżej, w twierdzeniu 1, warunek) i pokazano kilka interesujących zastosowań tego twierdzenia do dowodu niektórych znanych faktów. Na podkreślenie zasługuje to, że dość ogólne, ale stosunkowo łatwe, twierdzenie Robbinsa-Siegmunda obejmuje bardzo zróżnicowane przypadki, do tej pory rozważane i dowodzone oddzielnie, nieraz bardzo żmudnymi metodami.

TWIERDZENIE 1. *Niech* $(\Omega, F, P)$ *będzie pewną przestrzenią probabilistyczną,* $F_1 \subset F_2 \subset \ldots$ *ciągiem pod-σ-ciał* $F$. *Dla każdego* $n = 1, 2, \ldots$ *niech* $z_n, \beta_n, \xi_n, \zeta_n$ *będą nieujemnymi* $F_n$*-mierzalnymi zmiennymi losowymi takimi, że*

$$\mathrm{E}(z_{n+1} | F_n) \leqslant z_n(1+\beta_n) + \xi_n - \zeta_n .$$

*Wtedy* $\lim\limits_{n\to\infty} z_n$ *istnieje i jest skończona oraz* $\sum\limits_1^\infty \zeta_n < \infty$ *z prawdopodobieństwem* 1 *na zbiorze* $\left\{\sum\limits_1^\infty \beta_n < \infty,\ \sum\limits_1^\infty \xi_n < \infty\right\}$.

**Zastosowanie 1** (*mocne prawo wielkich liczb*). Niech $y_1, y_2, \ldots$ będzie ciągiem zmiennych losowych takich, że $\mathrm{E}(y_{n+1} | F_n) = 0$, $n = 1,2,\ldots$ Niech $c_n$ $(n = 1,2,\ldots)$ będą dodatnimi $F_{n-1}$ mierzalnymi zmiennymi losowymi takimi, że $c_1 \leqslant c_2 \leqslant \ldots$ Wtedy $\lim c_n^{-1} |\sum\limits_{k=1}^n y_k|$ istnieje i jest skończona na zbiorze $A = \left\{\sum c_k^{-2}\, \mathrm{E}(y_k^2 | F_{k-1}) < \infty\right\}$.

Ta granica jest równa zeru z prawdopodobieństwem jeden na zbiorze $\{c_n \uparrow \infty\} \cap A$.

Z tego ostatniego twierdzenia w szczególności otrzymujemy:

WNIOSEK 1. *Niech* $y_1, y_2, \ldots$ *będzie ciągiem niezależnych zmiennych losowych takich, że* $\mathrm{E}y_k = 0$, $k = 1,2,\ldots$ *Jeżeli* $\sum n^{-2}\, \mathrm{E}y_n^2 < \infty$, *to* $\lim\limits_n n^{-1} \sum\limits_{k=1}^n y_k = 0$ *z prawdopodobieństwem jeden.*

---

* *A convergence theorem for non negative almost supermartingales and some applications.* W książce: *Optimizing Methods in Statistics*, J.S. Rustagi (ed). New York 1971.

WNIOSEK 2. *Niech* $r_1, r_2, \ldots$ *będzie ciągiem zmiennych losowych przyjmujących wartości* 0 *lub* 1 *i niech* $p_k = \mathrm{P}(r_k = 1 \mid r_1, \ldots, r_{k-1}), p_1 > 0$. *Wtedy* $\lim_n (r_1 + \ldots + r_n)/(p_1 + \ldots + p_n)$ *istnieje i jest skończona z prawdopodobieństwem jeden. Ta granica jest równa* 1 *gdy* $\sum p_k = \infty$.

**Zastosowanie 2** (*proces Robbinsa-Monro*). Niech $\{y(x),\ x \in R^1\}$ będzie rodziną zmiennych losowych takich, że $\mathrm{E}y(x) = M(x)$ oraz $\mathrm{Var}\, y(x) = \sigma^2(x) < \infty$. Zadanie polega na wyznaczeniu pierwiastka $\vartheta$ równania $M(x) = 0$ na podstawie obserwacji zmiennych losowych $y(x)$ w różnych (dowolnie wybieranych) punktach $x \in R^1$. Rozważa się następujący algorytm

$$x_{n+1} = x_n - a_n y_n$$

gdzie $x_1$ — dowolna zmienna losowa, $a_n$ — nieujemne zmienne losowe mierzalne względem $\sigma$-ciała generowanego przez $x_1, y_1, \ldots, y_{n-1}$ oraz $\mathrm{P}(y_n < t \mid x_1, y_1, \ldots, y_{n-1}) = \mathrm{P}(y(x_n) < t)$.

Jeżeli:

(i) dla pewnych $a, b > 0$

$$\sigma(x) + |M(x)| \leqslant a + b|x|,$$

(ii) istnieje takie $\vartheta$, że dla każdego $\varepsilon > 0$

$$\inf_{\varepsilon < x - \vartheta < \varepsilon^{-1}} M(x) > 0 \qquad \text{oraz} \qquad \sup_{\varepsilon < \vartheta - x < \varepsilon^{-1}} M(x) < 0,$$

(iii) $\sum a_n^2 < \infty$ dla każdego ciągu $x_1, y_1, \ldots,$

(iv) $\sum a_n = \infty$ dla każdego ciągu $x_1, y_1, \ldots$ takiego, że $\sup |x_n| < \infty$, to $\lim x_n = \vartheta$ z prawdopodobieństwem jeden.

**Zastosowanie 3** (*zmodyfikowany proces Robbinsa-Monro*). W zmodyfikowanym procesie rozpatruje się ciąg kolejnych przybliżeń $\tilde{x}_1, \tilde{x}_2, \ldots$ określony w następujący sposób:

$$\tilde{x}_{n+1} = \begin{cases} (\tilde{x}_n - a_n y_n) \cdot I\{\tilde{x}_n - a_n y_n \geqslant \underline{\vartheta}\} + \underline{\vartheta} \cdot I\{\tilde{x}_n - a_n y_n < \underline{\vartheta}\} & \text{gdy } \tilde{x}_n > \bar{\vartheta}, \\ \tilde{x}_n - a_n y_n & \text{gdy } \underline{\vartheta} \leqslant \tilde{x}_n \leqslant \bar{\vartheta}, \\ (\tilde{x}_n - a_n y_n) \cdot I\{\tilde{x}_n - a_n y_n \leqslant \bar{\vartheta}\} + \bar{\vartheta} \cdot I\{\tilde{x}_n - a_n y_n > \bar{\vartheta}\} & \text{gdy } \tilde{x}_n \leqslant \underline{\vartheta}, \end{cases}$$

gdzie $\underline{\vartheta}, \bar{\vartheta}$ ($\underline{\vartheta} < \bar{\vartheta}$) są pewnymi ustalonymi liczbami (o których przypuszcza się, że szukane $\vartheta \in (\underline{\vartheta}, \bar{\vartheta})$; jest to więc procedura uwzględniająca pewną informację a priori). Symbol $I\{x_1 > x_2\}$ oznacza funkcję charakterystyczną zbioru tych $\omega \in \Omega$ dla których $x_1 > x_2$.

Jeżeli funkcja $M$ jest lokalnie ograniczona oraz dla pewnych $a, b > 0$

$$\sigma^2(x) \leqslant a + bx^2$$

i spełnione są założenia (ii), (iii) i (iv) wymienione w zastosowaniu 2, to $\mathrm{P}(\lim \tilde{x}_n = \vartheta) = 1$.

**Zastosowanie 4** (*aproksymacja stochastyczna, schemat Dvoretzky'ego*)**.** Niech $X$ będzie przestrzenią Hilberta oraz $X_n = \underbrace{X \times X \times \ldots \times X}_{n \text{ razy}}$. Niech $T_n : X_n \to X$ będzie przekształceniem borelowskim spełniającym dla pewnego $\vartheta$ warunek:

$$\| T_n(x_1, \ldots, x_n) - \vartheta \| \leqslant \max \{ a_n, (1 + b_n) \| x_n - \vartheta \| - c_n \},$$

gdzie $a_n$, $b_n$, $c_n$ są nieujemnymi, borelowskimi funkcjami na $X_n$ takimi, że $\lim a_n = 0$ dla wszystkich $(x_1, x_2, \ldots)$, $\sum b_n < \infty$ dla wszystkich $(x_1, x_2, \ldots)$ oraz $\sum c_n = \infty$ dla wszystkich $(x_1, x_2, \ldots)$ takich, że $\sup \| x_n \| < \infty$.

Niech $x_1, y_1, y_2, \ldots$ będą zmiennymi losowymi o wartościach w $X$ i zdefiniujmy

$$x_{n+1} = T_n(x_1, \ldots, x_n) + y_n, \qquad n = 1,2,\ldots$$

Jeżeli $\mathrm{E}(y_n \mid x_1, y_1, \ldots, y_{n-1}) = 0$, $n = 1,2, \ldots$ oraz $\sum \mathrm{E}(\| y_n \|^2 \mid x_1, y_1, \ldots, y_{n-1}) < \infty$, to $\mathrm{P}(\lim \| x_n - \vartheta \| = 0) = 1$.

**Zastosowanie 5** (*teoria gier*)**.** Rozpatruje się grę dwuosobową, w której każdy z graczy ma skończoną liczbę strategii. Jeżeli gracz I wybierze strategię $i$ oraz gracz II strategię $j$, to gracz I wygrywa sumę, która jest zmienną losową o rozkładzie $m(i, j)$, o wartościach w pewnym zbiorze $X$. Niech $\bar{x}_n$ będzie średnią wygraną gracza I w $n$ pierwszych grach. Gracze rozgrywają ciąg gier. Powstaje pytanie, czy istnieje taka strategia dla gracza I, żeby przy dowolnym postępowaniu gracza II ciąg $\bar{x}_n$ był zbieżny do pewnego ustalonego podzbioru $S \subset X$. Niech $d_n$ będzie odległością $\bar{x}_n$ od zbioru $S$. Przy pewnych założeniach o zbiorach $X$, $S$ i macierzy $(m(i, j))$ Blackwell dowiódł, że gracz ma taką strategię, dla której przy pewnych $a$, $b$, $c$:

(i) $$\mathrm{E}(d_n \mid d_1, \ldots, d_{n-1}) \leqslant \left(1 - \frac{2}{n}\right) d_{n-1} + \frac{c}{n^2}, \qquad \text{gdy } d_{n-1} > 0,$$

(ii) $$0 \leqslant d_n \leqslant a,$$

(iii) $$| d_n - d_{n-1} | \leqslant b/n.$$

Korzystając z twierdzenia 1 dowodzi się, że ciąg $(d_n)$ spełniający warunki (i), (ii), (iii) jest zbieżny do zera oraz że dla każdego $\varepsilon > 0$ istnieje takie $n_0 = n_0(\varepsilon, a, b, c)$, że

$$\mathrm{P}\{ \sup_{n \geqslant n_0} d_n \geqslant \varepsilon \} \leqslant \varepsilon$$

jednostajnie ze względu na wszystkie ciągi $(d_n)$.

**Zastosowanie 6** (*analiza skupień*)**.** Niech zmienna losowa $y$ ma ciągły rozkład $p$, którego nośnikiem jest domknięty, ograniczony i wypukły zbiór $R \subset E^N$. Niech $S = (S_1, \ldots, S_k)$, $k$ — ustalona liczba, będzie podziałem zbioru $R$ oraz niech $v = (v_1, \ldots, v_k)$, $v_i \in S_i$, będzie punktem związanym z tym podziałem. Niech

$$w(S, v) = \sum_{i=1}^{k} \int_{S_i} \| y - v_i \|^2 p(dy).$$

Należy tak dobrać podział $S$ i punkt $v$, żeby $w(S, v)$ było możliwie małe. W zastosowaniach zadanie to jest znane jako zadanie optymalnego podziału na klasy.

Ustalmy $x = (x_1, x_2, ..., x_k)$, $x_i \in E^N$ i zdefiniujemy podział $S(x) = (S_1(x), ..., S_k(x))$ w następujący sposób:

$$S_i(x) = \{ \xi \in E^N : \| \xi - x_i \| \leqslant \| \xi - x_j \|, \ j = 1,2, ..., k \}.$$

Niech

$$W(x) = \sum_{i=1}^{k} \int_{S_i(x)} \| y - x_i \|^2 p(dy),$$

$$V(x) = \sum_{i=1}^{k} \int_{S_i(x)} \| y - u_i(x) \|^2 p(dy),$$

gdzie

$$u_i(x) = \int_{S_i(x)} y p(dy) / p(S_i(x)).$$

Rozpatruje się następującą sekwencyjną procedurę budowy podziału $S$.

Niech $y_1, ..., y_k$ będzie $k$-elementową próbką dla zmiennej losowej $y$. Zdefiniujemy

$$x^1 = (x_1^1, ..., x_k^1), \qquad x_i^1 = y_i$$

oraz liczby

$$w_1^1, ..., w_k^1, \qquad w_i^1 = 1$$

i skonstruujemy podział $S_1 = S(x^1)$ – będzie to pierwsze przybliżenie poszukiwanego optymalnego podziału.

Niech $y_{k+n}$, $n = 1,2,...$ będą dalszymi kolejnymi elementami próbki. Dla $y_{k+n} \in S_i(x^n)$ definiujemy

$$x_i^{n+1} = \frac{w_i^n x_i^n + y_{k+n}}{w_i^n + 1}, \qquad x_j^{n+1} = x_j^n \qquad \text{dla } j \neq i,$$

$$w_i^{n+1} = w_i^n + 1, \qquad w_j^{n+1} = w_j^n \qquad \text{dla } j \neq i$$

i konstruujemy kolejne podziały $S(x^n)$, $n = 1,2, ...$

Korzystając z twierdzenia 1 dowodzi się, że $\lim_n W(x^n)$ istnieje z prawdopodobieństwem jeden i jest równe $V(x)$ dla pewnego $x$ w klasie takich $x^* = (x_1^*, ..., x_k^*)$, że $x_i^* \neq x_j^*$ gdy $i \neq j$ oraz $x_i^* = u_i(x^*)$. Ponadto

$$\lim_{m\to\infty} m^{-1} \sum_{n=1}^{m} \left( \sum_{i=1}^{k} p(S_i(x^n)) \cdot \| x_i^n - u_i(x^n) \| \right) = 0$$

z prawdopodobieństwem jeden.

**Zastosowanie 7.** Za pomocą twierdzenia 1 dowodzi się dwóch nierówności dla martyngałów, które podali Hájek-Rényi-Chow (1960) oraz Dubins-Freedman (1965).