AUTOREFERAT

IMIĘ I NAZWISKO: Anna Dembińska

DYPLOMY I STOPNIE NAUKOWE:

- tytuł magistra inżyniera uzyskany w 1995 roku po ukończeniu studiów na kierunku matematyka na Wydziale Fizyki Technicznej i Matematyki Stosowanej Politechniki Warszawskiej;
- stopień naukowy doktora nauk matematycznych w zakresie matematyki uzyskany w 2002 roku na Wydziale Matematyki i Nauk Informacyjnych Politechniki Warszawskiej; tytuł rozprawy doktorskiej "Charakteryzacje rozkładów prawdopodobieństwa związane z własnościami regresyjnymi statystyk porządkowych i rekordowych".

HISTORIA ZATRUDNIENIA:

1995-2002 asystent
Wydział Fizyki Technicznej i Matematyki Stosowanej
(w 1999 przekształcony w Wydział Matematyki i Nauk Informacyjnych)
Politechnika Warszawska

od 2002 adiunkt
Wydział Matematyki i Nauk Informacyjnych
Politechnika Warszawska

ROZPRAWA HABILITACYJNA:

Własności asymptotyczne liczby obserwacji w otoczeniach statystyk porządkowych

Rozprawa obejmuje jednotematyczny cykl ośmiu następujących prac:

R1. A. Dembińska, A. Stepanov, J. Wesołowski, How many observations fall in a neighborhood of an order statistic?, *Comm. Statist. – Theory & Methods* 36 (2007), 851–867.

R2. A. Dembińska, N. Balakrishnan, The asymptotic distribution of numbers of observations near order statistics, *J. Statist. Plann. Inf.* 138 (2008), 2552–2562.

R3. A. Dembińska, N. Balakrishnan, On the asymptotic independence of numbers of observations near order statistics, *Statistics* 44 (2010), 517–528.

R4. A. Dembińska, On numbers of observations near randomly indexed order statistics, *Statist. Probab. Lett.* 80 (2010), 309–317.

R5. G. Iliopoulos, A. Dembińska, N. Balakrishnan, Asymptotic properties of numbers of observations near sample quantiles, *Statistics* 46 (2012), 85–97.

R6. A. Dembińska, Asymptotic properties of numbers of observations in random regions determined by central order statistics, *J. Statist. Plann. Inf.* 142 (2012), 516–528.

R7. A. Dembińska, G. Iliopoulos, On the asymptotics of numbers of observations in random regions determined by order statistics, *J. Multivariate Anal.* 103 (2012), 151–160.

R8. A. Dembińska, Limit theorems for proportions of observations falling into random regions determined by order statistics, *Aust. N. Z. J. Stat.* (2012) DOI: 10.1111/j.1467-842X.2012.00667.x

OMÓWIENIE ROZPRAWY

1 Wstęp

Niech $(X_n, n \geq 1)$ będzie ciągiem zmiennych losowych o tym samym rozkładzie z ciągłą dystrybuantą F i niech $X_{1:n} \leq \ldots \leq X_{n:n}$ to statystyki porządkowe pochodzące z próby losowej $(X_1, \ldots, X_n)$. Dla $a > 0$ i $k \in \{1, \ldots, n\}$ definiujemy

$$K_-(n,k,a) := \#\{j \in \{1, \cdots, n\} : \ X_j \in (X_{k:n} - a, X_{k:n})\}$$

oraz

$$K_+(n,k,a) := \#\{j \in \{1, \cdots, n\} : \ X_j \in (X_{k:n}, X_{k:n} + a)\}.$$

$K_-(n,k,a)$ i $K_+(n,k,a)$ zliczają liczby obserwacji, które wpadły do, odpowiednio, lewostronnego i prawostronnego otoczenia k-tej statystyki porządkowej. Od momentu ukazania się pierwszej pracy [19] na ich temat, stały się one przedmiotem systematycznego zainteresowania badaczy. Rozważano także różne ich uogólnienia i modyfikacje, jak, na przykład, liczby obserwacji w otoczeniu maksimum zaobserwowane od momentu pojawienia się tego maksimum (zob. Khmaladze, Nadareishvili i Nikabadze [15]), liczby obserwacji w otoczeniu dwuwymiarowego maksimum (zob. Bairamov i Stepanov [1]), liczby obserwacji w otoczeniu progresywnie cenzurowanej statystyki porządkowej (zob. Balakrishnan and Stepanov [5]) czy liczby obserwacji w otoczeniu statystyki rekordowej (zob. Balakrishnan, Pakes i Stepanov [3], Gouet, López i Sanz [10] oraz Bairamov i Stepanov [2]).

Zmienne losowe $K_-(n,k,a)$ i $K_+(n,k,a)$ pojawiają się w różnorodnych problemach praktycznych. Na przykład znajdują zastosowanie w matematyce finansowej: Li i Pakes [16] oraz Hashorva [12] opisali model ubezpieczeniowy, w którym $K_-(n,n,a)$ to liczba roszczeń w pewnym portfelu, których wartości różnią się od wysokości maksymalnego roszczenia co najwyżej o a. Wtedy własności asyptotyczne zmiennej losowej $K_-(n,n,a)$ przy $n \to \infty$ dostarczają informacji o długoterminowym zachowaniu się tego modelu. Ponadto $K_-(n,k,a)$ i $K_+(n,k,a)$ posłużyły do konstrukcji estymatorów różnych wielkości opisujących dystrubuantę F - zob. Hashorva [12, 13], Hashorva i Hüsler [14] oraz Iliopoulos, Dembińska i Balakrishnan [R5]. Na inne zastosowanie zmiennych losowych $K_-(n,k,a)$ i $K_+(n,k,a)$ zwrócili uwagę Pakes i Steutel [19] oraz Dembińska, Stepanov i Wesołowski [R1]. Zauważyli oni, że rozkład rozstępu statystyk porządkowych można wyrazić za pomocą rozkładu zmiennej losowej $K_-(n,k,a)$ lub $K_+(n,k,a)$, co pozwala na wyprowadzenie własności asymptotycznych rozstępu z rozkładu granicznego dla $K_-(n,k,a)$ lub $K_+(n,k,a)$.

Wspólnym celem prac zawartych w rozprawie habilitacyjnej jest opisanie własności granicznych zmiennych losowych $K_-(n,k,a)$ i $K_+(n,k,a)$ oraz ich uogólnień, gdy $n \to \infty$.

W pracy [R1], badając asymptotyczne zachowanie się zmiennych losowych $K_-(n,k,a)$ i $K_+(n,k,a)$, rozpatrzono sytuację, w której $k = k_n$ zmienia się wraz z n w taki sposób, że $k_n/n \to \lambda \in [0,1]$, a także rozważono przypadek gdy $a = a_n$ zależy od n. Otrzymano twierdzenia graniczne dla zmiennych losowych $K_-(n,k_n,a_n)$ i $K_+(n,k_n,a_n)$, a w szczególności wyznaczono ciągi $\{a_n, n \geq 1\}$, dla których rozkład graniczny badanych zmiennych losowych jest rozkładem Poissona lub jego modyfikacją.

W artykułach [R2] i [R3] opisano asymptotyczne, łączne zachowanie się zmiennych losowych $K_-(n,k,a), K_+(n,k,b), K_-(n,n-r,c), K_+(n,n-r,d)$, gdy $n \to \infty$ a k i r są ustalone. Podano czterowymiarowy rozkład graniczny tych zmiennych losowych i pokazano, że przy pewnych założeniach są one asymptotycznie niezależne. W pracy [R3] rozpatrzono też asymptotyczne zachowanie się liczby obserwacji w otoczeniach centralnych statystyk porządkowych. Pokazano, że jeśli $\lim_{n\to\infty} k_n^{(i)}/n = \lambda_i \in (0,1)$, $i = 1,\ldots,s$, i wszystkie λ_i są różne, to, dla odpowiednio dobranych ciągów $\{a_n^{(i)}, n \geq 1\}$ i $\{b_n^{(i)}, n \geq 1\}$, $i = 1,\ldots,s$, zmienne losowe $K_-(n,k_n^{(1)},a_n^{(1)})$, $K_+(n,k_n^{(1)},b_n^{(1)})$, $\cdots$, $K_-(n,k_n^{(s)},a_n^{(s)})$, $K_+(n,k_n^{(s)},b_n^{(s)})$ są asymptotycznie niezależne. W [R4] uogólniono te wyniki na przypadek prób o losowej liczności.

W artykule [R5] pokazano, że częstości tych obserwacji, które wpadły do lewostronnego lub prawostronnego otoczenia centralnej statystyki porządkowej, czyli $K_-(n,k_n,a)/n$ i $K_+(n,k_n,b)/n$, gdzie $k_n/n \to \lambda \in (0,1)$, zbiegają według prawdopodobieństwa do pewnych wielkości nielosowych. Udowodniono także, że częstości te, po odpowiedniej standaryzacji, mają asymptotycznie łączny rozkład normalny jeśli tylko spełnione są pewne standardowe założenia. Następnie wynik ten uogólniono na przypadek częstości obserwacji, które wpadły do otoczeń dwóch lub większej liczby centralnych statystyk porządkowych.

W pracy [R6], rozszerzając pojęcia $K_-(n,k,a)$ i $K_+(n,k,b)$, wprowadzono zmienną losową $K(n,k,A)$ zliczającą liczbę obserwacji, które wpadły do zbioru losowego wyznaczonego przez k-tą statystykę porządkową i zbiór borelowski A:

$$K(n,k,A) := \#\{j \in \{1,\cdots,n\} : X_{k:n} - X_j \in A\}.$$

Dla $a > 0$ i $b > 0$ mamy $K(n,k,(0,a)) = K_-(n,k,a)$ i $K(n,k,(-b,0)) = K_+(n,k,b)$. Ponadto jeśli $k = n$ i $A = \{0\}$, to zmienna losowa $K(n,k,A) = K(n,n,\{0\})$ zlicza liczbę elementów próby o liczności n, które są równe obserwacji maksymalnej. Jeśli ponadto dystrybuanta F jest dyskretna o nośniku złożonym z liczb całkowitych nieujemnych, to $K(n,n,\{0\})$ może być interpretowana jako liczba zwycięzców w grze z n uczestnikami, których wygrane to $X_1, X_2, \ldots, X_n$. Odwołując się do tej interpretacji, własności asymptotyczne zmiennej losowej $K(n,n,\{0\})$, gdy $n \to \infty$, opisywali, między innymi: Eisenberg, Stengle i Strang [9], Brands, Steutel i Wilms [6], Qi [20], Bruss i Grübel [7], Eisenberg [8] oraz Gouet, López i Sanz [11]. Przy założeniu, że X_n, $n \geq 1$, są niezależne i mają ten sam rozkład skoncentrowany na nieujemnych liczbach całkowitych, autorzy ci badali istnienie granicy prawdopodobieństwa, że będzie dokładnie jeden zwycięzca, czyli istnienie $\lim_{n\to\infty} P(K(n,n,\{0\}) = 1)$ oraz, związane z tym zagadnieniem, problemy pokrewne.

W [R6] przedstawiono twierdzenia opisujące asymptotyczne zachowanie zmiennej losowej $K(n,k_n,A)$, uogólniając wyniki podane w [R5]. Twierdzenia te dotyczą sytuacji, gdy $k_n/n \to \lambda \in (0,1)$. Natomiast przypadkowi, gdy k_n albo $n - k_n$ nie zależy od n, poświęcona jest praca [R7]. Wyznaczono w niej wielowymiarowe rozkłady graniczne dla zmiennych losowych postaci $K(n,k,A)$ i $K(n,n-r,A)$, gdzie k i r są stałe.

Wszystkie opisane do tej pory wyniki były otrzymane przy założeniu, że dystrybuanta F jest ciągła albo dyskretna a obserwacje wchodzące w skład próbki są niezależne. W [R8], dowodząc zbieżności prawie na pewno dla częstości $K(n,k_n,A)/n$, osłabiono te restrykcyjne warunki. Przedstawiono twierdzenia zachodzące dla prób o losowej liczności, w których obserwacje, o niekoniecznie ciągłej dystrubuancie F, nie muszą być niezależne. Twierdzenia te dotyczą wszystkich trzech przypadków analizowanych w literaturze podczas badania granicznego zachowania się statystyk porządkowych: przypadku *centralnego* (odpowiadającego sytuacji gdy $k_n/n \to \lambda \in (0,1)$), przypadku *skrajnego* (gdzie k_n albo $n - k_n$ nie zależą od n) oraz przypadku *asymptotycznie skrajnego* (w którym $k_n/n \to \lambda \in \{0,1\}$, $k_n \to \infty$ oraz $n - k_n \to \infty$).

2 Obserwacje w otoczeniach statystyk porządkowych

Pierwsze wyniki dotyczące własności asymptotycznych liczby obserwacji, które wpadły do pewnego otoczenia statystyki porządkowej, dotyczyły przypadku, w którym statystyka porządkowa jest skrajna, por. na przykład [19], [17], [16] i [4]. W pracy [R1] po raz pierwszy podjęto badania asymptotyki w przypadku otoczeń statystyk centralnych i asymptotycznie skrajnych. W szczególności, dla przypadku centralnego wykazano, że jeśli $\lambda := \lim_{n\to\infty} k_n/n \in (0,1)$ oraz F jest funkcją ściśle rosnącą w pewnym otoczeniu kwantyla teoretycznego rzędu λ, to dla dowolnego $a > 0$ mamy

$$K_-(n, k_n, a) \xrightarrow{P} \infty \quad \text{i} \quad K_+(n, k_n, a) \xrightarrow{P} \infty \quad \text{gdy} \quad n \to \infty, \tag{1}$$

gdzie $\xrightarrow{P}$ oznacza zbieżność według prawdopodobieństwa. Wynika stąd, że aby dla zmiennych losowych $K_-(n, k_n, a)$ i $K_+(n, k_n, a)$ w granicy otrzymać rozkład niezdegenerowany, trzeba wraz ze wzrostem n zmniejszać a.

Twierdzenie 1 (R1, Theorems 5.1 i 5.2)**.**
Załóżmy, że $\lambda \in (0,1)$ i istnieje dokładnie jedna γ taka, że $F(\gamma) = \lambda$. Niech, dla pewnego $\alpha > 0$,

$$a_n = \gamma - F^{\leftarrow}(\lambda - \alpha/n),\ n = 1, 2, \ldots,$$

oraz

$$\lim_{(x,y)\to(\gamma,0^+)} \frac{F(x) - F(x-y)}{F(\gamma) - F(\gamma - y)} = 1. \tag{2}$$

Wówczas

$$K_-(n, k_n, a_n) \xrightarrow{d} \mathcal{P}(\alpha),$$

gdzie $F^{\leftarrow}(x) := \inf\{s : F(s) \geq x\}$, $x \in (0,1)$, $\xrightarrow{d}$ oznacza zbieżność według rozkładu a $\mathcal{P}(\alpha)$ to rozkład Poissona o średniej α. Ponadto jeśli dla pewnego $\alpha > 0$

$$b_n = F^{\leftarrow}(\lambda + \alpha/n) - \gamma,\ n = 1, 2, \ldots,$$

oraz

$$\lim_{(x,y)\to(\gamma,0^+)} \frac{F(x+y) - F(x)}{F(\gamma+y) - F(\gamma)} = 1, \tag{3}$$

to

$$K_+(n, k_n, b_n) \xrightarrow{d} \mathcal{P}(\alpha).$$

W [R1] przedstawiono także odpowiedniki powyższego twierdzenia dla przypadku skrajnego i asymptotycznie skrajnego. Zabrakło jednak analogicznego wyniku dla zmiennej losowej $K_-(n, k_n, b_n)$, gdy $\lambda = 1$, $n - k_n$ nie jest stałe i $r_F := \sup\{x \in \mathbb{R} : F(x) < 1\} = \infty$. Problem ten rozwiązano w [R2].

Twierdzenie 2 (R2, Theorems 2.1 i 2.2)**.**
Niech $r_F = \infty$, $\lambda = 1$, $n - k_n \to \infty$ i $b > 0$ będzie ustaloną liczbą taką, że

$$\beta(b) := \lim_{x\to\infty} \frac{\overline{F}(x)}{\overline{F}(x-b)} = \lim_{x\to\infty} \frac{\overline{F}(x+b)}{\overline{F}(x)} \in (0,1). \tag{4}$$

Wówczas

$$K_-(n, k_n, b) \xrightarrow{P} \infty.$$

Ponadto jeśli $\{b_n, n \geq 1\}$ jest ciągiem liczb dodatnich spełniającym warunek

$$\lim_{x\to\infty} (n - k_n) \left[\frac{\overline{F}(x - b_n)}{\overline{F}(x)} - 1 \right] = \alpha \ \ \textit{jednostajnie względem } n,$$

to

$$K_-(n, k_n, b_n) \xrightarrow{d} \mathcal{P}(\alpha).$$

W pracy [R2] wyznaczono także dwuwymiarowe rozkłady graniczne dla następujących par zmiennych losowych: $\big(K_+(n,k,a), K_-(n,n-r,b)\big)$, $\big(K_-(n,k,a),\ K_+(n,n-r,b)\big)$, $\big(K_-(n,k,a), K_-(n,n-r,b)\big)$ i $\big(K_+(n,k,a), K_+(n,n-r,b)\big)$, gdy $n\to\infty$ a k i r się nie zmieniają. Wynik ten został uogólniony w [R3], gdzie podano czterowymiarowy rozkład graniczny wektora losowego $\big(K_-(n,k,a), K_+(n,k,b), K_-(n,n-r,c), K_+(n,n-r,d)\big)$ i pokazano, że przy pewnych założeniach składowe tego wektora są asymptotycznie niezależne.

Twierdzenie 3 (R3, Section 2)**.**
Niech $l_F := \inf\{x\in\mathbb{R}:\ F(x)>0\} = -\infty$, $r_F=\infty$ *i* a,b,c,d *będą ustalonymi liczbami dodatnimi takimi, że* $\overline{\beta}(a)$, $\overline{\beta}(b)$, $\beta(c)$ *oraz* $\beta(d)$ *istnieją i wszystkie należą do* $(0,1)$, *gdzie* $\beta(\cdot)$ *została zdefiniowana w (4) natomiast*

$$\overline{\beta}(y) \ := \ \lim_{x\to-\infty}\frac{F(x)}{F(x+y)} = \lim_{x\to-\infty}\frac{F(x-y)}{F(x)}. \tag{5}$$

Wówczas dla dowolnych liczb całkowitych $i\ge 0$, $m\ge 0$, $j=0,1,\dots,k-1$ *i* $l=0,1,\dots,r$, *mamy*

$$\begin{aligned}
&\lim_{n\to\infty} P(K_-(n,k,a)=j, K_+(n,k,b)=i, K_-(n,n-r,c)=m, K_+(n,n-r,d)=l)\\
&\quad=\binom{k-1}{j}[\overline{\beta}(a)]^{k-1-j}[1-\overline{\beta}(a)]^j\binom{i+k-1}{i}[\overline{\beta}(b)]^k[1-\overline{\beta}(b)]^i\\
&\qquad\times\binom{m+r}{m}[\beta(c)]^{r+1}[1-\beta(c)]^m\binom{r}{l}[\beta(d)]^{r-l}[1-\beta(d)]^l,
\end{aligned}$$

co oznacza, że zmienne losowe $K_-(n,k,a)$, $K_+(n,k,b)$, $K_-(n,n-r,c)$ *i* $K_+(n,n-r,d)$ *są asymptotycznie niezależne.*

W drugiej części artykułu [R3] przedstawiono uogólnienie twierdzenia 1 na przypadek wielowymiarowy.

Twierdzenie 4 (R3, Theorem 3.1)**.**
Niech $\{k_n^{(i)}, n\ge 1\}$, $i=1,\dots,s$, *to ciągi liczb naturalnych takie, że* $\lim_{n\to\infty} k_n^{(i)}/n=\lambda_i\in(0,1)$, $i=1,\dots,s$, *gdzie wszystkie* λ_i *są różne. Załóżmy, że dla* $i=1,\dots,s$ *istnieją* γ_i *takie, że* $F(\gamma_i)=\lambda_i$ *oraz* F *jest funkcją ściśle rosnącą w pewnym otoczeniu każdej* γ_i. *Jeśli warunki (2) i (3) są spełnione z* $\gamma=\gamma_i$, $i=1,\dots,s$, *oraz dla ustalonych* $\alpha_{ij}>0$, $i=1,\dots,s$, $j=1,2$, *mamy*

$$a_n^{(i)}=\gamma_i-F^{\leftarrow}(\lambda_i-\alpha_{1i}/n),\quad b_n^{(i)}=F^{\leftarrow}(\lambda_i+\alpha_{2i}/n)-\gamma_i,\quad n=1,2,\dots,$$

to zmienne losowe $K_-(n,k_n^{(1)},a_n^{(1)})$, $K_+(n,k_n^{(1)},b_n^{(1)})$, …, $K_-(n,k_n^{(s)},a_n^{(s)})$, $K_+(n,k_n^{(s)},b_n^{(s)})$ *są asymptotycznie niezależne i*

$$K_-(n,k_n^{(i)},a_n^{(i)})\xrightarrow{d}\mathcal{P}(\lambda_{1i}),\quad K_+(n,k_n^{(i)},b_n^{(i)})\xrightarrow{d}\mathcal{P}(\lambda_{2i}),\quad i=1,\dots,s.$$

Wyniki przedstawione w pracach [R1] - [R3] dowodzono szacując całki wyrażające prawdopodobieństwo tego, że $K_-(n, k_n, a_n) = j$ i $K_+(n, k_n, b_n) = m$, gdzie j i m to ustalone liczby naturalne. Takie szacowania często wymagały żmudnych rachunków, które komplikowały się tym bardziej im ogólniejszy bądź głębszy wynik chciano uzyskać. W pracy [R5] zastosowano inne podejście: wykorzystano fakt, że łączny rozkład warunkowy zmiennych losowych $K_-(n, k_n, a)$ i $K_+(n, k_n, b)$ pod warunkiem $X_{k_n:n}$ to rozkład wielomianowy a następnie użyto własności rozkładu wielomianowego, jak, na przykład, centralnego twierdzenia granicznego dla tego rozkładu. Kolejnym krokiem było odtworzenie asymptotycznego rozkładu niewarunkowego z własności asymptotycznych rozkładu warunkowego i zmiennej losowej w warunku. Postępując w ten sposób, w [R5] opisano tempo zbieżności danej w (1), pokazując, że gdy $n \to \infty$, to

$$K_-(n, k_n, a)/n \xrightarrow{P} P(\gamma - a < X_1 \leq \gamma) \quad \text{i} \quad K_+(n, k_n, a)/n \xrightarrow{P} P(\gamma < X_1 \leq \gamma + a),$$

gdzie γ to jedyne rozwiązanie równania $F(\gamma) = \lambda$. Ponadto udowodniono, że, przy pewnych założeniach, zmienne losowe $K_-(n, k_n, a)$ i $K_+(n, k_n, a)$, po odpowiedniej standaryzacji, mają asymptotycznie dwuwymiarowy rozkład normalny (zob. twierdzenie 5). Następnie wynik ten uogólniono na przypadek liczby obserwacji w otoczeniach dwóch lub więcej cenatralnych statystyk porządkowych.

Twierdzenie 5 (R5, Theorem 2)**.**
Niech $k_n/n = \lambda + o(n^{-1/2})$, gdzie $\lambda \in (0, 1)$, F będzie funkcją różniczkowalną w punkcie γ oraz $F'(\gamma) \neq 0$. Ponadto załóżmy, że dla dystrybuanty F istnieje funkcja gęstości f i gęstość ta ma ciągłą pochodną w pewnym otoczeniu γ. Wówczas rozkład graniczny wektora losowego

$$\sqrt{n}\big(K_-(n, k_n, a)/n - P(\gamma - a < X_1 \leq \gamma), K_+(n, k_n, b)/n - P(\gamma < X_1 \leq \gamma + a)\big)$$

jest rozkładem dwuwymiarowym normalnym o średniej zero i macierzy kowariancji postaci

$$\begin{pmatrix} \lambda p(\gamma, a)\overline{p}(\gamma, a) + \dfrac{\lambda^3(1-\lambda)[p'(\gamma, a)]^2}{f^2(\gamma)} & \dfrac{\lambda^2(1-\lambda)^2 p'(\gamma, a) q'(\gamma, b)}{f^2(\gamma)} \\ \dfrac{\lambda^2(1-\lambda)^2 p'(\gamma, a) q'(\gamma, b)}{f^2(\gamma)} & (1-\lambda)q(\gamma, b)\overline{q}(\gamma, b) + \dfrac{\lambda(1-\lambda)^3[q'(\gamma, b)]^2}{f^2(\gamma)} \end{pmatrix},$$

gdzie

$$\overline{p}(x, a) := \frac{F(x-a)}{F(x)}, p(x, a) := 1 - \overline{p}(x, a), \overline{q}(x, b) := \frac{1 - F(x+b)}{1 - F(x)}, q(x, b) := 1 - \overline{q}(x, b)$$

i pochodne funkcji p oraz q są wyznaczone względem pierwszej zmiennej.

Warto zwrócić uwagę na fakt, że wyniki podane w pracy [R5] mają zastosowanie statystyczne - wynika z nich, że proporcje $K_-(n, k_n, a)/n$ i $K_+(n, k_n, b)/n$ są zgodnymi i asymptotycznie normalnymi estymatorami, odpowiednio, wielkości $F(\gamma) - F(\gamma - a)$ i $F(\gamma + b) - F(\gamma)$. Zatem w przypadku, gdy dystrybuanta F nie jest znana, mogą one posłużyć do estymacji prawdopodobieństwa przedziałów wyznaczonych przez nieznany kwantyl teoretyczny ustalonego rzędu.

3 Próby o losowej liczności

Zajmijmy się teraz przypadkiem, w którym obserwacje tworzą próbę o losowej liczności postaci $X_1, X_2, \ldots, X_{N(t)}$, gdzie $(N(t); t \geq 0)$ jest procesem liczącym określonym na $[0, \infty)$. Dla $a > 0$ zdefiniujmy $\mathcal{K}_-(t, k, a)$ i $\mathcal{K}_+(t, k, a)$, $t > 0$, jako procesy stochastyczne, takie, że dla każdego $t > 0$ mamy $\mathcal{K}_-(t, k, a) := K_-(N(t), k, a)$ i $\mathcal{K}_+(t, k, a) := K_+(N(t), k, a)$. Ponadto przyjmujemy, że $K_-(n, k, a) = 0$ i $K_+(n, k, a) = 0$ jeśli tylko $k > n$.

Próby o losowej liczności pojawiają się w różnych zagadnieniach praktycznych. Jednym z takich zagadnień jest modelowanie wysokości roszczeń wypłacanych w ramach pewnego portfela, w którym X_i to wysokość i-tej straty, natomiast $N(t)$ to liczba strat zgłoszonych do chwili t. Wówczas, na przykład, $X_{N(t):N(t)}$ jest największym roszczeniem zgłoszonym do chwili t zaś $\mathcal{K}_-(t, N(t), a)$ - liczbą roszczeń zgłoszonych do chwili t, których wyskość jest bliska roszczeniu największemu.

Kluczową rolę w dowodach wyników dotyczących prób o losowej liczności odgrywa podane poniżej twierdzenie 6. Pozwala ono wyprowadzać wielowymiarowe rozkłady graniczne dla liczby obserwacji w otoczeniach statystyk porządkowych z próby o losowej liczności, bezpośrednio z wielowymiarowych rozkładów granicznych odpowiadających przypadkowi próby o liczności deterministycznej.

Twierdzenie 6.
Niech $(\mathbb{Y}_n = (Y_n^{(1)}, \cdots, Y_n^{(r)}), n \geq 1)$ *będzie ciągiem r-wymiarowych wektorów losowych i niech* $(N(t), t \geq 0)$ *to proces stochastyczny złożony ze zmiennych losowych przyjmujących wartości naturalne. Załóżmy, że* $(\mathbb{Y}_n, n \geq 1)$ *i* $(N(t), t \geq 0)$ *są niezależne,* $N(t) \xrightarrow{P} \infty$, *gdy* $t \to \infty$ *i* $\mathbb{Y}_n \xrightarrow{d} \mathbb{Y} = (Y_1, \cdots, Y_r)$, *gdy* $n \to \infty$. *Wówczas* $\mathbb{Y}_{N(t)} \xrightarrow{d} \mathbb{Y}$, *gdy* $t \to \infty$.

Na przykład, korzystając z twierdzeń 3 i 6, otrzymujemy graniczny, łączny rozkład liczby obserwacji w otoczeniach $X_{k:N(t)}$ oraz $X_{N(t)-r:N(t)}$, czyli w otoczeniach skrajnych statystyk porządkowych z próby o losowej liczności.

Twierdzenie 7 (R4, Theorem 2.2)**.**
Niech $(N(t), t \geq 0)$ *będzie procesem liczącym, niezależnym od* $(X_n, n \geq 1)$. *Załóżmy, że* $N(t) \xrightarrow{P} \infty$, *gdy* $t \to \infty$ *i spełnione są założenia twierdzenia 3. Wówczas wektory losowe*

$$\big(\mathcal{K}_-(t, k, a), \mathcal{K}_+(t, k, b), \mathcal{K}_-(t, N(t) - r, c), \mathcal{K}_+(t, N(t) - r, d)\big)$$

oraz

$$\big(K_-(n, k, a), K_+(n, k, b), K_-(n, n - r, c), K_+(n, n - r, d)\big)$$

mają te same rozkłady graniczne przy, odpowiednio, $t \to \infty$ *i* $n \to \infty$.

Natomiast, używając twierdzeń 4 i 6, otrzymamy warunki przy których liczby obserwacji, które wpadły do lewostronnych i prawostronnych otoczeń centralnych statystyk porządkowych w próbie o losowej liczności, są asymptotycznie niezależne.

Twierdzenie 8 (R4, Theorem 3.3).
Załóżmy, że spełnione są założenia twierdzenia 4 oraz $(N(t), t \geq 0)$ *jest procesem liczącym, niezależnym od* $(X_n, n \geq 1)$ *i takim, że* $N(t) \xrightarrow{P} \infty$*, gdy* $t \to \infty$*. Wówczas zmienne losowe*

$$\mathcal{K}_-(t, k^{(1)}_{N(t)}, a^{(1)}_{N(t)}), \mathcal{K}_+(t, k^{(1)}_{N(t)}, b^{(1)}_{N(t)}), \ldots, \mathcal{K}_-(t, k^{(s)}_{N(t)}, a^{(s)}_{N(t)}), \mathcal{K}_+(t, k^{(s)}_{N(t)}, b^{(s)}_{N(t)})$$

są asymptotycznie niezależne i

$$\mathcal{K}_-(t, k^{(i)}_{N(t)}, a^{(i)}_{N(t)}) \xrightarrow{d} \mathcal{P}(\lambda_{1i}), \quad \mathcal{K}_+(t, k^{(i)}_{N(t)}, b^{(i)}_{N(t)}) \xrightarrow{d} \mathcal{P}(\lambda_{2i}), \quad i = 1, \cdots, s, \quad \text{gdy } t \to \infty.$$

W pracy [R4] pokazano także jak zmieni się rozkład graniczny podany w twierdzeniu 8, gdy losowe ciągi $(a_{N(t)}, t \geq 0)$ i $(b_{N(t)}, t \geq 0)$ zamienimy na nielosowe $(a_n, n \geq 1)$ i $(b_n, n \geq 1)$.

Twierdzenie 9 (R4, Theorem 3.4).
Załóżmy, że spełnione są założenia twierdzenia 4 oraz $(N(t), t \geq 0)$ *jest procesem liczącym, niezależnym od* $(X_n, n \geq 1)$ *i takim, że*

$$\frac{N(t)}{t} \xrightarrow{P} Z, \quad \text{gdy} \quad t \to \infty, \quad \text{gdzie } Z \text{ to nieujemna zmienna losowa.}$$

Wówczas

$$\lim_{n\to\infty} P\left(\mathcal{K}_-(n, k^{(i)}_{N(n)}, a^{(i)}_n) = j_i, \mathcal{K}_+(n, k^{(i)}_{N(n)}, b^{(i)}_n) = m_i, i = 1, \cdots, s\right)$$
$$= E\left(\prod_{i=1}^{s} e^{-\lambda_{1,i}Z} \frac{(\lambda_{1,i}Z)^{j_i}}{j_i!} \times \prod_{i=1}^{s} e^{-\lambda_{2,i}Z} \frac{(\lambda_{2,i}Z)^{m_i}}{m_i!}\right).$$

Twierdzenie 9 nie daje się wyprowadzić bezpośerdnio z odpowiedniego wyniku dla prób o deterministycznej liczności poprzez zastosowanie twierdzenia 6. Jego dowód uzyskano zauważając, że

$$P\left(\mathcal{K}_-(n, k^{(i)}_{N(n)}, a^{(i)}_n) = j_i, \mathcal{K}_+(n, k^{(i)}_{N(n)}, b^{(i)}_n) = m_i, i = 1, \cdots, s\right)$$
$$= E\left[P\left(\mathcal{K}_-(n, k^{(i)}_{N(n)}, a^{(i)}_n) = j_i, \mathcal{K}_+(n, k^{(i)}_{N(n)}, b^{(i)}_n) = m_i, i = 1, \cdots, s | N(n)\right)\right],$$

szacując prawdopodobieństwo warunkowe

$$P\left(\mathcal{K}_-(n, k^{(i)}_{N(n)}, a^{(i)}_n) = j_i, \mathcal{K}_+(n, k^{(i)}_{N(n)}, b^{(i)}_n) = m_i, i = 1, \cdots, s | N(n)\right)$$

i używając twierdzenia o zmajoryzowanym przechodzeniu do granicy pod znakiem wartości oczekiwanej.

4 Uogólnienia

Zauważmy, że zmienne losowe $K_-(n, k, a)$ i $K_+(n, k, a)$ można zapisać następująco:

$$K_-(n, k, a) = \#\{j \in \{1, \cdots, n\} : X_{k:n} - X_j \in (0, a)\}$$

oraz

$$K_+(n,k,a) = \#\{j \in \{1,\cdots,n\} : \ X_{k:n} - X_j \in (-a,0)\}.$$

Zastępując przedziały $(0,a)$ i $(-a,0)$ dowolnym zbiorem borelowskim A otrzymujemy zmienną losową

$$K(n,k,A) := \#\{j \in \{1,\cdots,n\} : \ X_{k:n} - X_j \in A\}$$

będącą uogólnieniem $K_-(n,k,a)$ i $K_+(n,k,a)$ i przyjmującą wartości równe liczbie obserwacji, które wpadły do losowego zbioru wyznaczonego przez k-tą statystykę porządkową oraz zbiór A. W ten sposób, dla ustalonych $1 \le k \le n$, definiujemy proces punktowy $K(n,k,\cdot)$ określony na $\mathbb{R}$.

Prace [R6], [R7] i [R8] poświęcone są asymptotyce zmiennej losowej $K(n,k,A)$ przy $n \to \infty$. W [R6] przedstawiono własności graniczne dla $K(n,k_n,A)$, gdzie $X_{k_n:n}$ to centralna statystyka porządkowa. Główne wyniki [R6] orzekają, że, przy odpowiednich założeniach, częstości $K(n,k_n,A)/n$ są zbieżne według prawdopodobieństwa (zob. twierdzenie 10) oraz, że właściwie wystandaryzowana zmienna losowa $K(n,k_n,A)$ ma asymptotycznie rozkład normalny (zob. twierdzenie 11).

Twierdzenie 10 (R6, Theorem 2.3)**.**
Niech $k_n/n \to \lambda \in (0,1)$ i niech istnieje dokładnie jedna γ spełniająca $F(\gamma) = \lambda$. Wówczas dla dowolnego zbioru borelowskiego $A \subset \mathbb{R}$, którego brzeg ma miarę Lebesgue'a zero,

$$K(n,k_n,A)/n \xrightarrow{P} P\big(X_1 \in \{\gamma - a : a \in A\}\big) \ \ \textit{przy} \ \ n \to \infty.$$

Twierdzenie 11 (R6, Theorem 3.3)**.**
Niech $k_n/n = \lambda + o(n^{-1/2})$, gdzie $\lambda \in (0,1)$, F będzie różniczkowalna w punkcie γ i $f(\gamma) := F'(\gamma) \neq 0$. Załóżmy, że $A \subset \mathbb{R}$ jest zbiorem borelowskim ze skończoną ale niezerową miarą Lebesgue'a i na pewnym zbiorze otwartym zawierającym domknięcie zbioru $\{\gamma - a : a \in A\}$ dystrybuanta F posiada ograniczoną gęstość. Wówczas

$$\sqrt{n}\Big[K(n,k_n,A)/n - P\big(X_1 \in \{\gamma - a : a \in A\}\big)\Big] \xrightarrow{d} \mathcal{N}(0,\sigma^2) \ \ \textit{przy } n \to \infty,$$

gdzie $\sigma^2 = \lambda p_A(\gamma)\overline{p}_A(\gamma) + (1-\lambda)q_A(\gamma)\overline{q}_A(\gamma) + \lambda(1-\lambda)\{\lambda p'_A(\gamma) + (1-\lambda)q'_A(\gamma)\}^2/f^2(\gamma)$ oraz

$$p_A(x) = \frac{P\big(X_1 \in \{x - a : a \in A \cap \mathbb{R}_+\}\big)}{F(x)} \quad , \quad \overline{p}_A(x) = 1 - p_A(x),$$
$$q_A(x) = \frac{P\big(X_1 \in \{x - a : a \in A \cap \mathbb{R}_-\}\big)}{1-F(x)} \quad , \quad \overline{q}_A(x) = 1 - q_A(x).$$

W pracy [R6] podano także uogólnienia twierdzenia 11 na przypadek wielowymiarowy z dwiema lub większą liczbą centralnych statystyk porządkowych oraz wersje twierdzenia 11 ze słabszymi założeniami w szczególnym przypadku, gdy zbiór A daje się przedstawić jako skończona suma przedziałów.

Natomiast w artykule [R7] opisano zachowanie asymptotyczne zmiennych losowych $K(n,m,A)$ i $K(n-k,n,B)$ zliczających liczby obserwacji, które wpadły do zbiorów losowych wyznaczonych przez skrajne statystyki porządkowe $X_{m:n}$ i $X_{n-k:n}$ oraz przez zbiory borelowskie A i B, których brzegi mają miarę Lebesgue'a zero. Wyrażono łączny rozkład graniczny tych zmiennych losowych za pomocą niezależnych rozkładów wielomianowych i ujemnych wielomianowych.

Twierdzenie 12 (R7, Theorem 4.1)**.**
Niech $l_F = -\infty$, $r_F = \infty$ i, dla pary liczb $a_1, a_2 > 0$ takich, że ich iloraz a_1/a_2 jest niewymierny, istnieją granice $\beta(a_i)$ i $\overline{\beta}(a_i)$, $i = 1, 2$, zdefiniowane w (4) i (5) i ich wartości mieszczą się w przedziale $(0, 1)$. Jeśli $A_j, C_j \subset \mathbb{R}_+$, $B_j, D_j \subset \mathbb{R}_-$, $j = 1, 2$, to zbiory borelowskie o brzegach miary Lebesgue'a zero, $A_1 \cap A_2 = B_1 \cap B_2 = C_1 \cap C_2 = D_1 \cap D_2 = \emptyset$, i B_j, C_j, $j = 1, 2$, są ograniczone, to dla dowolnych, ustalonych $m \geq 1$ i $k \geq 0$

$$\begin{aligned}&\big(K(n,m,A_1), K(n,m,A_2), K(n,m,B_1), K(n,m,B_2),\\ &\ K(n,n-k,C_1), K(n,n-k,C_2), K(n,n-k,D_1), K(n,n-k,D_2)\big) \xrightarrow{d}\\ &\mathcal{M}(m-1; \mu_\lambda(A_1), \mu_\lambda(A_2)) \otimes \mathcal{NM}(m; r_\lambda(B_1), r_\lambda(B_2)) \otimes\\ &\mathcal{NM}(k+1; r_\gamma(-C_1), r_\gamma(-C_2)) \otimes \mathcal{M}(k; \mu_\gamma(-D_1), \mu_\gamma(-D_2)),\end{aligned}$$

gdzie

- *dla dowolnego zbioru $E \subset \mathbb{R}$ definiujemy $-E := \{-d : d \in E\}$;*
- *$\mathcal{M}(m; p_1, p_2)$ i $\mathcal{NM}(m; p_1, p_2)$ oznaczają, odpowiednio, rozkład wielomianowy i ujemny wielomianowy z indeksem m i prawdopodobieństwami $p_1, p_2, 1 - p_1 - p_2$;*
- *$r_\lambda(B_j) := \frac{\mu_\lambda(B_j)}{1+\sum_{i=1}^{2}\mu_\lambda(B_i)}$, $r_\lambda(-C_j) := \frac{\mu_\lambda(-C_j)}{1+\sum_{i=1}^{2}\mu_\lambda(-C_i)}$, $\quad j = 1, 2$;*
- *$\mu_\lambda(E) := \int_E \lambda e^{-\lambda x}\mathrm{d}x$ dla $E = A_1, A_2, B_1, B_2, -C_1, -C_2, -D_1, -D_2$;*
- *"$\otimes$" oznacza, że miary prawdopodobieństwa są niezależne.*

Dla prostoty sformułowania, twierdzenie 12 podano dla par zmiennych losowych: $K(n, m, A_1), K(n, m, A_2)$; $K(n, m, B_1), K(n, m, B_2)$; $K(n, n-k, C_1), K(n, n-k, C_2)$ oraz $K(n, n-k, D_1), K(n, n-k, D_2)$. Jest ono także prawdziwe w ogólniejszej wersji, a mianowicie, dla dowolnej liczby zmiennych losowych: $K(n, m, A_i)$, $i = 1, \ldots, s_1$, $K(n, m, B_i)$, $i = 1, \ldots, s_2$, $K(n, n-k, C_i)$, $i = 1, \ldots, s_3$, oraz $K(n, n-k, D_i)$, $i = 1, \ldots, s_4$, przy założeniach, że A_i, $C_i \subset \mathbb{R}_+$, $B_i, D_i \subset \mathbb{R}_-$ to zbiory borelowskie o brzegach miary Lebesgue'a zero, $A_i \cap A_j = B_i \cap B_j = C_i \cap C_j = D_i \cap D_j = \emptyset$ dla $i \neq j$, oraz $B_i, i = 1, \ldots, s_2$, $C_i, i = 1, \ldots, s_3$, są ograniczone. Wynika stąd, że jeśli spełnione są założenia zawarte w pierwszym zdaniu twierdzenia 12, to, dla ustalonych $m \geq 1$ i $k \geq 0$, ciąg czterowymiarowych procesów punktowych

$$\big(K(n, m, \cdot), K(n, m, \cdot), K(n, n-k, \cdot), K(n, n-k, \cdot), n \geq 1\big),$$

określonych na $\mathbb{R}_+ \times \mathbb{R}_- \times \mathbb{R}_+ \times \mathbb{R}_-$, jest słabo zbieżny do czterowymiarowego procesu punktowego, którego składowe są niezależne.

W artykule [R8], uogólniając wyniki z prac [19], [16], [12], [14], [R5] i [R6], podano warunki, przy których

$$K(n, k_n, A)/n \xrightarrow{\text{p.n.}} P\big(X_1 \in \{\gamma_\lambda - a : a \in A\}\big) \text{ gdy } n \to \infty, \tag{6}$$

gdzie $\xrightarrow{\text{p.n.}}$ oznacza zbieżność z prawdopodobieństwem 1, $k_n/n \to \lambda \in [0, 1]$ oraz γ_λ to jednoznacznie wyznaczony kwantyl teoretyczny rzędu λ odpowiadający dystrybuancie F jeśli $\lambda \in (0, 1)$, natomiast $\gamma_0 := l_F$ i $\gamma_1 := r_F$. Wcześniejsze, częściowe rozwiązania tego problemu, były otrzymane przy założeniu, że dystrybuanta F jest ciągła. W [R8] pozbyto się założenia dotyczącego ciągłości, ponadto osłabiono warunek wymagający by obserwacje w próbie były niezależne.

Twierdzenie 13 (R8, Theorem 3.2)**.**
Niech $(X_n, n \geq 1)$ będzie ciągiem zmiennych losowych o tym samym rozkładzie z niekoniecznie ciągłą dystrybuantą F, takim, że dla każdego x należącego do nośnika F

$$\#\{i \in \{1, \ldots, n\} : X_i \leq x\}/n \xrightarrow{\text{p.n.}} F(x), \quad \textit{gdy} \;\; n \to \infty.$$

Załóżmy, że $k_n/n \to \lambda \in [0, 1]$ i spełniony jest jeden z poniższych warunków:

(a) *$\lambda \in (0, 1)$ i kwantyl teoretyczny rzędu λ odpowiadający dystrybuancie F jest wyznaczony jednoznacznie,*

(b) *$\lambda = 0$ i $\gamma_0 > -\infty$,*

(c) *$\lambda = 1$ i $\gamma_1 < \infty$.*

Wtedy zbieżność (6) zachodzi dla dowolnego zbioru borelowskiego $A \subset \mathbb{R}$ takiego, że

$$P\big(X_1 \in \{\gamma_\lambda - b : b \textit{ należy do brzegu zbioru } A\}\big) = 0. \tag{7}$$

Ponadto w [R8] pokazano, że warunek (7) można osłabić, jeśli dodatkowo założy się, że dystrybuanta F jest ciągła albo γ_λ nie jest punktem skupienia nośnika F. Uogólniono także twierdzenie 13 na przypadek prób o losowej liczności.

Gdy wiemy, że zachodzi zbieżność podana w (6), nasuwa się pytanie czy przy odpowiednich założeniach wystandaryzowane proporcje $K_{k_n:n}(A)/n$ mają asymptotyczny rozkład normalny. Odpowiedz na to pytanie jest twierdząca - szczegóły podano w [18] oraz w pracy [S24], która jeszcze nie ukazała się w druku.

Literatura

[1] Bairamov, I., Stepanov, A., Numbers of near-maxima for the bivariate case, *Statist. Probab. Lett.* 80 (2010), 196–205.

[2] Bairamov, I., Stepanov, A., Numbers of near bivariate record-concomitant observations, *J. Multivariate Anal.* 102 (2011), 908–917.

[3] Balakrishnan, N., Pakes, A.G., Stepanov, A., On the number and sum of near-record observations, *Adv. Appl. Prob.* 37 (2005), 765–780.

[4] Balakrishnan, N., Stepanov, A., A note on the number of observations near an order statistic, *J. Statist. Plann. Inf.* 134 (2005), 1–14.

[5] Balakrishnan, N., Stepanov, A., Asymptotic properties of numbers of near minimum observations under progressive Type-II censoring, *J. Statist. Plann. Inf.* 138 (2008), 1010–1020.

[6] Brands, J.J.A.M., Steutel, F.W., Wilms, R.J.G., On the number of maxima of a discrete sample. *Statist. Probab. Lett.* 20 (1994), 209–218.

[7] Bruss F.T., Grübel, R., On the multiplicity of the maximum in a discrete random sample. *Ann. Appl. Probab.* 13 (2003), 1252–1263.

[8] Eisenberg, B., The numbers of players tied for the record. *Statist. Probab. Lett.* 79 (2009), 283–288.

[9] Eisenberg, B., Stengle, G., Strang, G., The asymptotic probability of a tie for first place. *Ann. Appl. Probab.* 3 (1993), 731–745.

[10] Gouet, R., López, F. J., Sanz G., Asymptotic normality for the counting process of weak records and d-records in discrete models. *Bernoulli* 13 (2007), 754–781.

[11] Gouet, R., López, F. J., Sanz G., Limit laws for the cumulative number of ties for the maximum in a random sequence. *J. Statist. Plann. Inf.* 139 (2009), 2988–3000.

[12] Hashorva, E., On the number of near-maximum insurance claim under dependence, *Insurance Math. Econom.* 32 (2003), 37–49.

[13] Hashorva, E., Bivariate maximum insurance claim and related point processes, *Statist. Probab. Lett.* 69 (2004), 117–128.

[14] Hashorva, E., Hüsler, J., Estimation of tails and related quantities using the number of near-extremes, *Comm. Statist. – Theory & Methods* 34 (2004), 337–349.

[15] Khmaladze, E., Nadareishvili, M., Nikabadze, A., Asymptotic behaviour of a number of repeated records, *Statist. Probab. Lett.* (1997) 35, 49–58.

[16] Li, Y., Pakes, A., On the number of near-maximum insurance claims, *Insurance Math. Econom.* 28 (2001), 309–323.

[17] Pakes, A., The number and sum of near-maximamfor thin-tailed populations, *Adv. Appl. Probab.* 32 (2000), 1100-1116.

[18] Pakes, A., Numbers of observations near order statistics, *Aust. N. Z. J. Stat.* 51 (2009), 375–395.

[19] Pakes, A., Steutel, F.W., On the number of records near the maximum, *Austral. J. Statist.* 39 (1997), 179–193.

[20] Qi, Y., A note on the number of maxima in a discrete sample. *Statist. Probab. Lett.* 33 (1997), 373–377.

OMÓWIENIE POZOSTAŁYCH OSIĄGNIĘĆ NAUKOWO-BADAWCZYCH

1 Publikacje naukowe spoza rozprawy habilitacyjnej

S1. A. Dembińska, J. Wesołowski, On characterizing the exponential distribution by linearity of regression for non-adjacent order statistics, *Demonstratio Math.* 30 (1997), 945-952.

S2. A. Dembińska, J. Wesołowski, Linearity of regression for non-adjacent order statistics, *Metrika* 48 (1998), 215-222.

S3. A. Dembińska, J. Wesołowski, Linearity of regression for non-adjacent record values, *J. Statist. Plann. Inf.* 90 (2000), 195-205.

S4. A. Dembińska, Linearity of regression for adjacent order statistics - discrete case, *Demonstratio Math.*, 34 (2001), 711–721.

S5. A. Dembińska, J. Wesołowski, Constancy of regression for size-two record spacings, *Pakistan J. Statist.*, 19 (2003), 143–149

S6. A. Dembińska, F. López-Blázquez, kth records from discrete distributions, *Statist. Probab. Lett.*, 71 (2005), 203-214.

S7. A. Dembińska, F. López-Blázquez, A characterization of geometric distribution through kth weak records, *Comm. Statist. – Theory & Methods* 34 (2005), 2345-2351.

S8. F. López-Blázquez, B. Salamanca Miño, A. Dembińska, A note on the distribution of kth records from discrete distributions, *Statist. Probab. Lett.* 75 (2005), 325–330.

S9. A. Dembińska, A. Stepanov, Limit theorems for the ratio of weak records, *Statist. Probab. Lett.* 76 (2006), 1454–1464.

S10. N. Balakrishnan, A. Dembińska, Ordered random variables from discontinuous distributions, *J. Iran. Stat. Soc.* 6 (2007), 31–46.

S11. K. Danielak, A. Dembińska, Some characterizations of discrete distributions based on weak records, *Statist. Papers* 48 (2007), 479–489.

S12. K. Danielak, A. Dembińska, On characterizing discrete distributions via conditional expectations of weak record values, *Metrika* 66 (2007), 129–138.

S13. A. Dembińska, A review on characterizations of discrete distributions based on records and kth records, *Comm. Statist. – Theory & Methods*, 36 (2007), 1381–1387.

S14. A. Dembińska, Comments on: Progressive censoring methodology: an appraisal, *Test* 16 (2007), 262-264.

S15. A. Dembińska, Records from discrete distributions. W: *Recent Developments in Ordered Random Variables* pod redakcją M. Ahsanullaha i M.Z. Raqaba, strony 77–95, Nova Science Publishers, New York 2007.

S16. A. Dembińska, Characterizations of discrete distributions based on record values. W: *Extreme Value Distributions* pod redakcją M. Ahsanullaha i S. Kirmaniego, strony 63–75, Nova Science Publishers, New York 2007.

S17. N. Balakrishnan, A. Dembińska, Progressively Type-II right censored order statistics from discrete distributions, *J. Statist. Plann. Inf.* 138 (2008), 845–856.

S18. N. Balakrishnan, A. Dembińska, A. Stepanov, Precedence-type tests based on record values, *Metrika* 68 (2008), 233–255.

S19. A. Dembińska, Discrete order statistics. W: *Encyclopedia of Statistical Sciences* pod redakcją S. Kotza, C.B. Reada, N. Balakrishnana i B. Vidakovica, Online version, John Wiley & Sons, Hoboken, New Jersey 2008.

S20. A. Dembińska, kth records from geometric distribution, *Statist. Probab. Lett.* 78 (2008), 1662–1670.

S21. A. Dembińska, K. Danielak, On moments of k-records from discrete distributions, *Comm. Statist. – Theory & Methods* 37 (2008), 2516–2531.

S22. N. Balakrishnan, A. Dembińska, Erratum to 'Progressively Type-II right censored order statistics from discrete distributions' [J. Statist. Plann. Inference 138 (2008) 845–856], *J. Statist. Plann. Inference* 139 (2009), 1572–1574.

S23. N. Balakrishnan, E. Cramer, A. Dembińska, Characterizations of geometric distribution through progressively Type-II right censored order statistics, *Statistics* 45 (2011), 559–573.

S24. A. Dembińska, Asymptotic normality of numbers of observations in random regions determined by order statistics. Złożone do druku.

S25. A. Dembińska, Asymptotic behavior of central order statistics from stationary processes. Złożone do druku.

2 Krótki opis publikacji naukowych spoza rozprawy

Artykuły spoza rozprawy związane są z szeroko pojętą tematyką analizy statystycznych danych uporządkowanych. Po pierwsze, dotyczą rozmaitych modeli próbkowania statystycznego: zwykłych statystyk porządkowych, wartości rekordowych i ich uogólnień, progresywnie cenzurowanych statystyk porządkowych i uogólnionych statystyk porządkowych. Po drugie, analizowane są w nich rozmaite aspekty matematyczne i zastosowania praktyczne rozpatrywanych modeli: własności rozkładów i ich momentów, charakteryzacje, analiza asymptotyczna i konstrukcje testów statystycznych. W pracach tych poświęcono wiele uwagi badaniu statystycznych danych uporządkowanych pochodzących z rozkładów dyskretnych oraz takich, które zawierają komponentę dyskretną i ciągłą. Założenie ciągłości znacznie upraszcza wnioskowanie, a uwolnienie się od tego restrykcyjnego warunku prowadzi do wielu zaskakujących uogólnień wyników klasycznego modelu ciągłego.

Pięć pierwszych prac, [S1-S5], stanowi część rozprawy doktorskiej pt. "Charakteryzacje rozkładów prawdopodobieństwa związane z własnościami regresyjnymi statystyk porządkowych i rekordowych". W artykułach [S1] i [S2] podano klasę ciągłych rozkładów prawdopodobieństwa, dla których regresja jednej statystyki porządkowej względem innej, przeszłej lub przyszłej ma własność liniowości - zamykając problem postawiony przez Fergusona w 1967 roku. Rozwiązanie analogicznego problemu dla statytyk rekordowych przedstawiono w [S3]. W [S4] uzyskano charakteryzacje rozkładów dyskretnych związane z własnością liniowości regresji statystyk porządkowych, natomist w [S5] scharakteryzowano zmodyfikowany rozkład geometryczny za pomocą liniowości regresji dla statystyk rekordowych o odstępie dwa.

Kolejną grupę artykułów tworzą prace [S6], [S7], [S8], [S13], [S15], [S16], [S20] i [S21] poświęcone k-tym rekordom pochodzącym z rozkładów dyskretnych. Kluczową wśród nich jest praca [S6], w której pokazano, że funkcjonujące w literaturze definicje k-tych rekordów, uznawane do tej pory za równoważne, w rzeczywistości w przypadku dyskretnym równoważne *nie* są. Aby uporządkować terminologię, wprowadzono różne nazwy dla k-tych rekordów generowanych przez trzy nierównoważne definicje: mocne k-te rekordy, k-te rekordy i słabe k-te rekordy. Omówiono także podstawowe własności wyżej wymienionych trzech rodzajów k-tych rekordów i skorygowano pewne błędy i nieścisłości, które znaleźć można w literaturze dotyczącej dyskretnych k-tych rekordów. W powiązanej z pracą [S6] pracy [S8] wykazano, że znana charakteryzacja k-tych rekordów, wyrażająca ich łączny rozkład za pomocą rozkładu łącznego rekordów, nie jest prawdziwa w przypadku dyskretnym gdy $k > 1$.

W pracy [S7] pokazano, że rozkład geometryczny jest jedynym rozkładem dyskretnym, dla którego regresja pierwszego k-tego słabego rekordu względem zerowego k-tego słabego rekordu jest funkcją liniową warunku ze współczynnikiem kierunkowym równym 1, uogólniając w ten sposób znane w literaturze twierdzenia dla $k = 1$. Natomiast praca [S20] zawiera opis struktury zależności k-tych rekordów pochodzących z rozkładu geometrycznego a także charakteryzacje tego rozkładu i jego ogona za pomocą własności k-tych rekordów. Przegląd wyników dotyczących charakteryzacji dyskretnych rozkładów prawdopodobieństwa za pomocą własności różnych rodzajów rekordów i k-tych rekordów został podany w pracy [S13].

Pozycje [S15] i [S16] to rozdziały w książkach, napisane na zaproszenie redaktorów tych książek. Rozdział [S15] to przegląd wyników dotyczących dyskretnych rekordów i k-tych rekordów. Oprócz znanych w literaturze wyników z tej dziedziny, rozdział ten zawiera także opis własności asymptotycznych k-tych, mocnych k-tych oraz słabych k-tych rekordów pochodzących z rozkładu geometrycznego. Natomiast rozdział [S16] przedstawia znane w literaturze własności rekordów, słabych rekordów, k-tych rekordów oraz słabych i mocnych k-tych rekordów pochodzących z rozkładów dyskretnych. Zostały w nim także podane i udowodnione nowe twierdzenia charakteryzujące rozkład geometryczny w klasie rozkładów dyskretnych poprzez niezależność oraz częściową niezależność zerowego k-tego słabego rekordu i różnicy pomiędzy pierwszym i zerowym k-tym słabym rekordem.

W pracy [S21] wyprowadzono rekurencyjne wzory opisujące zwykłe i mieszane momenty k-tych rekordów pochodzących z rozkładów dyskretnych i uzyskano ich uproszczenia w przypadku populacji o rozkładzie geometrycznym. Rozpatrzono trzy rodzaje k-tych rekordów: mocne k-te rekordy, k-te rekordy i słabe k-te rekordy.

Prace [S9], [S11] i [S12] stanowią kolejną grupę bliskich tematycznie artykułów. Poświęcone są one badaniu własności słabych rekordów pochodzących z populacji o rozkładzie dyskretnym. W pracy [S9] przedstawiono twierdzenia graniczne opisujące zbieżność z prawdopodobieństwem jeden oraz zbieżność według prawdopodobieństwa ilorazu słabych rekordów, niekoniecznie sąsiednich. Pokazano, że graniczne zachowanie się takiego ilorazu zależy od tego czy dystrybuanta wyjściowego rozkładu jest wolno, regularnie czy szybko zmieniającą się funkcją w nieskończoności. Ponadto podano i udowodniono asymptotyczne własności zmiennej losowej, która zlicza liczbę słabych rekordów, których wartość równa się n. Prace [S11] i [S12] poświęcone są charakteryzacjom rozkładów dyskretnych opartych na własnościach słabych rekordów. W pracy [S11] udowodniono szereg charakteryzacji różnych rozkładów dyskretnych - w tym także rozkładu geometrycznego - opartych na zależnościach pomiędzy funkcjami liniowymi słabych rekordów. W pracy [S12] pokazano, że warunkowa wartość oczekiwana dowolnej rosnącej funkcji słabego rekordu pod warunkiem następnego słabego rekordu, rozpatrywana jako funkcja zmiennej losowej w warunku, jednoznacznie wyznacza wyjściowy rozkład dyskretny z dokładnością do prawdopodobieństwa najmniejszego punktu nośnika. Przedstawiono także wzór pozwalający wyznaczyć ten rozkład.

Kolejna grupa prac, złożona z [S10], [S14], [S17], [S22] i [S23], dotyczy uogólnionych statystyk porządkowych pochodzących z próby o niekoniecznie ciągłej dystrybuancie, w tym prace [S14], [S17] i [S22] poświęcone są szczególnemu przypadkowi uogólnionych statystyk porządkowych - progresywnie cenzurowanym statystykom porządkowym.

W artykule [S10] wyznaczono największe klasy rozkładów, dla których rekordy i słabe rekordy są szczególnym przypadkiem uogólnionych statystyk porządkowych. Pokazano także, że progresywnie cenzurowane statystyki porządkowe pochodzące z próbki o dowolnym rozkładzie to szczególny przypadek uogólnionych statystyk porządkowych i wykorzystując ten fakt wyprowadzono pewne własności progresywnie cenzurowanych statystyk porządkowych. Przedstawiono także charakteryzacje rozkładu geometrycznego i zmodyfikowanego rozkładu geometrycznego oparte na własnościach progresywnie cenzurowanych statystyk porządkowych.

Praca [S14] to komentarz do artykułu przeglądowego N. Balakrishnana *Progressive censoring methodology: an appraisal*, napisany na zaproszenie redaktorów pisma *Test*, w którym ten artykuł przeglądowy się ukazał. Podano w niej rozwiązanie problemu otwartego postawionego w tym artykule, a dokładniej pokazano, że różnego typu oszacowania dla wartości oczekiwanych progresywnie cenzurowanych statystyk porządkowych, otrzymane przy założeniu, że wyjściowy rozkład jest ciągły, pozostają prawdziwe także po opuszczeniu założenia dotyczącego ciągłości rozkładu. Przeanalizowano dokładniej jeden typ takich oszacowań i wyznaczono dla niego rozkład, dla którego oszacowanie jest osiągane.

W [S17] poddano gruntownej analizie jedno- i wielowymiarowe rozkłady progresywnie cenzurowanych statystyk porządkowych pochodzących z populacji o niekoniecznie ciągłej dystrybuancie. Podano reprezentację, która pozwala wyrazić rozkład cenzurowanych statystyk porządkowych z dowolnego rozkładu za pomocą cenzurowanych statystyk porządkowych pochodzących z rozkładu jednostajnego. Dowód tej reprezentacji zawierał błąd - poprawioną jego wersję zamieszczono w [S22]. W [S17] pokazano także, ze jeśli wyjściowy rozkład jest dyskretny a jego nośnik zawiera więcej niż dwa punkty, to progresywnie cenzurowane statystyki porządkowe nie tworzą łańcucha Markowa. Przedstawiono również kilka charakteryzacji rozkładu geometrycznego opartych na własnościach progresywnie cenzurowanych statystyk porządkowych.

W pracy [S23] otrzymano nowe charakteryzacje rozkładu geometrycznego i jego ogona poprzez własności progresywnie cenzurowanych statystyk porządkowych. Dokładniej podano twierdzenia charakteryzacyjne wykorzystujące warunkową wartość oczekiwaną, niezależność oraz równość rozkładów zmiennych losowych będących pewnymi funkcjami progresywnie cenzurowanych statystyk porządkowych. Twierdzenia te przedstawiono także w ogólniejszej wersji - w terminach uogólnionych statystyk porządkowych.

Trzy prace [S18], [S19] i [S24] dotyczą różnych dziedzin: konstrukcji testów statystycznych, własności probabilistycznych statystyk porządkowych pochodzących z rozkładów dyskretnych oraz centralnych twierdzeń granicznych dla liczby obserwacji, które wpadły do pewnego zbioru losowego wyznaczonego przez statystykę porządkową. Dokładniej [S18] przedstawia konstrukcję i analizę trzech testów jednorodności dwóch prób wobec alternatyw o ich stochastycznym uporządkowaniu, opartych na różnych statystykach testowych zdefiniowanych za pomocą wartości rekordowych. Opisano ich dokładne rozkłady przy założeniu prawdziwości hipotezy zerowej oraz oszacowano ich moc w przypadku nieparametrycznych alternatyw Lehmanna. Pozycja [S19] to opracowanie hasła "Discrete Order Statistics" w *Encyclopedia of Statistical Sciences* przygotowane na zamówienie redakcji tej encyklopedii. Artykuł [S24] tematycznie przynależy do rozprawy habilitacyjnej; nie został jednak do niej dołączony, ponieważ nadal jest w trakcie recenzji. Pokazano w nim, że liczba obserwacji, które wpadły do pewnego zbioru wyznaczonego przez statystykę porządkową, po odpowiedniej standaryzacji, ma asymptotycznie rozkład normalny. Rozpatrzono zarówno przypadek centralnych jak i skrajnych statystyk porządkowych w sytuacji, gdy próbka pochodzi z dowolnego rozkładu.

Ostatnia spośród prac [S25] stanowi odzwierciedlenie najnowszych zainteresowań autorki twierdzeniami granicznymi dla ciągów zależnych zmiennych losowych. W pracy tej pokazano, że centralne statystyki porządkowe, pochodzące ze ściśle stacjonarnego i ergodycznego ciągu obserwacji, są mocno zgodnymi estymatorami kwantyli teoretycznych jeśli tylko kwantyle te są wyznaczone jednoznacznie. Następnie opisano trzy możliwe wersje zachowania asymptotycznego centralnych statystyk porządkowych w sytuacji, gdy odpowiedni kwantyl teoretyczny nie jest wyznaczony jednoznacznie. Wyniki te użyto do uzyskania twierdzeń granicznych opisujących zachowanie centralnych statystyk porządkowych z procesów liniowych zarówno z absolutnie ciągłymi jak i z dyskretnymi zaburzeniami.

3 Referaty wygłoszone na konferencjach naukowych

- *Charakteryzacje związane z liniowością regresji niesąsiednich statystyk porządkowych i rekordów*, *VI Międzynarodowa Konferencja z Probabilistyki*, Poraj, Polska, 5-9.06.2000;
- *On characterizing distributions via linearity of regression for non-adjacent record values*, *Internatinal Conference on Mathematical Statistics STAT'2000*, Szklarska Poręba, Polska, 21-25.08.2000;
- *Charakteryzacje rozkładów prawdopodobieństwa związane z własnościami regresyjnymi statystyk porządkowych i rekordowych*, *XXVII Konferencja Statystyka Matematyczna*, Wisła, Polska, 3-7.12.2001;
- *k-te rekordy w przypadku dyskretnym i związane z nimi charakteryzacje*, *XXVIII Konferencja Statystyka Matematyczna*, Wisła, Polska, 2-6.12.2002;
- *Distributional Properties of Discrete kth Records*, warsztaty *Characterizations and Bounds for Ordered Statistical Data*, Warszawa, Polska, 12-17.05.2003;
- *K-th records from discrete distributions*, konferencja *Ordered Statistical Data: Approximations, Bounds and Characterizations*, Warszawa, Polska, 24-28.05.2004;
- *Characterizations of discrete distributions based on record values*, konferencja *Ordered Statistical Data: Approximations, Bounds and Characterizations*, Izmir University of Economics, Izmir, Turcja, 15-18.06.2005;
- *On record statistics from discrete distributions*, *6th Summer Mini-Conference on Statistics*, McMaster University, Hamilton, Kanada, 5.08.2005;
- *O liczbie obserwacji, które wpadły do pewnego otoczenia statystyki porządkowej*, konferencja *Statystyka matematyczna – Wisła 2005*, Wisła, Polska, 5-9.12.2005;
- *On characterizing discrete distributions via conditional expectations of weak records*, konferencja *Ordered Statistical Data and Related Topics*, Ferdowsi University of Mashhad, Mashhad, Iran, 15-17.06.2006;
- *Progressively Type-II right censored order statistics from discrete distributions*, XII międzynarodowa konferencja *Applied Stochastic Models and Data Analysis* (ASDMA 2007), Technical University of Crete, Kreta, Grecja, 29.05.2007-1.06.2007;
- *Precedence-type tests based on record values*, konferencja *Ordered Statistical Data and Inequalities: Theory and Applications*, University of Jordan, Amman, Jordania, 12-14.06.2007;
- *Distributional properties of progressively Type-II right censored order statistics from discrete distributions*, *2007 Mini-Conference on Statistics*, McMaster University, Hamilton, Kanada, 20.07.2007;

- *The asymptotic distribution of numbers of observations near order statistics*, *International Conference on Ordered Statistical Data and Its Applications*, Akwizgran, Niemcy, 6.03.2008-12.03.2008;

- *On asymptotic independence of numbers of observations near order statistics*, *8th Mini-Conference on Statistics*, McMaster University, Hamilton, Kanada, 24.07.2008;

- *Ordered Random Variables from Discrete Distributions*, konferencja *6th St. Petersburg Workshop on Simulation*, St. Petersburg State University, Petersburg, Rosja, 29.06.2009-2.07.2009;

- *On Quantile Representation for Ordered Random Variables from Discontinuous Distributions*, *2009 Summer Mini-Conference*, McMaster University, Hamilton, Kanada, 31.07.2009;

- *Asymptotic properties of numbers of observations near order statistics*, konferencja *45th Scientific Meeting of the Italian Stastical Society*, Padwa, Włochy, 16-18.06.2010;

- *Limit Theorems for Numbers of Observations Near Ordered Statistics*, *9th Inernational Conference on Ordered Statistical Data and its Applications*, Zagazig, Egipt, 11-13.07.2010;

- *Limit theorems for numbers of observations falling into random regions determined by order statistics*, *2011 Mini-Conference on Statistics*, McMaster University, Hamilton, Kanada, 25.07.2011;

- *On the asymptotics of numbers of observations in random regions determined by order statistics*, *International Conference on Advances in Probability and Statistics – Theory and Applications: A Celebration of N. Balakrishnan's 30 years of Contributions to Statistics*, Hong Kong, Chiny, 28.12.2011-31.12.2011;

- *Asymptotic behavior of central order statistics from stationary processes*, *10th Inernational Conference on Ordered Statistical Data and its Applications OSDA2012*, Murcia, Hiszpania, 23-25.05.2012;

- *Własności asymptotyczne liczby obserwacji w otoczeniach statystyk porządkowych*, *XII Konferencja z Probabilistyki*, Będlewo, Polska, 28.05-1.06.2012;

- *On limiting properties of central order statistics*, *11th Iranian Statistical Conference*, Iran University of Science and Technology, Teheran, Iran, 28-30.08.2012.

4 Nagrody, wyróżnienia i stypendia

- wyróżnienie Rady Wydziału Matematyki i Nauk Informacyjnych Politechniki Warszawskiej za rozprawę doktorską, 2002 rok;
- 5 nagród Rektora Politechniki Warszawskiej za osiągnięcia naukowe w latach 2002, 2005-2006, 2007, 2008 i 2009-2010;
- Złota Kreda - nagroda za zajęcie I miejsca w plebiscycie na osobę najlepiej prowadzącą ćwiczenia na Wydziale Matematyki i Nauk Informacyjnych Politechniki Warszawskiej w roku akademickim 2010/2011;
- stypendium doktorskie Politechniki Warszawskiej (czerwiec 2000 - maj 2001);
- stypendium habilitacyjne Politechniki Warszawskiej (luty 2011 - styczeń 2012).

5 Granty badawcze:

- 7 grantów Politechniki Warszawskiej w latach 1996, 1997, 1998, 1999, 2000, 2003, 2004, wykonawca;
- grant KBN (promotorski), marzec 2001 - luty 2003, główny wykonawca;
- 8 grantów Politechniki Warszawskiej w latach 2005, 2006, 2007-2008, 2008-2009, 2009-2010, 2010-2011, 2011-2012 i 2012, kierownik.

6 Wyjazdy w ramach współpracy naukowej

- Universidad de Sevilla, Hiszpania, wrzesień 2003;
- University of Nothern Iowa, USA, luty 2005;
- McMaster University, Kanada, lipiec-sierpień 2005;
- McMaster University, Kanada, lipiec-sierpień 2006;
- McMaster University, Kanada, lipiec-sierpień 2007;
- McMaster University, Kanada, lipiec 2008;
- RWTH Aachen University, Niemcy, wrzesień 2008;
- McMaster University, Kanada, lipiec-sierpień 2009;
- McMaster University, Kanada, lipiec 2011;
- Ferdowsi University of Mashhad, Iran, sierpień 2012.

7 Udział w komitetach redakcyjnych czasopism

Członek komitetu redakcyjnego czasopism:

- Communications in Statistics – Theory and Methods, od 2009;
- Communications in Statistics – Simulation and Computation, od 2009.

8 Wskaźniki służące do oceny dorobku naukowego

Poniżej podane wartości wskaźników zostały wyznaczone przez Oddział Informacji Naukowej Biblioteki Głównej Politechniki Warszawskiej - zestawienie przygotowane przez ten oddział dołączono do wniosku jako jeden z załączników.

a) Sumaryczny *impact factor* publikacji naukowych według listy *Journal Citation Reports*, zgodnie z rokiem opublikowania (ponieważ nie są jeszcze znane wartości *impact factor* z 2012 roku, w przypadku publikacji z tego roku zastosowano *impact factor* z roku 2011): **11,084**.

b) Liczba cytowań publikacji według bazy *Web of Science*:

- liczba cytowań: **110**,
- liczba cytowań bez samocytowań: **63**.

c) Indeks Hirscha opublikowanych publikacji według bazy *Web of Science*: **6**.

Anna Dembińska