ROCZNIKI POLSKIEGO TOWARZYSTWA MATEMATYCZNEGO
Seria II: WIADOMOŚCI MATEMATYCZNE XXIX (1990)

wmptm 1897

Ministerstwo Nauki i Szkolnictwa Wyższego

WIESŁAW DZIUBDZIELA (Wrocław, Kielce)

O czasach rekordowych i liczbie rekordów w ciągu zmiennych losowych

1. Wstęp. W ostatnich dziesięcioleciach nastąpił burzliwy rozwój matematycznej teorii rekordów, będącej częścią teorii wartości ekstremalnych. W niedawno opublikowanym artykule przeglądowym Niewzorowa [28] poświęconym rekordom odnotowane jest 166 pozycji bibliograficznych, z których więcej niż połowa ukazała się w ostatnim dziesięcioleciu.

Zainteresowanie rekordami jest spowodowane dwiema przyczynami. Po pierwsze, w teorii rekordów jest jeszcze wiele problemów matematycznych czekających na rozwiązanie, a otrzymywane dotychczas wyniki są zazwyczaj eleganckie matematycznie i często zaskakujące. Po drugie, możliwych jest wiele zastosowań teorii rekordów w problemach, w których interesujące jest badanie wielkości będących największymi lub najmniejszymi ze wszystkich zaobserwowanych do danego momentu czasu. Najbardziej interesujące są zastosowania w meteorologii (np. rekordowe opady miesięczne, rekordowe nasłonecznienie, rekordowa szybkość wiatru itp.), w sporcie, w teorii wytrzymałości materiałów.

Pierwszą pracą poświęconą rekordom jest praca Chandlera [7], która ukazała się w 1952 roku. Uważa się, że pracami, które najwięcej wniosły w rozwój teorii, są prace: Rényiego [29]–[31], Tata [38] oraz cykl prac Shorrocka [33]–[35]. Podstawy matematycznej teorii rekordów można znaleźć w pracy przeglądowej Glicka [20] i monografii Galambosa [19]. Bardziej zaawansowana teoria przedstawiona jest w monografii Resnicka [32].

W niniejszym opracowaniu ze zrozumiałych względów ograniczymy się jedynie do przedstawienia wybranych zagadnień teorii rekordów i jej zastosowań.

2. Definicje i podstawowe własności czasów rekordowych. Niech $\{X_n, n \geqslant 1\}$ będzie ciągiem zmiennych losowych. Definiujemy ciąg zmiennych losowych

(2.1) $L(0) = 0, \quad L(1) = 1, \quad L(n+1) = \min\{j: X_j > X_{L(n)}\}, \quad n = 1, 2, \ldots$

Zmienne losowe $L(n)$, $n = 1, 2, \ldots$, nazywamy *czasami rekordowymi*. Interesować nas będą także zmienne losowe $\Delta(n) = L(n) - L(n-1)$, $n = 1, 2, \ldots$, nazywane *czasami międzyrekordowymi* i $R(n) = X_{L(n)}$, $n = 1, 2, \ldots$, nazywane *wartościami rekordowymi* lub krótko *rekordami*.

We wspomnianej już pracy [7] Chandler badał ciągi $\{L(n)\}$, $\{\Delta(n)\}$ i $\{R(n)\}$ przy założeniu, że ciąg $\{X_n, n \geqslant 1\}$ jest ciągiem niezależnych zmiennych losowych o jednakowej ciągłej dystrybuancie F, i udowodnił między innymi następujący lemat i twierdzenie:

LEMAT 2.1. *Rozkład prawdopodobieństwa zmiennej losowej* $L(n)$, $n = 1,2, \ldots$, *nie zależy od dystrybuanty* F.

Dowód. Oznaczmy przez $L^*(n)$, n = 1, 2, ..., czasy rekordowe w ciągu $\{F(X_n), n \geqslant 1\}$. Ponieważ F jest ciągłą dystrybuantą, więc ciąg $\{F(X_n), n \geqslant 1\}$ jest ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie jednostajnym na przedziale (0, 1). Zauważmy także, że jeżeli $X_i \leqslant X_j$, to $F(X_i) \leqslant F(X_j)$. Z powyższego wynika, że $P\{L(n) = k\} = P\{L^*(n) = k\}$, $k = 0, 1, \ldots$, czyli rozkłady prawdopodobieństwa zmiennych losowych $L(n)$ i $L^*(n)$ są jednakowe. Zatem rozkład prawdopodobieństwa zmiennej losowej $L(n)$ nie zależy od postaci dystrybuanty F i może być obliczany przy dodatkowym założeniu, że F jest dystrybuantą rozkładu jednostajnego na przedziale (0, 1).

TWIERDZENIE 2.1. *Rozkład zmiennej losowej* $L(2)$ *jest postaci*

$$P\{L(2) = j\} = \frac{1}{j(j-1)}, \quad j = 2, 3, \ldots \tag{2.2}$$

Dowód. Korzystając z lematu 2.1, możemy przyjąć, że $\{X_n, n \geqslant 1\}$ jest ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie jednostajnym na przedziale (0, 1). Wtedy dla $j = 2, 3, \ldots$ mamy

$$P\{L(2) = j\} = \int_0^1 P\{L(2) = j \mid X_1 = x\}\, dx$$

$$= \int_0^1 x^{j-2}(1-x)\, dx = \frac{1}{j-1} - \frac{1}{j} = \frac{1}{j(j-1)}.$$

Interesujący jest następujący wniosek z twierdzenia 2.1.

WNIOSEK. *Dla* $n = 2, 3, \ldots$

$$E(L(n)) = +\infty. \tag{2.3}$$

Dowód. Z (2.2) otrzymujemy

$$E(L(2)) = \sum_{j=2}^{\infty} jP\{L(2) = j\} = \sum_{j=2}^{\infty} \frac{1}{j-1} = +\infty.$$

Stąd i z nierówności

$$L(n) \geqslant L(2), \quad n = 2, 3, \ldots,$$

otrzymujemy (2.3).

Znalezienie rozkładów prawdopodobieństwa czasów rekordowych $L(n)$ dla dowolnego $n > 2$ jest bardziej skomplikowane. Rényi [29] udowodnił, że ciąg $\{L(n), n \geqslant 2\}$ tworzy łańcuch Markowa z prawdopodobieństwami przejść

$$P\{L(n) = k \mid L(n-1) = l\} = \frac{l}{k(k-l)},$$

gdy $k > l \geqslant n-1$, $n = 2, 3, \ldots$ Własność tę wykorzystał on do znalezienia dwuwymiarowych łącznych rozkładów czasów rekordowych i w konsekwencji do znalezienia rozkładów brzegowych.

Praca Chandlera [7] spowodowała pojawienie się w drugiej połowie lat pięćdziesiątych serii prac [17], [18], [36], [37], w których rekordy wykorzystano w zagadnieniach związanych z testowaniem hipotez statystycznych. Wkrótce jednak nastąpił spadek zainteresowania statystyków teorią rekordów. Wydaje się to związane z faktem, że $E(L(2)) = +\infty$, tzn. czas czekania na rekord ma wartość oczekiwaną nieskończoną. Feller w drugim tomie swojej ogólnie znanej książki [16] (Rozdział I, § 5. *Prześladowanie przez pech*) podaje ten wynik jako przykład ilustrujący nieoczekiwane własności fluktuacji losowych i wyjaśniający pech, który zwykle prześladuje nas w życiu.

Impulsem, który pobudził dalszy rozwój teorii rekordów i jej zastosowań, była praca Rényiego [29] opublikowana w roku 1962. Główny jej wynik omówimy w następnym rozdziale. Obecnie przedstawimy naturalne rozszerzenie pojęcia rekordu.

Dziubdziela i Kopociński [14] (zob. także [12]) w roku 1976 zdefiniowali tzw. k-te rekordy w sposób następujący: Niech $M_n^{(k)}$, $k = 1, \ldots, n$, będzie *k-tą statystyką pozycyjną* w ciągu $X_1, \ldots, X_n$, tzn. zmienną losową, która dla danego $\omega \in \Omega$ przyjmuje k-tą z kolei spośród wartości $X_j(\omega)$. Oczywiście

$$M_n^{(n)} = \min(X_1, \ldots, X_n) \leqslant M_n^{(n-1)} \leqslant \ldots \leqslant M_n^{(1)} = \max(X_1, \ldots, X_n).$$

Dla ustalonego $k \geqslant 1$ weźmy pod uwagę ciąg $\{M_m^{(k)}, m \geqslant k\}$ k-tych statystyk pozycyjnych. Ciąg $\{M_m^{(k)}\}$ jest ciągiem niemalejącym. Przez skreślenie w nim powtórzeń otrzymujemy ciąg ściśle rosnący, który nazywamy ciągiem *k-tych rekordów*. Dokładniej, definiujemy zmienne losowe $l^{(k)}(n)$, $n = 1, 2, \ldots$, będące *numerami rekordów* w ciągu $\{M_m^{(k)}\}$:

$$l^{(k)}(1) = 1, \quad l^{(k)}(n+1) = \min\{j: M^{(k)}_{l^{(k)}(n)+k-1} < M^{(k)}_{j+k-1}\}, \quad n = 1, 2, \ldots$$

Zmienne losowe

$$R^{(k)}(n) = M^{(k)}_{l^{(k)}(n)+k-1}, \quad n = 1, 2, \ldots,$$

nazywamy *k-tymi rekordami.* Oczywiście dla $k = 1$ otrzymujemy czasy rekordowe i wartości rekordowe.

Na początku lat osiemdziesiątych Deheuvels [8]–[10] definiuje i bada tzw. *k-te czasy rekordowe*

$$L^{(k)}(1) = k, \quad L^{(k)}(n+1) = \min(j: j > L^{(k)}(n), X_j > M^{(k)}_{j-1}\}, \quad n = 1, 2, \ldots$$

Zauważmy, że

$$L^{(k)}(n) = l^{(k)}(n)+k-1, \quad n = 1, 2, \ldots,$$

oraz

$$R^{(k)}(n) = M^{(k)}_{L^{(k)}(n)}, \quad n = 1, 2, \ldots$$

Ciąg $\{L^{(k)}(n), n \geqslant 1\}$ jest ciągiem monotonicznym, a można go interpretować jako ciąg indeksów n, dla których w ciągu $\{X_n, n \geqslant 1\}$ otrzymuje się nowy k-ty rekord. Nie są to jednak miejsca k-tych rekordów w ciągu $\{X_n\}$. Na przykład, niech

$$X_1(\omega) = 8, \quad X_2(\omega) = 12, \quad X_3(\omega) = 7, \quad X_4(\omega) = 10, \quad X_5(\omega) = 15, \ldots,$$

będzie realizacją ciągu $\{X_n, n \geqslant 1\}$ dla ustalonego zdarzenia elementarnego ω. Wtedy

$$L^{(2)}(1) = 2, \quad L^{(2)}(2) = 4, \quad L^{(2)}(3) = 5, \quad \ldots$$

oraz

$$R^{(2)}(1) = X_1 = 8, \quad R^{(2)}(2) = X_4 = 10, \quad R^{(2)}(3) = X_2 = 12, \quad \ldots,$$

natomiast miejsca k-tych rekordów w ciągu $\{X_n, n \geqslant 1\}$ są następujące: 1, 4, 2, ... Zatem można mieć wątpliwości co do słuszności nazwy „k-te czasy rekordowe".

3. Lemat Rényiego i jego konsekwencje. Jak już wspomnieliśmy, pod koniec lat pięćdziesiątych zaznaczył się spadek zainteresowania teorią rekordów. Nowy impuls do badań dały opublikowane w 1962 roku wyniki Rényiego [29]–[31].

Niech $\{X_n, n \geqslant 1\}$ będzie ciągiem zmiennych losowych. Definiujemy ciąg $\{\xi_n, n \geqslant 1\}$ zmiennych losowych przyjmujących wartości 0 i 1 w sposób następujący:

$$\xi_n = \begin{cases} 1, & \text{jeżeli } X_n \text{ jest rekordem,} \\ 0, & \text{w przeciwnym wypadku,} \end{cases}$$

$n = 1, 2, \ldots$ W celu bardziej precyzyjnego zapisu powyższej definicji wpro-

wadzimy następujące pojęcie: *Rangą zmiennej losowej* X_n *wśród* $X_1, \ldots, X_n$ nazywamy zmienną losową

$$\varrho_n = \sum_{j=1}^{n} 1_{\{X_j \leqslant X_n\}},$$

gdzie 1_A jest indykatorem zdarzenia losowego A, tzn.

$$1_A(\omega) = \begin{cases} 1, & \text{jeżeli } \omega \in A, \\ 0, & \text{jeżeli } \omega \notin A. \end{cases}$$

Innymi słowy, $\varrho_n(\omega)$ jest liczbą wskaźników j takich, że $X_j(\omega) \leqslant X_n(\omega)$. Oczywiście

$$X_n = M_n^{(n-\varrho_n+1)},$$

czyli jeżeli $X_n(\omega) = \max(X_1(\omega), \ldots, X_n(\omega)) = M_n^{(1)}(\omega)$, to $\varrho_n(\omega) = n$; jeżeli $X_n(\omega) = M_n^{(2)}(\omega)$, to $\varrho_n(\omega) = n-1$ itd. Teraz zmienne losowe ξ_n możemy zdefiniować następująco:

$$\xi_n = 1_{\{\varrho_n = n\}} = 1_{\{X_n \text{ jest rekordem}\}}, \qquad n = 1, 2, \ldots$$

LEMAT 3.1 (A. Rényi). *Jeżeli* $\{X_n, n \geqslant 1\}$ *jest ciągiem niezależnych zmiennych losowych o jednakowej ciągłej dystrybuancie, to zmienne losowe* $\xi_1, \xi_2, \ldots$ *są niezależne oraz*

$$P\{\xi_n = 1\} = 1/n, \qquad n = 1, 2, \ldots$$

Powyższy lemat, który odegrał kluczową rolę w teorii rekordów, jest natychmiastowym wnioskiem z następującego twierdzenia, którego znaczenie wykracza poza zakres teorii rekordów:

TWIERDZENIE 3.1 ([29], [11], [6]). *Jeżeli* $\{X_n, n \geqslant 1\}$ *jest ciągiem niezależnych zmiennych losowych o jednakowej ciągłej dystrybuancie* F, *to rangi* $\varrho_1, \varrho_2, \ldots$ *są niezależnymi zmiennymi losowymi oraz*

$$P\{\varrho_n = j\} = 1/n, \qquad j = 1, \ldots, n, \; n = 1, 2, \ldots$$

Dowód twierdzenia. Z założenia ciągłości dystrybuanty F wynika, że równość wśród realizacji zmiennych losowych $X_1, \ldots, X_n$ występuje z prawdopodobieństwem zero. Zatem każde z $n!$ uporządkowań

$$X_{i_1} < \ldots < X_{i_n}$$

jest jednakowo prawdopodobne ($i_1, \ldots, i_n$ oznacza permutację liczb $1, \ldots, n$). Istnieje wzajemnie jednoznaczna odpowiedniość między tymi uporządkowaniami a realizacjami rang $\varrho_1, \ldots, \varrho_n$; np. dla $n = 3$ uporządkowanie $X_3 < X_1 < X_2$ odpowiada realizacji $\varrho_1 = 1$, $\varrho_2 = 2$, $\varrho_3 = 1$. Z powyższego wynika, że każda realizacja $\varrho_1, \ldots, \varrho_n$ ma jednakowe prawdopodobieństwo, tzn.

$$P\{\varrho_1 = r_1, \ldots, \varrho_n = r_n\} = 1/n!$$

dla $r_i = 1, \ldots, i$, $i = 1, \ldots, n$. Zauważmy, że

$$P\{\varrho_n = r_n\} = \sum_{r_1,\ldots,r_{n-1}} P\{\varrho_1 = r_1, \ldots, \varrho_n = r_n\}$$

$$= \sum_{r_1,\ldots,r_{n-1}} \frac{1}{n!} = \frac{1\cdot 2\cdot \ldots \cdot (n-1)}{n!} = \frac{1}{n},$$

czyli

$$P\{\varrho_1 = r_1, \ldots, \varrho_n = r_n\} = \prod_{i=1}^{n} P\{\varrho_i = r_i\},$$

zatem $\varrho_1, \ldots, \varrho_n$ są niezależnymi zmiennymi losowymi.

Zdefiniujmy dla $n = 1, 2, \ldots$ zmienną losową

$$N_n = \sum_{j=1}^{n} \xi_j.$$

Interpretuje się ją jako liczbę rekordów wśród pierwszych n zmiennych losowych $X_1, \ldots, X_n$.

Badanie zmiennej losowej N_n odgrywa ważną rolę w teorii rekordów. Nie jest ono specjalnie trudne, gdyż N_n jest sumą niezależnych zmiennych losowych $\xi_1, \xi_2, \ldots$ przyjmujących jedynie wartości 0 i 1, przy czym

$$E(\xi_n) = \frac{1}{n}, \qquad \operatorname{Var}(\xi_n) = \frac{1}{n}\left(1 - \frac{1}{n}\right).$$

Zatem

$$E(N_n) = \sum_{j=1}^{n} \frac{1}{j} \sim \log n, \qquad \text{gdy } n \to \infty$$

(„$a_n \sim b_n$, gdy $n \to \infty$" oznacza, że $\lim_{n\to\infty}(a_n/b_n) = 1$).

Z elementarnego rachunku prawdopodobieństwa wynika, że

$$P\{\lim_{n\to\infty} N_n/\log n = 1\} = 1$$

oraz

$$\lim_{n\to\infty} P\left\{\frac{N_n - \log n}{\sqrt{\log n}} < x\right\} = \Phi(x), \qquad -\infty < x < \infty,$$

gdzie

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{t^2}{2}\right) dt$$

jest dystrybuantą standardowego rozkładu normalnego. Stąd już łatwo udowodnić (zob. np. [19]), korzystając po drodze z równości

$$P\{L(n) > m\} = P\{N_m < n\},$$

odpowiednie twierdzenia dla czasów rekordowych $L(n)$ i czasów międzyrekordowych $\Delta(n)$:

(i) Z prawdopodobieństwem równym 1

$$\lim_{n\to\infty}\frac{\log L(n)}{n}=\lim_{n\to\infty}\frac{\log \Delta(n)}{n}=1,$$

(ii) $$\lim_{n\to\infty} P\left\{\frac{\log L(n)-n}{\sqrt{n}}<x\right\}$$

$$=\lim_{n\to\infty} P\left\{\frac{\log \Delta(n)-n}{\sqrt{n}}<x\right\}=\Phi(x), \qquad -\infty<x<\infty.$$

Niewzorow [24] uogólnił lemat 3.1 następująco. Dla $k\geqslant 2$ definiujemy ciąg $\{\xi_n^{(k)}, n\geqslant 1\}$ zmiennych losowych przyjmujących wartości 0 i 1 wzorem

$$\xi_n^{(k)}=1_{\{n-k+1\leqslant \varrho_n\leqslant n\}}, \qquad n=1,2,\ldots \tag{3.1}$$

Z twierdzenia 3.1 wynika

LEMAT 3.2 (Niewzorow). *Jeżeli* $\{X_n, n\geqslant 1\}$ *jest ciągiem niezależnych zmiennych losowych o jednakowej ciągłej dystrybuancie, to dla ustalonego* $k\geqslant 2$ *zmienne losowe* $\xi_1^{(k)}, \xi_2^{(k)}, \ldots$ *są niezależne oraz*

$$P\{\xi_n^{(k)}=1\}=k/n, \qquad n=k, k+1, \ldots$$

Ponieważ dla k-tych czasów rekordowych mamy równość

$$L^{(k)}(n)=\min\{j>L^{(k)}(n-1): j-k+1\leqslant \varrho_j\leqslant j\}, \qquad n=2,3,\ldots,$$

więc zmienną losową

$$N_n^{(k)}=\sum_{j=k}^{n}\xi_j^{(k)}, \qquad n=k, k+1, \ldots,$$

można interpretować jako liczbę k-tych rekordów wśród pierwszych n $(n\geqslant k)$ zmiennych losowych. Podobnie jak w przypadku $k=1$, korzystając z lematu 3.2, można otrzymać różne twierdzenia graniczne dotyczące liczby k-tych rekordów $N_n^{(k)}$, k-tych czasów rekordowych $L^{(k)}(n)$ i k-tych czasów międzyrekordowych $\Delta^{(k)}(n)=L^{(k)}(n)-L^{(k)}(n-1)$ (zob. [24], a także [9], gdzie twierdzenia graniczne dowodzone są inną metodą).

Na zakończenie rozdziału omówimy problem istnienia skończonej wartości oczekiwanej k-tego czasu rekordowego $L^{(k)}(n)$. W rozdziale 2 pokazaliśmy, że wartość oczekiwana czasu rekordowego $L(n)=L^{(1)}(n)$ jest nieskończona:

$$E(L(n))=+\infty, \qquad n=1,2,\ldots$$

Niewzorow [23], [26] pokazał, korzystając z lematu 3.2, że dla ustalonego $r=1,\ldots,k-1$ ciąg

$$V_r(n) = \frac{(k-r)^{n-1}(k-r)!}{k^{n-1}k!} L^{(k)}(n)\left(L^{(k)}(n)-1\right)\cdot\ldots\cdot\left(L^{(k)}(n)-r+1\right),$$

$n = 1, 2, \ldots$, jest martyngałem oraz $E\left(V_r(n)\right) = 1$. Jako wniosek udowodnił między innymi równość

$$E\left(L^{(k)}(n)\right) = \frac{k^n}{(k-1)^n}, \qquad n = 1, 2, \ldots, \; k = 2, 3, \ldots$$

Zatem wartości oczekiwane k-tych czasów rekordowych $L^{(k)}(n)$ dla $k \geqslant 2$ są skończone; co więcej, wiadomo, że skończone są także ich momenty rzędu $k-1$.

4. Rekordy w sporcie. Model Yanga. Na ogół panuje przekonanie, że model, w którym bada się rekordy w ciągu niezależnych zmiennych losowych o jednakowym rozkładzie, jest wystarczający do opisu rekordowych zjawisk meteorologicznych, np. rekordowo dużych lub małych miesięcznych opadów, rekordowo dużej lub małej liczby godzin nasłonecznienia w danym miesiącu.

Glick [20] przytacza dane dotyczące opadów i godzin nasłonecznienia w danym miesiącu w Vancouver. W każdym przypadku podaje 65 obserwacji (opady w latach od 1900 do 1966, nasłonecznienie w latach od 1909 do 1973). Interesuje go liczba rekordów N_{65} wśród tych obserwacji. Otrzymane wyniki przedstawiamy w tabeli 1.

Tabela 1. Liczba rekordów w ciągu 65 obserwacji

	miesiąc												średnia
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	
opady:													
rekordowo duże	4	5	2	3	6	4	5	2	6	4	4	5	4,17
rekordowo małe	10	3	4	4	6	7	8	4	4	4	7	5	5,50
nasłonecznienie:													
rekordowo duże	4	5	3	3	4	5	6	4	1	7	5	4	4,25
rekordowo małe	4	6	7	5	7	5	3	5	4	6	4	7	5,25

Teoretyczna, wynikająca z założeń modelu, wartość oczekiwana liczby rekordów N_{65} wynosi

$$E(N_{65}) = 1 + \tfrac{1}{2} + \tfrac{1}{3} + \ldots + \tfrac{1}{65} = 4{,}76.$$

Widać więc pewną zgodność danych empirycznych z modelem, gdyż rekordy w 65 obserwacjach pojawiały się średnio co $\frac{1}{4}(4{,}17+5{,}50+4{,}25+5{,}25) = 4{,}7925$ lat.

Z rekordami w sporcie sytuacja przedstawia się inaczej. Już Foster i Stuart [17] zauważyli na podstawie danych dotyczących opadów atmosferycznych w Oxfordzie i wyników sportowych w Wielkiej Brytanii, że rekordowe wyniki w sporcie zdarzają się o wiele częściej niż rekordowe opady. Fakt ten łatwo wytłumaczyć. Podczas gdy fluktuacje pogody w obecnym stuleciu są losowe, bez żadnych wyraźnych trendów, sport rozwija się intensywnie, zmienia się liczba uprawiających go, a także udoskonalają się techniki treningu. Trudno więc przyjąć, że wyniki w sporcie są realizacjami ciągów niezależnych zmiennych losowych o jednakowym rozkładzie. Powstaje zatem potrzeba poszukiwania modeli matematycznych pozwalających opisać rekordy w sporcie.

Pierwszy taki model zaproponował Yang [39] w 1975 roku. Na podstawie wyników uzyskanych na igrzyskach olimpijskich wykazał on, że przyjęcie założenia jednakowości rozkładów zmiennych losowych $X_1, X_2, \ldots$ jest nieuzasadnione. Tłumaczy to faktem, że liczba ludności świata, a więc i sportowców, jest rosnąca. Proponuje zatem przyjąć, że

$$X_n = \max\{X_{n1}, X_{n2}, \ldots, X_{nm_n}\}, \tag{4.1}$$

gdzie $\{X_{nj}, j = 1, \ldots, m_n, n = 1, 2, \ldots\}$ jest tablicą niezależnych zmiennych losowych o jednakowej ciągłej dystrybuancie F, a

$$m_n = [\lambda^{n-1} m_1 + \tfrac{1}{2}], \quad \lambda \geqslant 1, m_1 > 0, n = 2, 3, \ldots$$

([a] oznacza tu część całkowitą liczby a). Liczby m_n, interpretowane jako rozmiar populacji, wzrastają geometrycznie, a m_1 jest początkowym rozmiarem populacji.

Yang badał między innymi rozkład graniczny czasu, jaki upływa między chwilami ustanawiania rekordów w danej dyscyplinie sportu.

TWIERDZENIE 4.1. *Jeżeli* $\{X_n, n \geqslant 1\}$ *jest ciągiem zmiennych losowych zdefiniowanych przez* (4.1), *to*

$$p_j = \lim_{n\to\infty} P\{\Delta(n) = j\} = (\lambda - 1)\lambda^{-j}, \quad j = 1, 2, \ldots \tag{4.2}$$

Yang sprawdza empirycznie, czy jego model jest zgodny ze stanem faktycznym. Szacuje parametr λ, stwierdzając, że ludność świata podwaja się co 36 lat, zatem po uwzględnieniu czteroletniego odstępu między igrzyskami otrzymuje $\lambda^9 = 2$, czyli $\lambda = 1{,}08$.

Rozkład graniczny $\Delta(n)$ jest wtedy postaci

j	=	1	2	3	4	5	...
p_j	=	0,074	0,069	0,064	0,059	0,054	...

Zbieżność w (4.2) dla $\lambda = 1{,}08$ jest szybka (błąd względny aproksymacji $P\{\Delta(n) = j\}$ przez p_j jest mniejszy niż 5% dla $n > 7$). Zatem rozkład graniczny $\{p_j, j \geqslant 1\}$ powinien dobrze opisywać czas między rekordami. Niestety, rekordy na igrzyskach olimpijskich padają częściej, niż wynikałoby to z modelu

Yanga. Na przykład, dla ostatnich 5 rekordów olimpijskich w roku 1975 średni czas, jaki upłynął między nimi, wynosi: 1 igrzyska dla biegu na 200 metrów, 2,4 igrzyska dla maratonu, 2,6 igrzyska dla skoku w dal. Yang sugeruje, że rozbieżności te są spowodowane albo nieuwzględnieniem rozwoju metod treningu, albo tym, że liczba sportowców wzrasta szybciej niż ludność świata.

Pomimo niezgodności z danymi empirycznymi model Yanga jest interesujący z punktu widzenia teorii matematycznej. Jest on intensywnie badany ([1], [13], [21], [22], [27]), przy czym rozważa się następujące jego uogólnienie.

Niech $\{X_n, n \geqslant 1\}$ będzie ciągiem niezależnych zmiennych losowych o dystrybuantach

$$F_n(x) = P\{X_n \leqslant x\} = F^{m_n}(x), \qquad n = 1, 2, \ldots,$$

gdzie F jest ciągłą dystrybuantą, a $\{m_n, n \geqslant 1\}$ jest (dowolnym) ciągiem dodatnich stałych.

W pracach Niewzorowa [21], [22], a także w pracy Alpuim [1], podany jest odpowiednik lematu Rényiego.

LEMAT 4.1. *Jeżeli $\{X_n, n \geqslant 1\}$ jest ciągiem niezależnych zmiennych losowych o dystrybuantach F^{m_n}, $n = 1, 2, \ldots$, gdzie F jest ciągłą dystrybuantą, a m_n, $n = 1, 2, \ldots$, są dodatnimi liczbami rzeczywistymi, to zmienne losowe*

$$\xi_n = 1_{\{X_n \text{ jest rekordem}\}}, \qquad n = 1, 2, \ldots,$$

są niezależne oraz

$$P\{\xi_n = 1\} = \frac{m_n}{m_1 + \ldots + m_n}, \qquad n = 1, 2, \ldots$$

Podobnie jak poprzednio, lemat 4.1 można wykorzystać do badania asymptotyki liczby rekordów N_n wśród pierwszych n zmiennych losowych oraz asymptotyki czasów rekordowych $L(n)$ i międzyrekordowych $\Delta(n)$.

Dziubdziela w pracy [13], korzystając z metody wyłączania i włączania (zob. np. [15], [19]), znalazł dla każdego $k \geqslant 1$ rozkłady zmiennych losowych $\xi_n^{(k)}$, $n = 1, 2, \ldots$, zdefiniowanych przez (3.1):

$$P\{\xi_n^{(k)} = 1\} = \sum_{s=n-k+1}^{n} \sum_{j=0}^{n-s} (-1)^j \binom{j+s-1}{s-1} \sum_{1 \leqslant i_1 < \ldots < i_{j+s-1}} \frac{m_n}{m_{i_1} + \ldots + m_{i_{j+s-1}} + m_n}.$$

W pracy Niewzorowa [27] udowodniony jest interesujący fakt, że niezależność zmiennych losowych $\xi_1, \xi_2, \ldots$ charakteryzuje, przy pewnych dodatkowych założeniach, ciąg dystrybuant $F^{m_1}, F^{m_2}, \ldots$ zmiennych losowych $X_1, X_2, \ldots$

TWIERDZENIE 4.2. *Niech $X_1, \ldots, X_n$ będą niezależnymi zmiennymi losowymi o dystrybuantach $F_1, \ldots, F_n$ i gęstościach $f_1, \ldots, f_n$ takich, że $\prod_{i=1}^{n-1} f_i(x) \neq 0$*

dla wszystkich x z pewnego przedziału (a, b) $(-\infty \leqslant a < b \leqslant \infty)$. Jeżeli wektor $\xi = (\xi_1, \ldots, \xi_{n-1})$ oraz zmienna losowa ξ_n są niezależne dla dowolnego wyboru gęstości f_n, to istnieją stałe $\alpha_2, \ldots, \alpha_{n-1}$ takie, że

$$F_k(x) = F_1^{\alpha_k}(x), \quad -\infty < x < \infty, \quad k = 2, \ldots, n-1,$$

oraz składowe $\xi_1, \ldots, \xi_{n-1}$ wektora ξ są niezależne.

W pracy [25] podano przykłady pokazujące, że zmienne losowe $\xi_1, \xi_2, \ldots$ nie muszą być niezależne, gdy $X_1, X_2, \ldots$ są niezależne.

5. Rekordy w ciągach z liniowym trendem. Ostatnio Resnick i Ballerini opublikowali cykl prac [2]–[5], w których zaproponowali nowy model dla rekordów w sporcie. Wyniki zilustrowali danymi dotyczącymi biegów na 400 i 800 metrów oraz na 1 milę. Rozważają oni rekordy w ciągu

$$X_n = Y_n + cn, \quad n = 1, 2, \ldots,$$

gdzie $\{Y_n, -\infty < n < \infty\}$ jest ściśle stacjonarnym ciągiem zmiennych losowych oraz $c > 0$. Liniowy trend cn, $n = 1, 2, \ldots$, ma między innymi odpowiadać wzrostowi liczby sportowców i rozwojowi metod treningu.

Definiujemy jak poprzednio

$$\xi_n = 1_{\{X_n \text{ jest rekordem}\}}, \quad n = 1, 2, \ldots$$

Wtedy

$$E(\xi_n) \downarrow p = P\Big\{\bigcap_{i=1}^{\infty} \{Y_{-1} - ci < Y_0\}\Big\}, \quad \text{gdy } n \to \infty,$$

a stąd wynika, że

$$\lim_{n\to\infty} E(N_n)/n = p.$$

Przypomnijmy, że $N_n = \sum_{j=1}^{n} \xi_j$ jest liczbą rekordów wśród pierwszych n zmiennych losowych.

Stałą p nazywa się *asymptotyczną intensywnością rekordów.*

Obliczanie p nie jest łatwe. W pracy [3] rozważano przypadek, gdy $\{Y_n\}$ jest ciągiem niezależnych zmiennych losowych o jednakowej dystrybuancie F; wtedy

$$p = p(c) = \int_{-\infty}^{\infty} G_\infty(y; c)\, F(dy),$$

gdzie

$$G_\infty(y; c) = \prod_{j=1}^{\infty} F(y + cj).$$

Wzór ten jest nieefektywny i zwykle p wyznacza się numerycznie, np. dla $c = 1$ i standardowego rozkładu normalnego mamy $p = 0{,}72506$. Tylko w nielicznych przypadkach p wyznaczono analitycznie; np. jeśli

$$F(x) = \exp\{-\exp(-(x-a)/b)\}, \quad -\infty < a < \infty, \ b > 0,$$

to

$$p(c) = 1 - \exp(-c/b).$$

Empirycznie wyznaczono p dla kilku dyscyplin sportowych.

Tabela 2. Intensywności rekordów z próby

dyscyplina	bieg na		
	400 m	800 m	1 milę
intensywność rekordu z próby	0,16964	0,19643	0,293

W pracach [3], [4] badano asymptotykę N_n, gdy $n \to \infty$. Między innymi udowodniono:

TWIERDZENIE 5.1. *Jeżeli ciąg* $\{Y_n, -\infty < n < \infty\}$ *jest ergodyczny, to*

$$P\{\lim_{n\to\infty} N_n/n = p\} = 1.$$

Oczywiście ciąg niezależnych zmiennych losowych o jednakowym rozkładzie jest ergodyczny.

Wnioskiem z twierdzenia 5.1 jest równość

$$P\{\lim_{n\to\infty} L(n)/n = 1/p\} = 1.$$

Badano też asymptotyczną normalność N_n:

TWIERDZENIE 5.2 ([3]). *Jeżeli* $\{Y_n, -\infty < n < \infty\}$ *jest ciągiem niezależnych zmiennych losowych o jednakowej dystrybuancie oraz* $c > 0$ *i* $E(\max\{0, X_1\})^2 < \infty$, *to*

$$\lim_{n\to\infty} P\{\sqrt{n}(n^{-1}N_n - p) < x\} = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{x} \exp\left(-\frac{t^2}{2\sigma^2}\right)dt, \qquad -\infty < x < \infty,$$

gdzie

$$\sigma^2 = p - p^2 + 2\sum_{m=1}^{\infty}(E(1_0^* 1_m^*) - p^2),$$

$$1_m^* = 1_{\{Y_m > \max(Y_{m-1}-c,\, Y_{m-2}-2c,\, Y_{m-3}-3c,\ldots)\}}, \qquad m = 1, 2, \ldots$$

W [4] udowodniono podobne twierdzenie w przypadku, gdy $\{Y_n\}$ jest ciągiem ściśle stacjonarnym i silnie mieszanym.

Wydaje się, że model Balleriniego i Resnicka lepiej opisuje rekordy w sporcie niż model Yanga. Problem między innymi polega na takim dobraniu założeń modelu – typu zależności w ciągu $\{Y_n\}$, dystrybuanty F

(np. przyjęcie założenia, że Y_n jest procesem Gaussa) i stałej c — aby teoretyczna intensywność rekordów była bliska empirycznej. Na przykład w pracy [4] Ballerini i Resnick dla biegu na 400 m tak dobrali założenia, że intensywność rekordów wynosi $p \approx 0{,}1523$. W tym przypadku intensywność rekordów z próby jest równa 0,1696.

6. Uwagi końcowe. W opracowaniu omówiliśmy jedynie wybrane problemy teorii rekordów. Nie wspomnieliśmy np. nic o związkach teorii rekordów z teorią punktowych procesów przekroczeń wysokich poziomów i procesów ekstremalnych; o tych zagadnieniach można przeczytać w pięknie napisanej monografii Resnicka [32].

Prawie pełny przegląd wyników teorii rekordów znajduje się w pracy Niewzorowa [28]. Praca ta zawiera także listę otwartych problemów w teorii rekordów. Oczywiście problemów nie rozwiązanych jest znacznie więcej. Można zatem przewidywać w najbliższych latach dalszy intensywny rozwój metod probabilistycznych i statystycznych teorii rekordów, a także spodziewać się nowych interesujących zastosowań.

Literatura cytowana

[1] Maria Teresa Alpuim, *Record values in populations with increasing or random dimension*, Metron 43 (1985), 145–155.

[2] R. Ballerini, *Another characterization of the type I extreme value distribution*, Statist. Probab. Lett. 5 (1987), 83–85.

[3] R. Ballerini, S. I. Resnick, *Records from improving populations*, J. Appl. Probab. 22 (1985), 487–502.

[4] —, —, *Records in the presence of a linear trend*, Adv. in Appl. Probab. 19 (1987), 801–828.

[5] —, —, *Embedding sequences of successive maxima in extremal processes, with applications*, J. Appl. Probab. 24 (1987), 827–837.

[6] O. Barndorff-Nielsen, *On the limit behavior of extreme order statistics*, Ann. Math. Statist. 34 (1963), 992–1002.

[7] K. N. Chandler, *The distribution and frequency of record values*, J. Roy. Statist. Soc. B14 (1952), 220–228.

[8] P. Deheuvels, *Strong approximation in extreme value theory and applications*, Colloq. Math. Soc. János Bolyai 36 (1982), 369–403.

[9] —, *On record times associated with k-th extremes*, w: Proceedings of the 3rd Pannonian Symposium on Mathematical Statistics, Visegrad 1982, 43–51.

[10] —, *The strong approximation of extremal processes II*, Z. Wahrsch. Verw. Gebiete 62 (1983), 7–15.

[11] M. Dwass, *Some k sample rank order tests*, w: Contributions to Probability and Statistics, Essays in Honour of H. Hotelling, 1960, 198–202.

[12] W. Dziubdziela, *Rozkłady graniczne ekstremalnych statystyk pozycyjnych*, Mat. Stos. 9 (1977), 45–71.

[13] —, *On the distribution of the number of k-th record values in increasing populations* (w przygotowaniu).

[14] W. Dziubdziela, B. Kopociński, *Limiting properties of the k-th record values*, Zastos. Mat. 15 (1976), 187–190.

[15] W. Feller, *Wstęp do rachunku prawdopodobieństwa*, t. I, PWN, Warszawa 1966.
[16] —, *Wstęp do rachunku prawdopodobieństwa*, t. II, PWN, Warszawa 1969.
[17] F. G. Foster, A. Stuart, *Distribution free tests in time-series based on the breaking of records*, J. Roy. Statist. Soc. B16 (1954), 1–13.
[18] F. G. Foster, D. Teichroew, *A sampling experiment on the powers of the records tests for trend in a time series*, ibidem B17 (1955), 115–121.
[19] J. Galambos, *The Asymptotic Theory of Extreme Order Statistics*, wyd. 1, New York 1978; wyd. 2 poprawione, Malabar, Florida, 1987.
[20] N. Glick, *Breaking records and breaking boards*, Amer. Math. Monthly 85 (1978), 2–26.
[21] W. B. Niewzorow (В. Б. Невзоров), *Рекордные моменты в случае неодинаково распределенных случайных величин*, Теория вероятн. и ее примен. 29 (1984), 808–809.
[22] —, *О рекордных моментах и межрекордных временах для последовательности неодинаково распределенных случайных величин*, Записки научн. семин. ЛОМИ 142 (1985), 109–118.
[23] —, *Рекордные моменты и их обобщения*, Теория вероятн. и ее примен. 31 (1986), 629–630.
[24] —, *О k-ых рекордных моментах и их обобщениях*, Записки научн. семин. ЛОМИ 153 (1986), 115–121.
[25] —, *О числе рекордов в последовательности неодинаково распределенных величин*, в кн.: *Вероятностные распределения и математическая статистика*, Ташкент 1986, 373–388.
[26] —, *Моменты некоторых случайных величин связанных с рекордами*, Вестник ЛГУ 2 (1987), 33–37.
[27] — (V. B. Nevzorov), *Two characterizations using records*, w: *Stability Problems for Stochastic Models*, Proceedings of the 9th International Seminar held in Varna, Bulgaria, May 1985, Lecture Notes in Math. 1233, Springer, 1987, 79–85.
[28] —, *Рекорды*, Теория вероятн. и ее примен. 32 (1987), 219–251.
[29] A. Rényi, *On the extreme elements of observations*, w: *Selected Papers of Alfred Rényi*, tom 3, Budapest 1976 (tłumaczenie na język angielski artykułu: *Egy megfigyeléssorozat kiemelkedö elemeiröl*, Mag. Tud. Acad. III Oszt. Közl. 12 (1962), 105–121).
[30] —, *Théorie des-éléments saillants d'une suite d'observations*, w: Colloquium on Combinatorial Methods in Probability Theory, Aarhus 1962, 104–117.
[31] —, *Théorie des éléments saillants d'une suite d'observations*, Ann. Fac. Sci. Univ. Clermont-Ferrand 2 (1962), 7–12.
[32] S. I. Resnick, *Extreme Values, Regular Variation and Point Processes*, New York 1987.
[33] R. W. Shorrock, *A limit theorem for inter-record times*, J. Appl. Probab. 9 (1972), 219–223.
[34] —, *On record values and record times*, ibidem 9 (1972), 316–326.
[35] —, *Record values and inter-record times*, ibidem 10 (1973), 543–555.
[36] A. Stuart, *The efficiencies of tests of randomness against normal regression*, J. Amer. Statist. Assoc. 51 (1956), 285–287.
[37] —, *The efficiency of the records test for trend in normal regression*, J. Roy. Statist. Soc. B19 (1957), 149–153.
[38] M. N. Tata, *On outstanding values in a sequence of random variables*, Z. Wahrsch. Verw. Gebiete 12 (1969), 9–20.
[39] M. C. K. Yang, *On the distribution of the inter-record times in an increasing population*, J. Appl. Probab. 12 (1975), 148–154.