Wydział Matematyki, Fizyki i Informatyki
Uniwersytet Marii Curie-Skłodowskiej w Lublinie

Krystyna Maciąg

Charakteryzacje rozkładów prawdopodobieństwa przez regresje niekolejnych uogólnionych statystyk porządkowych oraz dyskretnych wartości rekordowych

Rozprawa doktorska napisana pod kierunkiem
dr. hab. Mariusza Bieńka, prof. UMCS

LUBLIN 2023

Składam serdeczne podziękowania Promotorowi
dr. hab. Mariuszowi Bieńkowi, prof. UMCS
za opiekę naukową, cenne wskazówki oraz pomoc
w trakcie przygotowania niniejszej pracy.

Pragnę również serdecznie podziękować Recenzentom
dr. hab. Annie Dembińskiej
i prof. dr. hab. Tomaszowi Rychlikowi
za wnikliwe uwagi, które przyczyniły się
do ulepszenia pierwotnego tekstu niniejszej pracy.

Spis treści

Wstęp

Jednym z podstawowych problemów statystyki matematycznej oraz stosowanej jest wyznaczenie na podstawie obserwacji pewnej wielkości losowej jej nieznanego rozkładu prawdopodobieństwa. Jednym z możliwych podejść do tego problemu jest poszukiwanie specyficznych własności probabilistycznych charakteryzujących dany rozkład prawdopodobieństwa. W rozważaniach tego typu szczególnie ważne są własności, które prowadzą do jednoznacznego określenia rozkładu prawdopodobieństwa. Na przykład wiadomo, że jedynym rozkładem ciągłym określonym na półprostej $[0,\infty)$ z własnością „braku pamięci" jest rozkład wykładniczy, a jedynym rozkładem skupionym na zbiorze liczb całkowitych nieujemnych z tą własnością – rozkład geometryczny.

W niniejszej rozprawie rozważany będzie problem charakteryzacji rozkładów prawdopodobieństwa na podstawie znajomości tylko pojedynczej funkcji regresji w różnych modelach uporządkowanych zmiennych losowych. Badania nad problemem charakteryzacyjnym tego typu zostały zapoczątkowane w roku 1967 przez Fergusona [28]. Scharakteryzował on rozkłady wykładniczy, Pareto i potęgowy jako jedyne, dla których regresja kolejnych statystyk porządkowych jest liniowa, tzn.

$$E\big(X_{r+1:n} \mid X_{r:n}\big) = aX_{r:n} + b,$$

dla pewnych $a,b \in \mathbb{R}$, gdzie $X_{1:n} \leq \ldots \leq X_{n:n}$ są statystykami porządkowymi z próby o rozmiarze n. Od tamtego czasu podobne zagadnienia były tematem wielu prac i monografii. Analogiczny problem charakteryzacyjny można wypowiedzieć dla innych modeli uporządkowanych zmiennych losowych, a obok zwykłych statystyk porządkowych są nimi m.in. sekwencyjne statystyki porządkowe, statystyki porządkowe progresywnie cenzurowane typu II, wartości rekordowe, k-te wartości rekordowe, jak również wartości rekordowe Pfeifera.

W trakcie badań nad tymi różnymi modelami uporządkowanych zmiennych losowych okazało się, że mają one wiele wspólnych własności. Spostrzeżenie to doprowadziło do wprowadzenia przez Kampsa w roku 1995 pojęcia uogólnionych statystyk porządkowych, które służy unifikacji podejścia do wielu zagadnień związanych z pozornie różnymi modelami uporządkowanych zmiennych losowych (zob. [36], [37]). W niniejszej pracy rozważamy w szczególności problem charakteryzacji rozkładów ciągłych przez funkcję regresji określoną wzorem

$$E\left(h\big(X_*^{(r+\ell)}\big) \mid X_*^{(r)} = x\right) = \xi(x),$$

gdzie h i ξ są danymi funkcjami, a $X_*^{(1)}, \ldots, X_*^{(n)}$ są uogólnionymi statystykami porządkowymi z pewnego rozkładu prawdopodobieństwa. Zaprezentowane zostanie nowatorskie podejście do problemu charakteryzacyjnego w oparciu o własność Markowa rozważanych modeli.

W rozdziale 1 przypomnimy podstawowe definicje i potrzebne w niniejszej rozprawie własności różnych modeli uporządkowanych zmiennych losowych, m.in. statystyk porządkowych, górnych wartości rekordowych, a także uogólnionych statystyk porządkowych. Omówimy dokładnie problem charakteryzacji rozkładów prawdopodobieństwa na podstawie znajomości tylko pojedynczej funkcji regresji tych modeli oraz przedstawimy zarys koncepcji nowego podejścia do problemu charakteryzacyjnego opartej na własności Markowa rozważanych modeli. Korzystając z tej własności wykażemy kluczowy lemat, w którym pokazana zostanie specyficzna struktura rekurencyjna regresji. W szczególności, własność ta pozwala zastąpić regresję niekolejnych uogólnionych statystyk porządkowych przez odpowiednio zmodyfikowaną regresję kolejnych zmiennych w tym modelu. Na koniec tego rozdziału wyprowadzimy istotne własności regresji modeli uporządkowanych zmiennych losowych.

W rozdziale 2 rozważane będą charakteryzacje rozkładów absolutnie ciągłych o ciągłych funkcjach gęstości przez pojedyncze regresje niekolejnych uogólnionych statystyk porządkowych. Przedstawimy nowe warunki jednoznaczności charakteryzacji rozkładu absolutnie ciągłego w terminach jednoznaczności rozwiązania odpowiedniego problemu różniczkowego (równania różniczkowego lub układu równań różniczkowych). Otrzymane wyniki posłużą do przedstawienia nowych charakteryzacji rozkładów absolutnie ciągłych oraz do wyprowadzenia nowych dowodów znanych charakteryzacji. Podamy charakteryzację dla pewnego rozkładu beta oraz wykażemy, że rozkład normalny, Gompertza oraz Weibulla z parametrem kształtu $\delta > 1$ są jednoznacznie charakteryzowane przez regresje niekolejnych uogólnionych statystyk porządkowych o odstępie dwa. Wyniki tego rozdziału są poprawioną wersją wyników pracy [13].

W rozdziale 3 wyniki rozdziału 2 zostaną uogólnione na klasę wszystkich rozkładów ciągłych, niekoniecznie mających funkcję gęstości. Wykażemy, że jednoznaczność charakteryzacji rozkładów ciągłych zachodzi wtedy i tylko wtedy, gdy odpowiednie problemy całkowe (równania całkowe lub układy równań całkowych) mają jednoznaczne rozwiązanie. Korzystając z otrzymanych wyników podamy nowe charakteryzacje rozkładów ciągłych. W szczególności wykażemy, że rozkład gamma, Gumbela i logistyczny są jednoznacznie charakteryzowane przez odpowiadające im regresje uogólnionych statystyk porządkowych o odstępie dwa. Wyniki tego rozdziału uogólniają wyniki pracy [13] i są poprawioną wersją wyników opublikowanych w pracy [14].

Przedmiotem rozważań w rozdziale 4 jest analogiczny problem charakteryzacyjny dla dyskretnych rozkładów prawdopodobieństwa z wykorzystaniem regresji słabych wartości rekordowych. Pokażemy, że jednoznaczność charakteryzacji w tym przypadku jest równoznaczna z jednoznacznością rozwiązania odpowiadającego równania różnicowego lub układu równań różnicowych z pewnymi niestandardowymi warunkami ograniczającymi. Jako przykład zastosowania podamy nowe charakteryzacje rozkładów dyskretnych oraz wykażemy, że rozkład Poissona i ujemny dwumianowy są jednoznacznie charakteryzowane przez odpowiadające im regresje słabych wartości rekordowych o odstępie dwa. Pierwotne wyniki tego rozdziału zostały przedstawione w pracy [15].

Na zakończenie rozprawy, podsumowując uzyskane w niej wyniki, przedstawiamy również zestawienie problemów związanych z tematyką rozprawy, które nadal pozostają otwarte i z pewnością zasługują na podjęcie nad nimi dalszych badań.

Rozdział 1

Regresje modeli uporządkowanych zmiennych losowych

Niech $\{X_n, n \geq 1\}$ będzie ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie danym przez dystrybuantę F. Przy ustalonym n zmienne losowe $X_1, \ldots, X_n$ mogą zostać dowolnie uporządkowane, co prowadzi do otrzymania zależnych (uporządkowanych) zmiennych losowych. Najprostsze sposoby uporządkowania zmiennych losowych dają w wyniku statystyki porządkowe oraz wartości rekordowe. Podkreślmy, że mimo tego, że modele te mogą być zdefiniowane bez żadnych założeń co do rozkładów oraz zależności poszczególnych obserwacji, w rozważaniach niniejszej rozprawy ograniczamy się jedynie do przypadku, gdy $X_1, \ldots, X_n$ są niezależnymi zmiennymi losowymi o jednakowym rozkładzie.

W rozdziale tym omówimy najpierw podstawowe modele uporządkowanych zmiennych losowych, a następnie przypomnimy pojęcie uogólnionych statystyk porządkowych oraz ich własności potrzebne w dalszym ciągu pracy. Przedstawimy problem charakteryzacji rozkładów prawdopodobieństwa przez regresje modeli uporządkowanych zmiennych losowych, który jest głównym tematem rozprawy. Przypomnimy własność Markowa zmiennych losowych, a następnie pokażemy, że regresje modeli uporządkowanych zmiennych losowych mających tę własność mają specyficzną strukturę rekurencyjną. Stanowi to punkt wyjścia do nowego spojrzenia na problem charakteryzacyjny, które prezentujemy w niniejszej rozprawie. Rozdział kończy się wyprowadzeniem własności regresji modeli uporządkowanych zmiennych losowych, które będą potrzebne w dalszym ciągu pracy.

1.1 Modele uporządkowanych zmiennych losowych

1.1.1 Statystyki porządkowe

Statystyki porządkowe stały się przedmiotem badań już na początku dwudziestego wieku. Pojawiają się one w wielu obszarach statystyki i mają wiele praktycznych zastosowaniach. Liczne własności i zastosowania statystyk porządkowych zostały omówione m.in. w monografiach [5] i [23]. Ich różne zastosowania np. w teorii niezawodności można znaleźć również w monografii Barlowa i Proschana [10].

Definicja 1.1. Niech $(X_1,\dots,X_n)$ będzie próbą prostą, tzn. wektorem złożonym z niezależnych zmiennych losowych o jednakowym rozkładzie. Przypuśćmy, że wartości wektora zostały ustawione w porządku niemalejącym

$$X_{1:n} \le \dots \le X_{r:n} \le \dots \le X_{n:n}.$$

Wówczas zmienną losową $X_{r:n}$ nazywamy r-tą statystyką porządkową z próby o rozmiarze n.

Jeżeli $X_1,\dots,X_n$ są niezależnymi zmiennymi losowymi o absolutnie ciągłej dystrybuancie F z funkcją gęstości f, to łączna gęstość wektora $(X_{1:n},\dots,X_{n:n})$ ma postać

$$f_{X_{1:n},\dots,X_{n:n}}(x_1,\dots,x_n) = n!\prod_{i=1}^{n} f(x_i), \quad x_1 < \dots < x_n, \tag{1.1}$$

oraz 0 poza tym, a rozkład statystyki porządkowej $X_{r:n}$ ma funkcję gęstości daną wzorem

$$f_{X_{r:n}}(x) = \frac{n!}{(r-1)!(n-r)!}\big(F(x)\big)^{r-1}\big(1-F(x)\big)^{n-r} f(x), \quad x \in \mathbb{R}. \tag{1.2}$$

Natomiast łączną gęstość r-tej i $(r+k)$-tej statystyki porządkowej $(1 \le r < r+k \le n)$ można przedstawić następująco

$$\begin{aligned} f_{X_{r:n},X_{r+k:n}}(x_1,x_2) =& \frac{n!}{(r-1)!(k-1)!(n-r-k)!} \\ & \cdot \big(F(x_1)\big)^{r-1} f(x_1)\big(F(x_2)-F(x_1)\big)^{k-1}\big(1-F(x_2)\big)^{n-r-k} f(x_2), \end{aligned} \tag{1.3}$$

dla $x_1 < x_2$ oraz 0 poza tym. Korzystając ze wzorów (1.2) i (1.3) można wyprowadzić wzór na warunkową funkcję gęstości $X_{r+k:n}$ pod warunkiem $X_{r:n} = x_1$

$$\begin{aligned} f_{X_{r+k:n}|X_{r:n}}(x_2 \mid x_1) =& \frac{(n-r)!}{(k-1)!(n-r-k)!} \\ & \cdot \left(\frac{F(x_2)-F(x_1)}{1-F(x_1)}\right)^{k-1}\left(\frac{1-F(x_2)}{1-F(x_1)}\right)^{n-r-k} \frac{f(x_2)}{1-F(x_1)}, \end{aligned}$$

dla $x_1 < x_2$ oraz 0 poza tym.

1.1.2 Wartości rekordowe

Pojęcie wartości rekordowych wprowadził Chandler w 1952 roku [17]. Od tamtego czasu badanie ich własności było tematem wielu prac i monografii. Wartości rekordowe są ściśle związane z ekstremalnymi statystykami porządkowymi. Są stosowane w modelowaniu zdarzeń katastroficznych, jak również w teorii niezawodności do opisu tzw. modeli szoku. Ich liczne własności i zastosowania można znaleźć m.in. w monografiach [1], [4] i [46].

Niech $\{X_n, n \geq 1\}$ będzie ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie danym przez dystrybuantę F.

Definicja 1.2. Dla ustalonej liczby naturalnej $r \geq 1$ określamy najpierw ciąg czasów rekordowych następująco

$$T_1 = 1, \quad T_r = \min\left\{ j > T_{r-1} : X_j > X_{T_{r-1}} \right\}, \quad r \geq 2.$$

Wówczas r-tą wartością rekordową (inaczej rekordem) ciągu $\{X_n, n \geq 1\}$ jest zmienna losowa

$$R_r = X_{T_r}, \quad r \geq 1.$$

Zauważmy, że dla rozkładów ciągłych, nawet o ograniczonym nośniku, powyższa definicja daje w wyniku nieskończony ciąg różnych wartości rekordowych. Jeżeli natomiast rozkład obserwacji ma atom w prawym końcu nośnika (tzn. istnieje $x_0 < \infty$ takie, że $F(x_0) - F(x_0^-) > 0$ i $F(x_0) = 1$), to z danego ciągu obserwacji można z dodatnim prawdopodobieństwem uzyskać skończoną liczbę różnych rekordów. W szczególności rekordy nie są poprawnie zdefiniowane dla rozkładów dyskretnych o skończonych nośnikach. Aby ominąć tę trudność Vervaat [56] wprowadził pojęcie słabych wartości rekordowych.

Definicja 1.3. Dla ustalonej liczby naturalnej $r \geq 1$ zdefiniujmy ciąg słabych czasów rekordowych następująco

$$U_1 = 1, \quad U_r = \min\left\{ j > U_{r-1} : X_j \geq X_{U_r-1} \right\}, \quad r \geq 2.$$

Wówczas r-tą słabą wartością rekordową (inaczej słabym rekordem) ciągu $\{X_n, n \geq 1\}$ jest zmienna losowa

$$W_r = X_{U_r}, \quad r \geq 1.$$

Zatem obserwacja jest rekordem, jeżeli jest **większa** od wszystkim poprzednich wartości, a słabym rekordem, gdy jest **większa lub równa** niż poprzedni słaby rekord. W obu przypadkach pierwsza obserwacja jest pierwszym rekordem ($R_1 = W_1 = X_1$), jest to tak zwany rekord trywialny. W przypadku zmiennych losowych o rozkładzie ciągłym definicje rekordów i słabych rekordów są równoważne z prawdopodobieństwem 1, natomiast dla zmiennych losowych o rozkładzie dyskretnym tak nie jest. Wtedy bowiem kolejne słabe rekordy z dodatnim prawdopodobieństwem mogą być sobie równe, podczas gdy zwykłe rekordy tworzą prawie pewnie ciąg ściśle rosnący.

W niniejszej pracy rekordy i słabe rekordy dla ciągu $\{X_n, n \geq 1\}$ niezależnych zmiennych losowych o jednakowym rozkładzie dyskretnym będą rozpatrywane jako odrębna klasa modeli. Z tego powodu poniżej prezentujemy własności rozkładów dyskretnych wartości rekordowych. Załóżmy, że nośnik rozkładu zmiennej losowej X_1 jest postaci $S = \{0, 1, \ldots, N\}$, gdzie $N \leq \infty$. Niech $p_k = P(X_1 = k)$ oraz $q_k = P(X_1 \geq k) = \sum_{j=k}^{N} p_j$ dla $k \in S$.

Jak już wyjaśniono powyżej, w przypadku zwykłych rekordów może być brany pod uwagę tylko nieskończony nośnik $S = \{0, 1, \ldots\}$. Wówczas funkcja rozkładu prawdopodobieństwa wektora $(R_1, \ldots, R_n)$ dana jest wzorem

$$P(R_1 = j_1, \ldots, R_n = j_n) = p_{j_n} \prod_{k=1}^{n-1} \frac{p_{j_k}}{q_{j_k+1}}, \quad 0 \leq j_1 < \ldots < j_n < \infty.$$

Następnie zauważmy, że skoro $R_2 > R_1 \geq 0$, to $R_2 \geq 1$. Podobnie $R_n \geq n-1$. Zatem warunkowy rozkład prawdopodobieństwa R_{r+1} pod warunkiem $R_r = j$ jest dany wzorem

$$P(R_{r+1} = k \mid R_r = j) = \frac{p_k}{q_{j+1}}, \quad r-1 \leq j < k < \infty. \tag{1.4}$$

Dla słabych rekordów mamy $N \leq \infty$, a funkcja rozkładu prawdopodobieństwa wektora $(W_1, \ldots, W_n)$ jest postaci

$$P(W_1 = j_1, \ldots, W_n = j_n) = p_{j_n} \prod_{k=1}^{n-1} \frac{p_{j_k}}{q_{j_k}}, \quad 0 \leq j_1 \leq \ldots \leq j_n \leq N.$$

Natomiast warunkowy rozkład prawdopodobieństwa W_{r+1} pod warunkiem $W_r = j$ jest postaci

$$P(W_{r+1} = k \mid W_r = j) = \frac{p_k}{q_j}, \quad 0 \leq j \leq k \leq N. \tag{1.5}$$

Zauważmy na koniec tego paragrafu, że warunkowe rozkłady (1.4) oraz (1.5) nie zależą od r. Powyższą własność zachowują również warunkowe rozkłady niekolejnych rekordów i słabych rekordów.

1.1.3 Uogólnione statystyki porządkowe

Statystyki porządkowe, jak i wartości rekordowe oparte na rozkładach ciągłych można wyrazić w terminach tego samego ogólnego modelu, mianowicie modelu uogólnionych statystyk porządkowych. Model ten został zdefiniowany przez Kampsa [36, 37] w następujący sposób.

Definicja 1.4. Niech $n \in \mathbb{N}$, $k > 0$, $m_1, \ldots, m_{n-1} \in \mathbb{R}$, będą parametrami takimi, że

$$\gamma_j = k + n - j + \sum_{i=j}^{n-1} m_i > 0, \quad 1 \leq j \leq n-1.$$

Mówimy, że zmienne losowe $U_*^{(1)}, \ldots, U_*^{(n)}$ są uogólnionymi statystykami porządkowymi z rozkładu jednostajnego z parametrami $m_1, \ldots, m_{n-1}$ oraz k, jeżeli ich łączna funkcja gęstości prawdopodobieństwa jest postaci

$$f^{U_*^{(1)}, \ldots, U_*^{(n)}}(u_1, \ldots, u_n) = k \left(\prod_{i=1}^{n-1} \gamma_i \right) \left(\prod_{i=1}^{n-1} (1 - u_i)^{m_i} \right) \big(1 - u_n\big)^{k-1},$$

dla $0 < u_1 \leq \ldots \leq u_n < 1$ oraz 0 poza tym.

Jeżeli teraz F oznacza dowolną dystrybuantę, to mówimy, że zmienne losowe $X_*^{(1)}, \ldots, X_*^{(n)}$ są uogólnionymi statystykami porządkowymi z rozkładu F, jeżeli

$$X_*^{(r)} = F^{-1}\big(U_*^{(r)}\big), \quad r = 1, \ldots, n,$$

gdzie F^{-1} oznacza funkcję kwantylową rozkładu F określoną jednoznacznie równościami

$$F^{-1}(y) = \inf\left\{x \in \mathbb{R} : F(x) \geq y\right\}, \quad y \in (0, 1], \quad F^{-1}(0) = F^{-1}(0^+).$$

Stosujemy tu standardową notację $f(a^+)$ na oznaczenie granicy prawostronnej funkcji f w punkcie a.

Zauważmy, że jeżeli F jest dystrybuantą absolutnie ciągłą, której odpowiada funkcja gęstości f, to łączna funkcja gęstości prawdopodobieństwa zmiennych losowych $X_*^{(1)}, \ldots, X_*^{(n)}$ jest następująca

$$f^{X_*^{(1)}, \ldots, X_*^{(n)}}(x_1, \ldots, x_n) = k \left(\prod_{i=1}^{n-1} \gamma_i \right) \left(\prod_{i=1}^{n-1} \big(1 - F(x_i)\big)^{m_i} f(x_i) \right) \big(1 - F(x_n)\big)^{k-1} f(x_n), \tag{1.6}$$

dla $F^{-1}(0) < x_1 \leq \ldots \leq x_n < F^{-1}(1)$ oraz 0 poza tym.

Stosowana jest również inna definicja uogólnionych statystyk porządkowych z rozkładu jednostajnego wprowadzona przez Cramera i Kampsa [18]. Podali oni równoważną definicję przy użyciu iloczynów niezależnych zmiennych losowych o odpowiednich rozkładach beta. Zauważmy najpierw, że parametry $m_1, \ldots, m_{n-1} \in \mathbb{R}$ oraz $k > 0$ wyznaczają jednoznacznie wielkości $\gamma_1, \ldots, \gamma_{n-1}$. Z drugiej strony, jeśli znane są liczby $\gamma_1, \ldots, \gamma_n > 0$, to parametry $m_1, \ldots, m_{n-1}$ oraz k mogą być równoważnie określone jako

$$m_j = \gamma_j - \gamma_{j+1} - 1, \quad k = \gamma_n.$$

Dlatego możemy mówić od razu, że parametrami modelu uogólnionych statystyk porządkowych są dowolne liczby dodatnie $\gamma_1, \ldots, \gamma_n > 0$.

Definicja 1.5. Niech $n \in \mathbb{N}$, $\gamma_1, \ldots, \gamma_n > 0$ oraz niech $B_1, \ldots, B_n$ będą niezależnymi zmiennymi losowymi takimi, że B_j ma rozkład $\mathrm{Beta}(\gamma_j, 1)$, $1 \leq j \leq n$. Mówimy, że zmienne losowe

$U_*^{(1)}, \dots, U_*^{(n)}$ są uogólnionymi statystykami porządkowymi z rozkładu jednostajnego z parametrami $\gamma_1, \dots, \gamma_n$, jeżeli

$$U_*^{(r)} \stackrel{d}{=} 1 - \prod_{i=1}^{r} B_i, \quad 1 \le r \le n,$$

gdzie $\stackrel{d}{=}$ oznacza równość rozkładów zmiennych losowych.

Pojęcie uogólnionych statystyk porządkowych pozwala na ujednolicenie podejścia do wielu zagadnień probabilistycznych i statystycznych związanych z pozornie różnymi modelami uporządkowanych zmiennych losowych. Przykładowo zamiast dowodzić pewne własności rozkładów najpierw dla statystyk porządkowych, a następnie dla wartości rekordowych i innych modeli, można spróbować udowodnić je od razu dla uogólnionych statystyk porządkowych.

Powyższy model zawiera jako szczególne przypadki, wiele znanych modeli uporządkowanych zmiennych losowych. Zacznijmy od przypadku, gdy dystrybuanta F jest ciągła. Wówczas model uogólnionych statystyk porządkowych obejmuje następujące szczególne przypadki:

Statystyki porządkowe $X_{1:n} \le \dots \le X_{n:n}$ z próby losowej $(X_1, \dots, X_n)$ o rozkładzie F są uogólnionymi statystykami porządkowymi z parametrami $m_1 = \dots = m_{n-1} = 0$ oraz $k = 1$ (por. wzory (1.1) i (1.6)). Wtedy $\gamma_j = n - j + 1$, $1 \le j < n$.

Wartości rekordowe $R_1, \dots, R_n$ ciągu $\{X_n, n \ge 1\}$ niezależnych zmiennych losowych o jednakowym rozkładzie F są uogólnionymi statystykami porządkowymi z parametrami $m_1 = \dots = m_{n-1} = -1$ oraz $k = 1$. Wtedy $\gamma_j = 1$, $1 \le j < n$.

Statystyki porządkowe progresywnie cenzurowane typu II stanowią uogólnienie statystyk porządkowych. W sposób naturalny pojawiają się one w eksperymentach dotyczących długości czasu pracy (lub czasu życia), gdzie nie są obserwowane awarie wszystkich elementów biorących udział w teście. Na początku eksperymentu mamy do dyspozycji N obiektów $X_1, \dots, X_N$ oraz ustalamy liczbę $n \le N$ awarii, które chcemy zaobserwować. Następnie z pewnych powodów (np. ekonomicznych lub etycznych) dla $r = 1, \dots, n$, po r-tej awarii w eksperymencie usuwamy z niego z góry ustaloną liczbę R_r obiektów spośród tych, które jeszcze pracują. Liczby $R_1, \dots, R_n$ są ustalane na początku doświadczenia w taki sposób, że zachodzi równość $N = n + R_1 + \dots + R_n$. Czasy kolejnych awarii

$$X_{1:n,N} \le \dots \le X_{n:n,N}$$

nazywamy *statystykami porządkowymi progresywnie cenzurowanymi (typu II)*. Model ten jest szczegółowo omówiony w monografiach [6, 7]. Balakrishnan i in. [8] zauważyli, że jeżeli $X_1, \dots, X_N$ są niezależnymi zmiennymi losowymi o jednakowym rozkładzie F, to $X_{1:n,N}, \dots, X_{n:n,N}$ tworzą uogólnione statystyki porządkowe z rozkładu F z parametrami $m_j = R_j$ oraz $k = R_n + 1$. Wtedy $\gamma_j = n - j + 1 + R_j + \dots + R_n$, $1 \le j < n$.

Sekwencyjne statystyki porządkowe zostały wprowadzone w teorii niezawodności jako rozszerzenie zwykłych statystyk porządkowych. Opisują one czasy kolejnych awarii w sekwencyjnych układach typu k-spośród-n. Załóżmy, że danych jest n dystrybuant $F_1, \ldots, F_n$ oraz, że na początku eksperymentu obserwujemy czasy uszkodzeń n elementów, których czasy pracy są niezależnymi zmiennymi losowymi o jednakowym rozkładzie opisanym dystrybuantą F_1. Po pierwszej awarii w chwili $X^*_{1:n} = x_1$ pozostałe $n-1$ elementów zostaje zastąpione przez nowe, których czasy życia są również niezależne i o jednakowym rozkładzie, ale opisanym rozkładem F_2 uciętym z lewej strony w punkcie x_1. Procedura ta jest kontynuowana, a więc po i-tym uszkodzeniu w chwili $X^*_{i:n} = x_i$, pozostałe $n-i$ elementów jest zastąpionych nowymi, których rozkład czasu życia jest opisany przez dystrybuantę F_{i+1} uciętą z lewej strony w punkcie x_i. Otrzymane w ten sposób czasy awarii $X^*_{1:n}, \ldots, X^*_{n:n}$ nazywamy *sekwencyjnymi statystykami porządkowymi*. Zaznaczmy, iż z sekwencyjnymi statystykami porządkowymi będziemy mieć do czynienia, również gdy działające elementy nie zostaną zastąpione przez nowe z innymi intensywnościami awarii, ale ich intensywności awarii ulegną zmianie ponieważ ich obciążenie będzie większe po awarii jednego z elementów systemu. Formalną definicję tego modelu można znaleźć w pracach [36, 37]. Jeżeli $F_i(x) = 1 - [1 - F(x)]^{\alpha_i}$ dla $\alpha_i > 0$, $1 \leq i \leq n$ oraz pewnej ustalonej dystrybuanty F, to sekwencyjne statystyki porządkowe $X^*_{1:n}, \ldots, X^*_{n:n}$ są uogólnionymi statystykami porządkowymi z rozkładu F z parametrami $m_j = (n-j+1)\alpha_j - (n-j)\alpha_{j+1} - 1$ oraz $k = \alpha_n$. Ponadto $\gamma_j = (n-j+1)\alpha_j$, $1 \leq j < n$.

k-te wartości rekordowe są uogólnieniem wartości rekordowych. Model ten został wprowadzony przez Dziubdzielę i Kopocińskiego w roku 1976 (por. [27]). Dla dowolnego naturalnego $k \geq 1$ definiujemy k-te czasy rekordowe dla ciągu $\{X_n, n \geq 1\}$ następująco

$$T_1^{(k)} = 1, \quad T_{n+1}^{(k)} = \min\left\{ j > T_n^{(k)} : X_{j:j+k-1} > X_{T_n^{(k)} :\, T_n^{(k)}+k-1} \right\}, \quad n \geq 1.$$

Wtedy ciąg k-tych wartości rekordowych określamy jako

$$R_n^{(k)} = X_{T_n^{(k)} :\, T_n^{(k)}+k-1}, \quad n \geq 1.$$

Dla $k = 1$ dostajemy zwykłe rekordy opisane w paragrafie 1.1.2. Jeżeli $\{X_n, n \geq 1\}$ są niezależnymi zmiennymi losowymi o jednakowym rozkładzie F, to k-te wartości rekordowe $R_1^{(k)}, \ldots, R_n^{(k)}$ ciągu $\{X_n, n \geq 1\}$ są uogólnionymi statystykami porządkowymi z parametrami $m_1 = \ldots = m_{n-1} = -1$ oraz $k \in \mathbb{N}$. Ponadto $\gamma_j = k$, $1 \leq j < n$.

Wartości rekordowe Pfeifera występują w sytuacji, gdy obserwujemy wartości rekordowe z ciągu obserwacji o niejednakowych rozkładach. Zakładamy, że dany jest ciąg dystrybuant $\{F_n, n \geq 1\}$. W pierwszym kroku zakładamy, że wszystkie obserwacje mają ten sam rozkład F_1 i czekamy na pojawienie się pierwszego rekordu, który oznaczamy przez $X^{(1)}_{\Delta_1}$.

Następnie zmieniamy rozkład obserwacji na F_2 i czekamy na kolejny rekord $X_{\Delta_2}^{(2)}$ przekraczający pierwszy rekord. Ogólnie, po zaobserwowaniu n-tego w kolejności rekordu $X_{\Delta_n}^{(n)}$, $n \geq 1$, zmieniamy wspólny rozkład następnych obserwacji na F_{n+1} (por. [47]). Jeżeli $F_i(x) = 1 - [1 - F(x)]^{\beta_i}$, gdzie F jest pewną ciągłą dystrybuantą oraz $\beta_i > 0$, $i \geq 1$, to wartości rekordowe Pfeifera $X_{\Delta_1}^{(1)}, \ldots, X_{\Delta_n}^{(n)}$ są uogólnionymi statystykami porządkowymi z parametrami $m_j = \beta_j - \beta_{j+1} - 1$ oraz $k = \beta_n$. Ponadto $\gamma_j = \beta_j$, $1 \leq j < n$.

k_n-wartości rekordowe są kombinacją k-tych wartości rekordowych i wartości rekordowych Pfeifera. Dla danego ciągu liczb naturalnych $\{k_n, n \geq 1\}$, interesuje nas k_1 największa wartość (z rozkładu F_1), następnie k_2 największa wartość (z rozkładu F_2), która jest większa od poprzedniej itd. (por. [36]). Wartości te oznaczamy $X_{\Delta_1,k_1}^{(1)}, X_{\Delta_2,k_2}^{(2)}, \ldots$. W przypadku, gdy $k_n = 1$, $n \geq 1$, otrzymujemy wartości rekordowe Pfeifera. Jeżeli $F_i(x) = 1 - [1 - F(x)]^{\beta_i}$, gdzie F jest pewną ciągłą dystrybuantą oraz $\beta_i > 0$, $i \geq 1$, to k_n-rekordy $X_{\Delta_1,k_1}^{(1)}, \ldots, X_{\Delta_n,k_n}^{(n)}$ są uogólnionymi statystykami porządkowymi z parametrami $m_j = \beta_j k_j - \beta_{j+1} k_{j+1} - 1$ oraz $k = \beta_n k_n$. Wtedy $\gamma_j = \beta_j k_j$, $1 \leq j < n$.

Uwaga 1.1. W przypadku rozkładów dyskretnych znanych jest znacznie mniej szczególnych przypadków uogólnionych statystyk porządkowych. Tran w pracy [55] wskazała tylko jeden model jako podmodel uogólnionych statystyk porządkowych z rozkładu dyskretnego, mianowicie dyskretne statystyki porządkowe. Jak dotąd wiadomo również, że progresywnie cenzurowane statystyki porządkowe typu II należą do tej klasy (patrz [9], Rozdział 3). Zaznaczmy, że w ogólnym przypadku nie należą do tej klasy ani dyskretne wartości rekordowe, ani słabe wartości rekordowe, które jako odrębna klasa modeli będą przedmiotem naszych rozważań w rozdziale 4.

Na zakończenie tego paragrafu przypomnimy znane własności rozkładów uogólnionych statystyk porządkowych. Cramer i Kamps [18] pokazali, że funkcja gęstości r-tej uogólnionej statystyki porządkowej z rozkładu jednostajnego $U_*^{(r)}$, $1 \leq r \leq n$, (tj. gdy $F(x) = x$ dla $x \in (0,1)$) może być zapisana w postaci

$$f_{*,r}(x) = c_{r-1} \mathrm{G}_{r,r}^{r,0}\left(1 - x \,\middle|\, \begin{matrix} \gamma_1, \ldots, \gamma_r \\ \gamma_1 - 1, \ldots, \gamma_r - 1 \end{matrix}\right), \quad x \in (0,1),$$

gdzie $c_{r-1} = \prod_{j=1}^{r} \gamma_j$ oraz

$$\mathrm{G}_{r,r}^{r,0}\left(s \,\middle|\, \begin{matrix} \gamma_1, \ldots, \gamma_r \\ \gamma_1 - 1, \ldots, \gamma_r - 1 \end{matrix}\right) = \frac{1}{2\pi i} \int_L \frac{s^z}{\prod_{j=1}^{r} (\gamma_j - 1 - z)} \mathrm{d}z \tag{1.7}$$

oznacza szczególną G-funkcję Meijera, a L jest odpowiednio dobranym konturem całkowania. Wiele własności G-funkcji można znaleźć w monografii [42] (patrz Rozdział 3). Dla uproszczenia, w dalszej części pracy będziemy pisać $G_r(x \mid \gamma_1, \ldots, \gamma_r)$ zamiast $\mathrm{G}_{r,r}^{r,0}\left(1 - x \,\middle|\, \begin{smallmatrix} \gamma_1, \ldots, \gamma_r \\ \gamma_1 - 1, \ldots, \gamma_r - 1 \end{smallmatrix}\right)$.

Uwaga 1.2. Równość (1.7) implikuje, że wartości funkcji $f_{*,r}$ nie zależą od porządku parametrów $\gamma_1, \ldots, \gamma_r$ (zob. [11], Uwaga 2).

Uwaga 1.3. W dalszym ciągu pracy będziemy stosować następującą konwencję i notację. Niech F oznacza dystrybuantę, $\bar{F} = 1 - F$ oznacza jej funkcję przeżycia oraz P_F oznacza miarę probabilistyczną na $\mathbb{R}$ wyznaczoną przez F. Zwykle, przez nośnik miary P_F w literaturze określa się najmniejszy zbiór borelowski $A \subset \mathbb{R}$ taki, że $P_F(A) = 1$. Jednakże w niniejszej pracy w przypadku dystrybuant ciągłych na $\mathbb{R}$ dla uproszczenia nośnikiem F będziemy nazywać przedział (α, β), gdzie

$$\alpha = \alpha(F) = \inf\{x \in \mathbb{R} : F(x) > 0\}, \quad \beta = \beta(F) = \sup\{x \in \mathbb{R} : F(x) < 1\}.$$

Inaczej będziemy również pisać, że F jest rozkładem na przedziale (α, β). W szczególności dopuszczamy więc możliwość, że F jest funkcją stałą na pewnym właściwym podprzedziale przedziału (α, β).

Jeżeli dystrybuanta F jest ciągła, to dla $\ell \geq 1$ gęstość $X_*^{(r+\ell)}$ pod warunkiem $X_*^{(r)} = x$ względem miary P_F może być przedstawiona w postaci

$$f^{X_*^{(r+\ell)}|X_*^{(r)}}(t \mid x) = \frac{c_{r+\ell-1}}{c_{r-1}} G_\ell\left(F_x(t) \mid \gamma_{r+1}, \ldots, \gamma_{r+\ell}\right) \frac{1}{\bar{F}(x)} I_{(x,\beta)}(t), \quad x, t \in \mathbb{R}, \tag{1.8}$$

gdzie funkcja I_A jest indykatorem zbioru A oraz dla $x \in [\alpha, \beta)$

$$F_x(t) = \begin{cases} \frac{F(t) - F(x)}{\bar{F}(x)}, & \text{gdy } t \geq x, \\ 0, & \text{gdy } t < x, \end{cases}$$

oznacza uciętą lewostronnie dystrybuantę F (por. [19]).

Zauważmy, że zgodnie z Uwagą 1.2 warunkowa funkcja gęstości (1.8) nie zależy od kolejności parametrów $\gamma_{r+1}, \ldots, \gamma_{r+\ell}$. Bez zmniejszania ogólności rozważań możemy założyć np. że $\gamma_1 \leq \ldots \leq \gamma_r$ oraz $\gamma_{r+1} \leq \ldots \leq \gamma_{r+\ell}$. Trzeba natomiast zwracać uwagę na to, czy $\gamma_{r+1} < \gamma_r$ lub $\gamma_{r+1} \geq \gamma_r$ itp.

Zaznaczmy również, że rozkłady brzegowe oraz warunkowe uogólnionych statystyk porządkowych wyrażają się przez skomplikowane funkcje specjalne, co znacznie utrudnia badanie ich własności. Ominięcie tego problemu jest jednym z celów niniejszej rozprawy.

1.2 Problem charakteryzacji rozkładów prawdopodobieństwa przez regresje modeli uporządkowanych zmiennych losowych

Jednym z podstawowych problemów statystyki jest wyznaczenie nieznanego rozkładu prawdopodobieństwa na podstawie obserwacji danych statystycznych. W klasycznym podejściu obserwowane są wartości próby losowej $(X_1, \ldots, X_n)$ rozmiaru n ciągu niezależnych zmiennych losowych o jednakowym rozkładzie określonym przez dystrybuantę F. Następnie próba stanowi podstawę wnioskowania statystycznego, estymacji parametrów, czy też testowania hipotez statystycznych. Mniej standardowym podejściem jest analiza różnych własności rozkładu, które determinują albo pewne rodziny rozkładów, albo w niektórych przypadkach określają dokładną postać szukanej dystrybuanty F. Liczne przykłady charakteryzacji różnych rozkładów prawdopodobieństwa można znaleźć np. w monografii [32].

Wiele prac zostało poświęconych problemowi charakteryzacji rozkładów prawdopodobieństwa przy użyciu własności statystyk porządkowych $X_{1:n} \leq \ldots \leq X_{n:n}$ z próby o rozmiarze n. Jednym z takich problemów jest charakteryzacja przez równanie regresji statystyk porządkowych. Równanie regresji, wokół którego będziemy się koncentrować ma postać

$$E\big(h(X_{s:n}) \mid X_{r:n} = x\big) = \xi(x), \tag{1.9}$$

dla ustalonych $1 \leq r < s \leq n$, gdzie h oraz ξ są znanymi funkcjami określonymi na nośniku bazowego rozkładu F. Problem charakteryzacyjny tego typu można wyrazić za pomocą pytania: czy znajomość funkcji ξ wystarcza do jednoznacznego wyznaczenia samego rozkładu F? Jako pierwszy tym problemem zajął się Ferguson, który w 1967 roku w pracy [28], dla $s = r+1$ oraz $h(x) = x$ scharakteryzował rozkłady wykładniczy, potęgowy i Pareto jako jedyne rozkłady absolutnie ciągłe określone przez liniową regresję $\xi(x) = ax + b$. Od tamtego czasu problem ten był tematem wielu prac i monografii. Na szczególną uwagę zasługuje zwłaszcza praca Dembińskiej i Wesołowskiego [25], w której autorzy uogólnili wynik Fergusona dla regresji niekolejnych statystyk porządkowych, a więc gdy $s \geq r+1$. Franco i Ruiz [29, 31] podobnie jak Ferguson rozważali regresję kolejnych statystyk porządkowych, $s = r+1$, ale dla ogólnego przypadku h i ξ, nawet obejmującego nieciągły rozkład F. Obszerny przegląd wyników dla tego zagadnienia można znaleźć na przykład w monografii [7]. Okazuje się, że wiele wyników dla statystyk porządkowych ma swoje odpowiedniki dla wartości rekordowych $R_1, \ldots, R_n$ ciągu $\{X_n, n \geq 1\}$ niezależnych zmiennych losowych o jednakowym rozkładzie. W literaturze zajmuje swoje miejsce analogiczny do (1.9) problem charakteryzacji rozkładów prawdopodobieństwa przez regresję wartości rekordowych

$$E\big(h(R_s) \mid R_r = x\big) = \xi(x). \tag{1.10}$$

Problem charakteryzacyjny tej postaci jako pierwszy rozważał Nagaraja. W pracy [43] scharakteryzował on przez liniową regresję kolejnych wartości rekordowych te same rozkłady co Ferguson. Wynik ten został uogólniony w pracy [26] na przypadek niekolejnych wartości rekordowych. Rozważania dotyczące dowolnej funkcji regresji kolejnych wartości rekordowych zostały przedstawione w pracy [30]. Jednakże dla przypadku niekolejnych wartości rekordowych podobnie jak dla niekolejnych statystyk porządkowych problem charakteryzacji przez równania (1.9), jak i (1.10) jest nadal otwarty.

Jak już wiemy statystyki porządkowe i wartości rekordowe oparte na rozkładach ciągłych są szczególnymi przypadkami tego samego modelu uogólnionych statystyk porządkowych. Zatem problemy charakteryzacyjne opisane powyżej można zunifikować. Dla ustalonych $r, \ell \geq 1$, niech $X_*^{(r)}, X_*^{(r+\ell)}$ będą uogólnionymi statystykami porządkowymi opartymi na rozkładzie określonym przez ciągłą dystrybuantę F o nośniku na przedziale (α, β), gdzie $-\infty \leq \alpha < \beta \leq \infty$. Załóżmy, że $h\colon (\alpha, \beta) \to \mathbb{R}$ jest funkcją ściśle rosnącą, ciągłą oraz taką, że $E\left|h\big(X_*^{(r+\ell)}\big)\right| < \infty$. Określmy funkcję regresji $h\big(X_*^{(r+\ell)}\big)$ względem $X_*^{(r)}$ jako funkcję $\xi : (\alpha, \beta) \to \mathbb{R}$ daną wzorem

$$\xi(x) = E\left(h\big(X_*^{(r+\ell)}\big) \;\middle|\; X_*^{(r)} = x\right), \quad x \in (\alpha, \beta). \tag{1.11}$$

Oczywiście równość ta jest bezpośrednim uogólnieniem zależności (1.9) i (1.10). Co więcej, lewa strona równania (1.11) zależy od r oraz ℓ, jak również od parametrów $\gamma_{r+1}, \ldots, \gamma_{r+\ell}$, ale aby uniknąć skomplikowanych oznaczeń opuszczamy te zależności w notacji.

Uwaga 1.4. Zauważmy, że wszystkie wyniki otrzymane przy założeniu, że funkcja h jest ściśle rosnąca mogą być prosto przeniesione na przypadek, gdy h jest funkcją ściśle malejącą. Wówczas wystarczy pomnożyć (1.11) przez -1 i dalej rozważać ściśle rosnącą funkcję $-h$ zamiast h oraz $-\xi$ zamiast ξ.

Cramer i in. [19] udowodnili, że dla każdej ciągłej dystrybuanty F istnieje jednoznacznie określona, ściśle rosnąca i ciągła funkcja regresji (1.11). Może być ona obliczona jako wartość oczekiwana względem regularnej wersji odpowiadającego rozkładu warunkowego $X_*^{(r+\ell)}$ pod warunkiem $X_*^{(r)} = x$. W niniejszej pracy rozważamy problem odwrotny, to znaczy problem jednoznacznej charakteryzacji rozkładu F przez znajomość pojedynczej funkcji regresji ξ określonej wzorem (1.11).

Tak określony problem jest kompletnie rozwiązany w przypadku $\ell = 1$, a więc dla regresji kolejnych uogólnionych statystyk porządkowych. Wówczas F można wyznaczyć znając ξ i h w następujący sposób (por. [19])

$$F(x) = 1 - \exp\left(-\frac{1}{\gamma_{r+1}} \int_\alpha^x \frac{\mathrm{d}\xi(t)}{\xi(t) - h(t)}\right), \quad x \in (\alpha, \beta).$$

Bieniek w pracy [11] udowodnił, że dystrybuanta F jest jednoznacznie określona również przez regresję dualną tj. $h\big(X_*^{(r)}\big)$ względem $X_*^{(r+1)}$. Jednakże, podejście przyjęte dla przypadku $\ell = 1$ nie może zostać rozszerzone na przypadek dowolnego $\ell > 1$, nawet dla szczególnych modeli uogólnionych statystyk porządkowych jak statystyki porządkowe czy wartości rekordowe.

Dla $\ell \geq 2$, gdy $h(x) = x$ oraz ξ jest funkcją liniową postaci $\xi(x) = ax + b$, to albo $a \in (0, 1)$ i F jest jednoznacznie określonym rozkładem potęgowym, albo $a = 1$ i F jest rozkładem wykładniczym, albo $a > 1$ i F jest rozkładem Pareto, patrz [16, 19]. Niestety metoda dowodu wprowadzona przez Dembińską i Wesołowskiego w pracach [25, 26], wykorzystująca rozwiązanie tak zwanego scałkowanego równania funkcyjnego Cauchy'ego, nie może zostać zastosowana w przypadku nieliniowej funkcji ξ. W przypadku dowolnej regresji ξ podejście to prowadzi do równania funkcyjnego, którego rozwiązanie jest nieznane. Jednakże, Bieniek w pracy [12] wykazał, że do jednoznacznej charakteryzacji rozkładu F wystarcza znajomość dwóch funkcji regresji

$$\xi(x) = E\left(h\big(X_*^{(r+\ell)}\big) \mid X_*^{(r)} = x\right), \quad \xi^*(x) = E\left(h\big(X_*^{(r+\ell)}\big) \mid X_*^{(r+1)} = x\right).$$

Wówczas dystrybuanta F jest jednoznacznie wyznaczona wzorem

$$F(x) = 1 - \exp\left(-\frac{1}{\gamma_{r+1}}\int_\alpha^x \frac{\mathrm{d}\xi(t)}{\xi(t)-\xi^*(t)}\right), \quad x \in (\alpha,\beta).$$

Powstaje stąd pytanie czy można wyznaczyć ξ^*, gdy znane są funkcje ξ i h. Pytanie to stanowi punkt wyjścia do naszych dalszych badań nad problemem charakteryzacyjnym zawartych w niniejszej rozprawie.

W kolejnych rozdziałach zaprezentujemy nowe podejście do problemu charakteryzacyjnego w oparciu o własność Markowa rozważanych modeli uporządkowanych zmiennych losowych. W przypadku rozkładów absolutnie ciągłych o ciągłej gęstości wykażemy, że jednoznaczność charakteryzacji bazowej funkcji rozkładu F przez regresję uogólnionych statystyk porządkowych (1.11) jest równoważna z jednoznacznością rozwiązania odpowiedniego układu $\ell - 1$ równań różniczkowych spełniającego pewne warunki ograniczające (zob. rozdział 2). W szczególności, dla $\ell = 2$ regresja ξ determinuje F jednoznacznie wtedy i tylko wtedy, gdy równanie różniczkowe

$$y' = \frac{\gamma_{r+2}}{\gamma_{r+1}}\frac{y-h(x)}{\xi(x)-y}\xi'(x) \tag{1.12}$$

ma dokładnie jedno rozwiązanie φ takie, że nierówność

$$h(x) < \varphi(x) < \xi(x)$$

zachodzi dla $x \in (\alpha,\beta)$, całka

$$I(x) = \int_\alpha^x \frac{\xi'(t)}{\xi(t)-\varphi(t)}\,\mathrm{d}t$$

jest zbieżna dla każdego $x \in (\alpha,\beta)$ oraz rozbieżna dla $x \to \beta^-$, a ponadto

$$\lim_{x\to\beta^-} \xi(x)\exp\left(-\int_\alpha^x \frac{\xi'(t)}{\xi(t)-\varphi(t)}\,\mathrm{d}t\right) = 0.$$

Wówczas dystrybuanta F jest jednoznacznie wyznaczona wzorem

$$F(x) = 1 - \exp\left(-\frac{1}{\gamma_{r+1}}\int_\alpha^x \frac{\xi'(t)}{\xi(t)-\varphi(t)}\,\mathrm{d}t\right).$$

W przypadku rozkładów ciągłych (dystrybuanta F jest tylko ciągła, a odpowiadająca jej funkcja gęstości f jest albo nieciągła albo nie istnieje), w naszym kryterium równania różniczkowe zostają zastąpione przez odpowiednie równania całkowe (zob. rozdział 3).

W niniejszej pracy będziemy również rozważać analogiczny problem charakteryzacyjny dla rozkładów dyskretnych. Niestety okazuje się, że w przypadku takich rozkładów wyżej opisane podejście nie może być zastosowane w modelu uogólnionych statystyk porządkowych. Może natomiast być efektywnie użyte w modelu słabych wartości rekordowych.

Dla ustalonych $r,\ell \geq 1$ rozważmy słabe wartości rekordowe $W_r, W_{r+\ell}$ z rozkładu o niezdegenerowanej dystrybuancie F określonej na przeliczalnym nośniku $S \subset \mathbb{R}$ takim, że elementy zbioru S tworzą ciąg ściśle rosnący $\alpha_0 < \alpha_1 < \ldots < \alpha_N$, gdzie $N \leq \infty$. Wtedy rozważamy

$$\xi(x) = E\big(h(W_{r+\ell}) \mid W_r = x\big), \quad x \in S, \tag{1.13}$$

gdzie $h\colon S \to \mathbb{R}$ jest funkcją ściśle rosnącą i taką, że $E\big|h(W_{r+\ell})\big| < \infty$. Powyższy problem był najczęściej rozpatrywany dla funkcji $h(x) = x$ oraz liniowej funkcji regresji $\xi(x) = ax+b$. Wiadomo, że wówczas są jednoznacznie charakteryzowane trzy rodziny rozkładów prawdopodobieństwa tj. rozkład geometryczny (gdy $a = 1$), ujemny rozkład hipergeometryczny pierwszego rodzaju (gdy $0 < a < 1$) oraz ujemny rozkład hipergeometryczny drugiego rodzaju (gdy $a > 1$). W przypadku liniowej regresji kolejnych słabych rekordów ($\ell = 1$) częściowy wynik otrzymał Stepanov [54], a uzupełnili go Wesołowski i Ahsanullah [57], którzy udowodnili również, że te same trzy rozkłady otrzymamy, jeśli założymy liniowość regresji słabych rekordów o odstępie dwa ($\ell = 2$). W ogólnym przypadku liniowym ($\ell \geq 1$) López-Blázquez [41] zaproponował podejście polegające na redukcji problemu do przypadku regresji kolejnych słabych rekordów, którego rozwiązanie jest znane. Jednakże, jak zauważyli Karczewski i Wesołowski [38], to podejście nie działa w przypadku, gdy $\ell \geq 5$ oraz $N = \infty$. Zatem problem charakteryzacji przez liniowość regresji słabych wartości rekordowych nie jest całkowicie rozwiązany. Nieliniowe regresje W_{r+1} względem W_r oraz W_{r+2} względem W_r były natomiast rozpatrywane przez Alieva odpowiednio w pracach [2] oraz [3]. W obu pracach autor zakładał, że nośnik bazowego rozkładu jest nieskończony, $N = \infty$.

Zaznaczmy w tym miejscu, że w przeciwieństwie do regresji słabych rekordów regresje zwykłych rekordów z rozkładów dyskretnych charakteryzują tylko ogony rozkładów prawdopodobieństwa. Problem charakteryzacyjny przez liniową regresję R_{r+1} względem R_r badali Srivastava [53] oraz Korwar [40], którzy otrzymali charakteryzacje ogona rozkładu geometrycznego oraz ogona ujemnego rozkładu hipergeometrycznego drugiego rodzaju. Powiązane rozważania można również znaleźć w monografii [4] (patrz Rozdział 4.6). Obszerny przegląd wyników charakteryzacji rozkładów dyskretnych z wykorzystaniem rekordów oraz słabych rekordów można znaleźć w pracy [24].

W przypadku charakteryzacji dyskretnych rozkładów prawdopodobieństwa przez regresje słabych wartości rekordowych w naszym kryterium zamiast (1.12) pojawiają się odpowiednie równania różnicowe (zob. rozdział 4).

1.3 Własność Markowa i rekurencyjna struktura regresji

W dalszej części pracy kluczową rolę odgrywa własność Markowa rozważanych modeli uporządkowanych zmiennych losowych.

Definicja 1.6. Mówimy, że ciąg zmiennych losowych $\{Y_n, n \geq 1\}$ o takim samym zbiorze wartości $\mathscr{Y}$ ma własność Markowa, jeżeli dla każdego $r \geq 1$ rozkład warunkowy Y_{r+1} pod warunkiem $Y_1, \ldots, Y_r$ jest funkcją wyłącznie zmiennej Y_r

$$P\big(Y_{r+1} \leq y \mid Y_1, \ldots, Y_r\big) = P\big(Y_{r+1} \leq y \mid Y_r\big), \quad y \in \mathscr{Y}.$$

Taki ciąg nazywamy łańcuchem Markowa.

W szczególności, z powyższej definicji otrzymujemy następującą ważną własność. Jeśli ciąg $\{Y_n, n \geq 1\}$ spełnia własność Markowa, to dla ustalonych $1 \leq s < r$ oraz dowolnej funkcji mierzalnej $h\colon \mathscr{Y} \to \mathbb{R}$ takiej, że $E\big|h(Y_{r+1})\big| < \infty$ zachodzi

$$E\big(h(Y_{r+1}) \mid Y_s, Y_r\big) = E\big(h(Y_{r+1}) \mid Y_r\big). \tag{1.14}$$

Uwaga 1.5. Oczywiście powyższa równość zachodzi prawie pewnie. W dalszej części pracy będą również pojawić się równości warunkowych wartości oczekiwanych tego typu. O wszystkich będziemy milcząco zakładać, że zachodzą z prawdopodobieństwem 1.

Kamps w pracy [36] (patrz Lemat 3.2) pokazał, że uogólnione statystyki porządkowe z rozkładu absolutnie ciągłego posiadają własność Markowa. Wynik ten został uogólniony przez Keseling [39] (patrz Twierdzenie 3.1) na przypadek rozkładów ciągłych.

Twierdzenie 1.1. *Jeżeli F jest dystrybuantą rozkładu ciągłego określonego na nośniku (α, β), to dla dowolnego $r \geq 2$ warunkowy rozkład $X_*^{(r+1)}$ pod warunkiem $X_*^{(1)}, \ldots, X_*^{(r)}$ jest taki sam jak warunkowy rozkład $X_*^{(r+1)}$ pod warunkiem $X_*^{(r)}$*

$$P\left(X_*^{(r+1)} > y \,\middle|\, X_*^{(r)} = x\right) = \left(\frac{\bar{F}(y)}{\bar{F}(x)}\right)^{\gamma_{r+1}}, \quad y \geq x \geq \alpha.$$

Inaczej wygląda sytuacja w przypadku uogólnionych statystyk porządkowych z rozkładu dyskretnego. Takie modele w ogólności nie tworzą łańcuchów Markowa. Wiemy, że należą do tej klasy dwa szczególne modele: statystyki porządkowe i progresywnie cenzurowane statystyki porządkowe typu II. Brak własności Markowa statystyk porządkowych z dyskretnego rozkładu o nośniku zawierającym co najmniej trzy atomy pokazał Nagaraja w pracy [45]. Odpowiednie rozważania można znaleźć również w monografii [5] (patrz Rozdział 3.4). Analogiczny wynik dla progresywnie cenzurowanych statystyk porządkowych typu II otrzymali Balakrishnan i Dembińska w pracy [9] (patrz Twierdzenie 3.1.1). W ogólności, Cramer i Tran w [21] wykazali, że własność Markowa nie zachodzi dla uogólnionych statystyk porządkowych z rozkładu mającego atomy.

Z drugiej strony zarówno dyskretne wartości rekordowe, jak i dyskretne słabe wartości rekordowe tworzą łańcuchy Markowa o prawdopodobieństwach przejścia odpowiednio (1.4) i (1.5). Oznacza to oczywiście, że rekordy z rozkładów dyskretnych nie są zawarte w modelu dyskretnych uogólnionych statystyk porządkowych. Jest to główna przyczyna, dla której w rozważaniach dotyczących rozkładów dyskretnych ograniczamy się do słabych wartości rekordowych.

Udowodnimy teraz lemat będący najważniejszą obserwacją, na której bazuje nowe podejście do problemu charakteryzacyjnego zaprezentowane w niniejszej pracy. W lemacie przedstawiona została rekurencyjna zależność między odpowiednimi funkcjami regresji.

Lemat 1.1. *Niech $\{Y_n, n \geq 1\}$ będzie ciągiem zmiennych losowych określonych na nośniku $\mathscr{Y}$ i spełniającym własność Markowa. Ustalmy $r \geq 1$, $\ell \geq 2$ oraz niech $\varphi_\ell \colon \mathscr{Y} \to \mathbb{R}$ będzie daną funkcją mierzalną taką, że $E\big|\varphi_\ell\big(Y_{r+\ell}\big)\big| < \infty$. Ponadto dla $i = 0, 1, \ldots, \ell-1$ zdefiniujmy funkcje $\varphi_{\ell-1}, \ldots, \varphi_1, \varphi_0$ rekurencyjnie*

$$\varphi_i(x) = E\big(\varphi_{i+1}(Y_{r+i+1}) \,\big|\, Y_{r+i} = x\big), \quad x \in \mathscr{Y}. \tag{1.15}$$

Wówczas

$$E\big(\varphi_\ell(Y_{r+\ell}) \,\big|\, Y_r = x\big) = \varphi_0(x), \quad x \in \mathscr{Y}.$$

Dowód. Dowód jest indukcyjny ze względu na $\ell \geq 2$. Dla $\ell = 2$ ustalmy funkcję $\varphi_2 \colon \mathscr{Y} \to \mathbb{R}$ taką, że $E\big|\varphi_2\big(Y_{r+2}\big)\big| < \infty$ i zdefiniujmy

$$\varphi_1(x) = E\big(\varphi_2(Y_{r+2}) \,\big|\, Y_{r+1} = x\big), \quad \varphi_0(x) = E\big(\varphi_1(Y_{r+1}) \,\big|\, Y_r = x\big),$$

dla $x \in \mathscr{Y}$. Wówczas korzystając z dobrze znanej własności warunkowej wartości oczekiwanej

$$\mathscr{G}_1 \subset \mathscr{G}_2 \Rightarrow E\left[E\left(X \mid \mathscr{G}_2\right) \mid \mathscr{G}_1\right] = E\left(X \mid \mathscr{G}_1\right), \tag{1.16}$$

przyjmując $\mathscr{G}_1 = \sigma(Y_r)$ i $\mathscr{G}_2 = \sigma(Y_r, Y_{r+1})$ otrzymujemy

$$E\big(\varphi_2(Y_{r+2}) \,\big|\, Y_r\big) = E\big[E\big(\varphi_2(Y_{r+2}) \,\big|\, Y_r, Y_{r+1}\big) \,\big|\, Y_r\big]. \tag{1.17}$$

Następnie z własności Markowa, dostajemy

$$E\big(\varphi_2(Y_{r+2}) \,\big|\, Y_r, Y_{r+1}\big) = E\big(\varphi_2(Y_{r+2}) \,\big|\, Y_{r+1}\big) = \varphi_1(Y_{r+1}). \tag{1.18}$$

Wstawiając (1.18) do prawej strony równości (1.17) mamy

$$E\big(\varphi_2(Y_{r+2}) \,\big|\, Y_r\big) = E\big(\varphi_1(Y_{r+1}) \,\big|\, Y_r\big) = \varphi_0(Y_r),$$

co kończy dowód lematu dla $\ell = 2$. Dla $\ell \geq 2$ analogicznie dowodzimy, że

$$\begin{aligned} E\big(\varphi_{\ell+1}(Y_{r+\ell+1}) \,\big|\, Y_r\big) &= E\big[E\big(\varphi_{\ell+1}(Y_{r+\ell+1}) \,\big|\, Y_r, Y_{r+\ell}\big) \,\big|\, Y_r\big] \\ &= E\big[E\big(\varphi_{\ell+1}(Y_{r+\ell+1}) \,\big|\, Y_{r+\ell}\big) \,\big|\, Y_r\big] \\ &= E\big(\varphi_\ell(Y_{r+\ell}) \,\big|\, Y_r\big), \end{aligned}$$

gdzie pierwsza równość wynika z własności (1.16), druga z własności Markowa (1.14), a trzecia wprost w definicji funkcji φ_ℓ (1.15). Powyższa równość, zgodnie z zasadą indukcji matematycznej kończy dowód lematu. □

Z Twierdzenia 1.1 i Lematu 1.1 otrzymujemy następujący wniosek.

Wniosek 1.1. Niech $X_*^{(1)},\dots,X_*^{(n)}$ będą uogólnionymi statystykami porządkowymi z rozkładu określonego przez ciągłą dystrybuantę F o nośniku na przedziale (α,β). Ustalmy $r \geq 1$, $\ell \geq 2$ takie, że $r+\ell \leq n$ oraz niech $\varphi_\ell \colon (\alpha,\beta) \to \mathbb{R}$ będzie funkcją mierzalną taką, że $E\left|\varphi_\ell\big(X_*^{(r+\ell)}\big)\right| < \infty$. Ponadto dla $i = 0,1,\dots,\ell-1$ zdefiniujmy $\varphi_{\ell-1},\dots,\varphi_1,\varphi_0$ rekurencyjnie

$$\varphi_i(x) = E\left(\varphi_{i+1}\big(X_*^{(r+i+1)}\big) \mid X_*^{(r+i)} = x\right), \quad x \in (\alpha,\beta). \tag{1.19}$$

Wówczas

$$E\left(\varphi_\ell\big(X_*^{(r+\ell)}\big) \mid X_*^{(r)} = x\right) = \varphi_0(x), \quad x \in (\alpha,\beta). \tag{1.20}$$

Analogiczny wniosek można wypowiedzieć dla dyskretnych wartości rekordowych i słabych wartości rekordowych, które jak już zauważyliśmy powyżej tworzą łańcuchy Markowa.

Podsumowując, struktura regresji uogólnionych statystyk porządkowych jest taka, że regresja funkcji niekolejnych uogólnionych statystyk porządkowych jest równa odpowiednio zmodyfikowanej regresji kolejnych zmiennych w tym modelu.

1.4 Własności regresji modeli uporządkowanych zmiennych losowych

1.4.1 Własności regresji uogólnionych statystyk porządkowych

Zbadamy najpierw własności regresji kolejnych uogólnionych statystyk porządkowych z rozkładu ciągłego. Załóżmy, że zachodzi relacja

$$\xi_1(x) = E\left(h\big(X_*^{(r+1)}\big) \mid X_*^{(r)} = x\right), \quad x \in (\alpha,\beta). \tag{1.21}$$

Z własności G-funkcji Meijera mamy $G_1(x \mid \gamma) = (1-x)^{\gamma-1}$. Zatem z równości (1.8) dla $\ell = 1$ otrzymujemy następującą postać funkcji regresji

$$\xi_1(x) = \frac{\gamma_{r+1}}{\bar{F}(x)^{\gamma_{r+1}}} \int_x^\beta h(t)\bar{F}(t)^{\gamma_{r+1}-1}\,\mathrm{d}F(t). \tag{1.22}$$

Korzystając z powyższej formuły wykażemy ważne własności funkcji regresji (1.21).

Lemat 1.2. *Dla ustalonego $r \geq 1$, niech $X_*^{(r)},X_*^{(r+1)}$ będą uogólnionymi statystykami porządkowymi z ciągłego rozkładu o dystrybuancie F określonej na przedziale (α,β). Załóżmy, że $h\colon (\alpha,\beta) \to \mathbb{R}$ jest funkcją ściśle rosnącą, ciągłą i taką, że $E\left|h\big(X_*^{(r+1)}\big)\right| < \infty$ oraz niech funkcja regresji $\xi_1\colon (\alpha,\beta) \to \mathbb{R}$ będzie określona wzorem* (1.21). *Wówczas ξ_1 ma następujące własności:*

(i) ξ_1 *jest ciągła na* (α,β)*;*

(ii) jeżeli F *jest różniczkowalna w sposób ciągły na* (α,β)*, to* ξ_1 *jest różniczkowalna na tym przedziale;*

(iii) $\xi_1(x) > h(x)$ *dla wszystkich* $x \in (\alpha,\beta)$*;*

(iv) ξ_1 *jest rosnąca, a ponadto jest ona stała na przedziale* $I \subset (\alpha,\beta)$ *wtedy i tylko wtedy, gdy* F *jest stała na tym przedziale;*

(v) jeżeli $h(\beta) = \lim_{x\to\beta^-} h(x) < \infty$*, to* $\xi_1(\beta) = h(\beta)$*.*

Dowód. W dowodzie użyjemy argumentów podobnych do [31] (por. Lemat 3.4 tam zawarty). Ustalmy $r \in \mathbb{N}$ i przyjmijmy dla uproszczenia zapisu $\gamma = \gamma_{r+1}$. Z równości (1.22) oraz z podstawowego twierdzenia rachunku całkowego otrzymujemy, że ξ_1 jest funkcją ciągłą. Co więcej, jeżeli F jest różniczkowalna na (α,β), to ξ_1 jest również różniczkowalna na tym przedziale.

Zauważmy, że równanie (1.22) można zapisać równoważnie w postaci

$$\xi_1(x) = -\int_x^\beta h(t)\,\mathrm{d}\left(\frac{\bar{F}(t)}{\bar{F}(x)}\right)^\gamma = -\int_x^\beta h(t)\,\mathrm{d}G(t), \tag{1.23}$$

gdzie $G(t) = \left(\frac{\bar{F}(t)}{\bar{F}(x)}\right)^\gamma$. Z twierdzenia o wartości średniej dla całki Riemanna-Stieltjesa istnieje punkt $\theta \in (x,\beta)$ taki, że

$$\xi_1(x) = -h(\theta)[G(\beta) - G(x)] = h(\theta),$$

gdyż $G(\beta) = 0$ oraz $G(x) = 1$. Stąd i z monotoniczności funkcji h otrzymujemy

$$h(x) \le \xi_1(x) \le h(\beta), \quad x \in (\alpha,\beta). \tag{1.24}$$

Co więcej, jeżeli $\xi_1(x) = h(x)$ dla pewnego $x < \beta$, to

$$\xi_1(x) - h(x) = -\int_x^\beta [h(t) - h(x)]\,\mathrm{d}G(t) = 0.$$

Powyższa całka jest równa zero wtedy i tylko wtedy, gdy funkcja podcałkowa jest równa zero, czyli $h(t) = h(x)$ dla pewnego $t \in (x,\beta)$, ale to jest sprzeczne z założeniem ścisłej monotoniczności funkcji h. Zatem $h(x) < \xi_1(x)$, co dowodzi punktu (iii).

Aby wykazać punkt (iv) załóżmy, że x_1, x_2 są dwoma dowolnymi, ale różnymi punktami z przedziału (α,β). Bez straty ogólności możemy założyć, że $x_1 < x_2$. Wtedy z równości (1.23) dostajemy

$$\xi_1(x_1)\bar{F}(x_1)^\gamma - \xi_1(x_2)\bar{F}(x_2)^\gamma = \int_{x_2}^\beta h(t)\,\mathrm{d}\bar{F}(t)^\gamma - \int_{x_1}^\beta h(t)\,\mathrm{d}\bar{F}(t)^\gamma = -\int_{x_1}^{x_2} h(t)\,\mathrm{d}\bar{F}(t)^\gamma.$$

Stosując ponownie twierdzenie o wartości średniej dla całki Riemanna-Stieltjesa otrzymujemy

$$\xi_1(x_1)\bar{F}(x_1)^\gamma - \xi_1(x_2)\bar{F}(x_2)^\gamma = -h(\theta)\,[\bar{F}(x_2)^\gamma - \bar{F}(x_1)^\gamma]\,,$$

dla pewnego $\theta \in (x_1, x_2)$. Stąd

$$\bar{F}(x_1)^\gamma [\xi_1(x_1) - \xi_1(x_2)] = [\xi_1(x_2) - h(\theta)]\,[\bar{F}(x_2)^\gamma - \bar{F}(x_1)^\gamma]\,.$$

Z punktu (iii) otrzymujemy, że $\xi_1(x_2) - h(\theta) > 0$ dla każdego $\theta < x_2$. Zatem albo ξ jest stała, a więc F jest stała na przedziale (x_1, x_2), albo F jest ściśle rosnąca i ξ jest ściśle rosnąca, co kończy dowód punktu (iv).

Dowód punktu (v) wynika z nierówności (1.24) i twierdzenia o trzech ciągach. □

Wychodząc od Lematu 1.2 i korzystając z rekurencyjnej struktury regresji uogólnionych statystyk porządkowych (patrz Wniosek 1.1) wykażemy teraz analogiczne własności regresji (1.11) dla dowolnego $\ell \geq 1$.

Lemat 1.3. *Dla ustalonych $r, \ell \geq 1$, niech $X_*^{(r)}, X_*^{(r+\ell)}$ będą uogólnionymi statystykami porządkowymi z ciągłego rozkładu o dystrybuancie F określonej na nośniku (α, β). Załóżmy, że $h\colon (\alpha, \beta) \to \mathbb{R}$ jest funkcją ściśle rosnącą, ciągłą i taką, że $E\big|h\big(X_*^{(r+\ell)}\big)\big| < \infty$ oraz niech funkcja regresji $\xi : (\alpha, \beta) \to \mathbb{R}$ będzie określona wzorem* (1.11). *Wówczas ξ ma następujące własności:*

(i) ξ jest ciągła na (α, β);

(ii) jeżeli F jest różniczkowalna w sposób ciągły na (α, β), to ξ jest różniczkowalna na tym przedziale;

(iii) $\xi(x) > h(x)$ dla wszystkich $x \in (\alpha, \beta)$;

(iv) ξ jest rosnąca, a ponadto jest ona stała na przedziale $I \subset (\alpha, \beta)$ wtedy i tylko wtedy, gdy F jest stała na tym przedziale;

(v) jeżeli $h(\beta) < \infty$, to $\xi(\beta) = h(\beta)$.

Dowód. Oznaczmy $\varphi_\ell = h$ i rozważmy funkcje $\varphi_{\ell-1}, \ldots, \varphi_0$ zdefiniowane jak we Wniosku 1.1. Lemat 1.2 implikuje, że φ_i jest funkcją ciągłą na (α, β), o ile φ_{i+1} jest funkcją ciągłą, $\varphi_i > \varphi_{i+1}$ na (α, β), φ_i jest funkcją ściśle rosnącą pod warunkiem, że φ_{i+1} jest funkcją rosnącą, ponadto jeżeli $\varphi_{i+1}(\beta) < \infty$, to $\varphi_i(\beta) = \varphi_{i+1}(\beta)$. Z założenia φ_ℓ jest funkcją ściśle rosnącą i ciągłą, a zatem funkcje φ_i, $0 \leq i < \ell$, są ściśle rosnące i ciągłe, $\varphi_0 > h$, a ponadto $\varphi_0(\beta) = h(\beta)$, gdy $h(\beta) < \infty$. Ale z Wniosku 1.1 mamy $\varphi_0 = \xi$, co kończy dowód. □

Lemat 1.4. *Niech $r, \ell \geq 1$. Załóżmy, że $h\colon (\alpha, \beta) \to \mathbb{R}$ oraz $g\colon (\alpha, \beta) \to (g(\alpha), g(\beta))$ są funkcjami ciągłymi i ściśle rosnącymi. Wówczas równość*

$$\xi(x) = E\left(h\big(X_*^{(r+\ell)}\big) \;\big|\; X_*^{(r)} = x\right), \quad x \in (\alpha, \beta), \tag{1.25}$$

gdzie $X_*^{(r)}$, $X_*^{(r+\ell)}$ są uogólnionymi statystykami porządkowymi z rozkładu określonego przez dystrybuantę F zachodzi wtedy i tylko wtedy, gdy zachodzi

$$E\left(h\circ g^{-1}\big(Y_*^{(r+\ell)}\big)\ \big|\ Y_*^{(r)}=y\right)=\xi\circ g^{-1}(y),\quad y\in(g(\alpha),g(\beta)),$$

gdzie $Y_*^{(r)}$ oraz $Y_*^{(r+\ell)}$ są uogólnionymi statystykami porządkowymi z tymi samymi parametrami z rozkładu $F\circ g^{-1}$.

Dowód. Niech $X_*^{(j)}$, $1\le j\le r+\ell$, będą uogólnionymi statystykami porządkowymi z parametrami $\gamma_1,\dots,\gamma_{r+\ell}$ z rozkładu określonego przez dystrybuantę F. Położmy

$$Y_*^{(j)}=g\big(X_*^{(j)}\big),\quad 1\le j\le r+\ell.$$

Wówczas $Y_*^{(j)}$, $1\le j\le r+\ell$, są uogólnionymi statystykami porządkowymi z rozkładu o dystrybuancie $F\circ g^{-1}$ określonej na nośniku $(g(\alpha),g(\beta))$. Co więcej, dla $y\in(g(\alpha),g(\beta))$ mamy

$$\begin{aligned}E\left(h\circ g^{-1}\big(Y_*^{(r+\ell)}\big)\ \big|\ Y_*^{(r)}=y\right)&=E\left(h\circ g^{-1}\big(g(X_*^{(r+\ell)})\big)\ \big|\ g\big(X_*^{(r)}\big)=y\right)\\&=E\left(h\big(X_*^{(r+\ell)}\big)\ \big|\ X_*^{(r)}=g^{-1}(y)\right)\\&=\xi\circ g^{-1}(y),\end{aligned}$$

co kończy dowód twierdzenia. □

Kładąc $g=h$ widzimy, że warunek (1.25), gdzie $X_*^{(r)}$ i $X_*^{(r+\ell)}$ są uogólnionymi statystykami porządkowymi opartymi na rozkładzie F, jest równoważny

$$E\left(Y_*^{(r+\ell)}\ \big|\ Y_*^{(r)}=y\right)=\xi\circ h^{-1}(y),\quad y\in(h(\alpha),h(\beta)),$$

gdzie $Y_*^{(r)}$ i $Y_*^{(r+\ell)}$ są uogólnionymi statystykami porządkowymi (z tymi samymi parametrami) opartymi na rozkładzie $F\circ h^{-1}$. Tak więc, jeśli tylko jest to pomocne, możemy bez straty ogólności założyć, że $h(x)=x$. Fakt ten został wcześniej zaobserwowany w pracy [19].

1.4.2 Własności regresji dyskretnych wartości rekordowych

Przedmiotem rozważań w tym paragrafie będą analogiczne własności funkcji regresji wartości rekordowych dla rozkładów dyskretnych. Niech $\{X_n, n\ge 1\}$ będzie ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie dyskretnym określonym przez niezdegenerowaną dystrybuantę F o przeliczalnym nośniku S. W dalszej części pracy, bez zmniejszenia ogólności rozważań, będziemy zakładać, że nośnik jest zbiorem następującej postaci $S=\{0,1,\dots,N\}$, skończonym dla $N<\infty$ bądź nieskończonym dla $N=\infty$ (patrz [22], Sekcja 2). Ponadto dla $N<\infty$ przyjmujemy $\bar{S}=S\setminus\{N\}$, a dla $N=\infty$ przyjmujemy $\bar{S}=S$. Przypomnijmy, że $p_k=P(X_1=k)$ oraz $q_k=P(X_1\ge k)=\sum_{j=k}^{N}p_j$ dla $k\in S$.

Rozpatrzmy najpierw funkcję regresji kolejnych słabych wartości rekordowych określoną równaniem (1.13). Dla $\ell = 1$, korzystając z równości (1.5) otrzymujemy

$$\xi_1(j) = E\left(h(W_{r+1}) \mid W_r = j\right) = \frac{1}{q_j}\sum_{k=j}^{N} p_k h(k), \quad j \in S. \tag{1.26}$$

Z powyższej równości widzimy, że regresja słabych rekordów nie zależy od r. Zatem zachodzi m.in. równość $E\big(h(W_{r+1}) \mid W_r\big) = E\big(h(W_2) \mid W_1\big)$.

Korzystając teraz z formuły (1.26) pokażemy trzy istotne własności regresji ξ_1 kolejnych słabych wartości rekordowych. Poniższy lemat jest odpowiednikiem Lematu 1.2.

Lemat 1.5. *Dla ustalonego $r \geq 1$, niech W_r, W_{r+1} będą słabymi rekordami z rozkładu o dystrybuancie F o nośniku $S = \{0, 1, \ldots, N\}$, gdzie $N \leq \infty$. Załóżmy, że $h\colon S \to \mathbb{R}$ jest funkcją ściśle rosnącą i taką, że $E\big|h(W_{r+1})\big| < \infty$ oraz niech funkcja regresji $\xi_1\colon S \to \mathbb{R}$ będzie określona wzorem* (1.26)*. Wówczas ξ_1 ma następujące własności:*

(i) $\xi_1(j) > h(j)$ dla dowolnego $j \in \bar{S}$;

(ii) $\xi_1(j) < \xi_1(j+1)$ dla dowolnego $j \in \bar{S}$;

(iii) jeżeli $h(N) < \infty$, to $\xi_1(N) = h(N)$.

Uwaga 1.6. Jeżeli $N < \infty$, to oczywiście mamy $h(N) < \infty$. Jeżeli $N = \infty$, to $h(N)$ jest zdefiniowana jako granica $h(N) = \lim_{j\to\infty} h(j)$, która może być zarówno skończona, jak i nieskończona.

Dowód. (i) Ustalmy $j \in \bar{S}$. Ponieważ h jest funkcją ściśle rosnącą, to dla dowolnego $k > j$ mamy $h(k) > h(j)$. Z równania (1.26) otrzymujemy

$$\begin{aligned}
\xi_1(j) &= \frac{p_j}{q_j}h(j) + \frac{1}{q_j}\sum_{k=j+1}^{N} p_k h(k) \\
&> \frac{p_j}{q_j}h(j) + h(j)\frac{1}{q_j}\sum_{k=j+1}^{N} p_k = h(j).
\end{aligned}$$

(ii) Wychodząc od równości (1.26) możemy zapisać

$$\xi_1(j) q_j = \sum_{k=j}^{N} h(k) p_k, \quad j \in S.$$

Wyznaczając teraz różnice pierwszego rzędu powyższego równania otrzymujemy dla $j \in \bar{S}$ równanie różnicowe

$$\xi_1(j+1) q_{j+1} - \xi_1(j) q_j = -h(j) p_j,$$

a w konsekwencji

$$[\xi_1(j) - h(j)] p_j = [\xi_1(j+1) - \xi_1(j)] q_{j+1}, \quad j \in \bar{S}. \tag{1.27}$$

Na mocy punktu (i), lewa strona równania (1.27) jest dodatnia, zatem $\xi_1(j+1) > \xi_1(j)$ dla $j \in \bar{S}$.

(iii) Jeżeli $h(N) < \infty$, to dla ustalonego argumentu $j \in S$ mamy

$$h(j) \le h(k) \le h(N), \quad j \le k \le N.$$

Mnożąc powyższe wyrażenie przez p_k, a następnie sumując po wszystkich k (od $k = j$ do N) i dzieląc przez q_j otrzymujemy

$$h(j) \le \xi_1(j) \le h(N), \quad j \in S,$$

co implikuje $\xi_1(N) < \infty$ oraz $\xi_1(N) = h(N)$. $\square$

Następny lemat podaje analogiczne własności regresji (1.13) dla dowolnego $\ell \ge 1$. Jego dowód z wykorzystaniem Lematu 1.5 oraz rekurencyjnej struktury regresji słabych rekordów może być przeprowadzony podobnie jak dowód Lematu 1.3.

Lemat 1.6. *Dla ustalonych $r, \ell \ge 1$, niech $W_r, W_{r+\ell}$ będą słabymi rekordami z rozkładu o dystrybuancie F określonej na nośniku $S = \{0, 1, \dots, N\}$, gdzie $N \le \infty$. Załóżmy, że $h\colon S \to \mathbb{R}$ jest funkcją ściśle rosnącą i taką, że $E\big|h(W_{r+\ell})\big| < \infty$ oraz niech funkcja regresji $\xi : S \to \mathbb{R}$ będzie określona wzorem* (1.13). *Wówczas ξ ma następujące własności:*

(i) $\xi(j) > h(j)$ *dla dowolnego* $j \in \bar{S}$;

(ii) $\xi(j) < \xi(j+1)$ *dla dowolnego* $j \in \bar{S}$;

(iii) jeżeli $h(N) < \infty$, *to* $\xi(N) = h(N)$.

Na koniec tego paragrafu podamy analogiczne własności regresji dyskretnych wartości rekordów, które można wykazać podobnie jak własności regresji słabych wartości rekordowych.

Lemat 1.7. *Dla ustalonych $r, \ell \ge 1$, niech $R_r, R_{r+\ell}$ będą rekordami z rozkładu o dystrybuancie F określonej na nośniku $S = \{0, 1, \dots\}$. Załóżmy, że $h\colon S \to \mathbb{R}$ jest funkcją ściśle rosnącą i taką, że $E\big|h(R_{r+\ell})\big| < \infty$ oraz niech funkcja regresji $\xi : \{r-1, r, \dots\} \to \mathbb{R}$ będzie określona wzorem*

$$\xi(j) = E\left(h(R_{r+\ell}) \mid R_r = j\right), \quad j \ge r-1. \tag{1.28}$$

Wówczas ξ ma następujące własności:

(i) $\xi(j) > h(j+\ell)$ *dla dowolnego* $j \ge r-1$;

(ii) $\xi(j) < \xi(j+1)$ *dla dowolnego* $j \ge r-1$;

(iii) jeżeli $\lim_{j\to\infty} h(j) < \infty$, *to* $\lim_{j\to\infty} \xi(j) = \lim_{j\to\infty} h(j)$.

Rozdział 2

Charakteryzacje rozkładów absolutnie ciągłych przez regresje uogólnionych statystyk porządkowych

Przedmiotem rozważań w tym rozdziale będą charakteryzacje rozkładów absolutnie ciągłych o ciągłych gęstościach przez regresje uogólnionych statystyk porządkowych (1.11). Korzystając z rekurencyjnej struktury regresji uogólnionych statystyk porządkowych (zob. podrozdział 1.3) w podrozdziale 2.1 wykażemy, że jednoznaczność charakteryzacji rozkładu absolutnie ciągłego zachodzi wtedy i tylko wtedy, gdy odpowiedni problem różniczkowy (równanie różniczkowe lub układ równań różniczkowych) ma jednoznaczne rozwiązanie.

Mimo że kryterium jednoznaczności, które otrzymamy jest zaskakująco proste, okazuje się ono trudne w zastosowaniu nawet dla szczególnych przypadków funkcji h i ξ. Trudności związane z zastosowaniem tego kryterium wynikają miedzy innymi z nałożonych ograniczeń (warunków brzegowych) na rozwiązanie otrzymanego nieliniowego równania różniczkowego. Warunki brzegowe nie mają klasycznej postaci typu $y(x_0) = y_0$, jak w problemie Cauchy'ego, ale są wyrażone w szczególności w postaci nierówności $h(x) < y(x) < \xi(x)$ dla $x \in (\alpha, \beta)$. W podrozdziale 2.2 badamy problem charakteryzacyjny z wykorzystaniem nowego kryterium dla regresji z $\ell = 2$. Wykażemy jednoznaczność charakteryzacji w szczególnym przypadku, gdy $\xi(x) - h(x) \to 0$, gdy $x \to \beta^-$. W podrozdziale 2.3 zastosujemy otrzymane wyniki w celu wyznaczenia nowych charakteryzacji rozkładów absolutnie ciągłych o ciągłych gęstościach przez regresje uogólnionych statystyk porządkowych. Podamy charakteryzację dla pewnego rozkładu beta oraz wykażemy, że rozkład normalny, Gompertza oraz Weibulla z parametrem kształtu $\delta > 1$ są jednoznacznie charakteryzowane przez regresję niekolejnych uogólnionych statystyk porządkowych z $\ell = 2$. Korzystając z naszego podejścia wykażemy również, że rozkłady wykładniczy, potęgowy i Pareto są charakteryzowane przez warunek (1.11) przy założeniu liniowości regresji $\xi(x) = ax + b$. Będzie to nowy elementarny dowód wyniku otrzymanego w pracy [16].

W rozdziale tym zakładamy, że F jest absolutnie ciągłą dystrybuantą o nośniku (α, β), a funkcja gęstości $f = F'$ jest ciągła na (α, β). Ponadto niech $\bar{F}(x) = 1 - F(x)$ oznacza funkcję przeżycia rozkładu F oraz niech

$$H(x) = -\log \bar{F}(x), \quad \lambda(x) = \frac{f(x)}{\bar{F}(x)} = H'(x) \tag{2.1}$$

oznaczają odpowiednio funkcję hazardową oraz intensywność awarii rozkładu F. Oczywiście znajomość funkcji H lub λ wystarcza do jednoznacznego określenia funkcji rozkładu F.

2.1 Kryteria jednoznaczności charakteryzacji rozkładów absolutnie ciągłych

Rozważymy najpierw regresję kolejnych uogólnionych statystyk porządkowych, tj. regresję (1.11) z odstępem $\ell = 1$. Załóżmy zatem, że zachodzi relacja

$$\xi(x) = E\left(h\big(X_*^{(r+1)}\big) \mid X_*^{(r)} = x\right), \quad x \in (\alpha, \beta). \tag{2.2}$$

Dla absolutnie ciągłej dystrybuanty F, wychodząc ze wzoru (1.22), możemy zapisać

$$\xi(x)\bar{F}(x)^{\gamma_{r+1}} = \gamma_{r+1} \int_x^\beta h(t)\bar{F}(t)^{\gamma_{r+1}-1} f(t)\,\mathrm{d}t. \tag{2.3}$$

Ponieważ funkcje h, $\bar{F}$ i f są ciągłe, to prawa strona ostatniego równania jest różniczkowalna, a zatem funkcja ξ jest również różniczkowalna na (α, β). Różniczkując wyrażenie (2.3) stronami, a następnie odpowiednio porządkując otrzymujemy

$$\xi'(x) = \gamma_{r+1}\,[\xi(x) - h(x)]\,H'(x). \tag{2.4}$$

Z Lematu 1.2(iii) mamy $\xi > h$. Zatem z ostatniego równania wnioskujemy, że ciągłość funkcji gęstości f pociąga za sobą ciągłość pochodnej ξ' oraz funkcja ξ jest stała na przedziale $I \subset (\alpha, \beta)$ wtedy i tylko wtedy, gdy F jest stała na I (por. Lemat 1.2). Stąd, w niniejszym rozdziale, bez straty ogólności będziemy przyjmować, że wszystkie rozpatrywane funkcje ξ są ściśle rosnące, ciągłe i różniczkowalne w sposób ciągły na przedziale (α, β).

Dzieląc teraz obie strony równania (2.4) przez $\gamma_{r+1}[\xi(x) - h(x)]$, a następnie całkując względem x wyznaczamy funkcję H

$$H(x) = \frac{1}{\gamma_{r+1}} \int_\alpha^x \frac{\xi'(t)}{\xi(t) - h(t)}\,\mathrm{d}t, \quad x \in (\alpha, \beta). \tag{2.5}$$

Stąd dla dowolnego rozkładu F, dla którego zachodzi równanie regresji (2.2) mamy

$$\int_\alpha^x \frac{\xi'(t)}{\xi(t) - h(t)}\,\mathrm{d}t < \infty, \quad x \in (\alpha, \beta) \tag{2.6}$$

oraz

$$\int_\alpha^\beta \frac{\xi'(t)}{\xi(t)-h(t)}\,\mathrm{d}t=\infty, \tag{2.7}$$

gdyż dla każdego rozkładu F, jego funkcja hazardowa H dąży do ∞, gdy $x\to\beta^-$. Zauważmy ponadto, że

$$\lim_{x\to\beta^-}\xi(x)\exp\left(-\int_\alpha^x \frac{\xi'(t)}{\xi(t)-h(t)}\,\mathrm{d}t\right)=0. \tag{2.8}$$

Istotnie, korzystając z (2.5) otrzymujemy

$$\xi(x)\exp\left(-\int_\alpha^x \frac{\xi'(t)}{\xi(t)-h(t)}\,\mathrm{d}t\right)=\xi(x)\exp\left(-\gamma_{r+1}H(x)\right)=\xi(x)\bar{F}(x)^{\gamma_{r+1}},$$

co na mocy (2.3) musi dążyć do 0, gdy $x\to\beta^-$.

Przejdźmy teraz do przypadku regresji (1.11) z odstępem $\ell=2$. W dowodzie głównego twierdzenia wykorzystamy następujący lemat.

Lemat 2.1. *Dla ustalonej liczby naturalnej $r\geq 1$, niech $\gamma_1,\ldots,\gamma_r>0$ będą dowolnymi parametrami oraz $0<\gamma_{r+1}\leq\gamma_{r+2}$. Przypuśćmy, że*

(a) $h,\xi\colon(\alpha,\beta)\to\mathbb{R}$ są funkcjami ściśle rosnącymi, ciągłymi i takimi, że $h(x)<\xi(x)$ dla $x\in(\alpha,\beta)$, a ponadto funkcja ξ jest różniczkowalna w sposób ciągły,

(b) $y=\tau(x)$ jest dowolnym rozwiązaniem równania różniczkowego

$$y'=\frac{\gamma_{r+2}}{\gamma_{r+1}}\frac{y-h(x)}{\xi(x)-y}\xi'(x),\quad x\in(\alpha,\beta), \tag{2.9}$$

takim, że

$$h(x)<\tau(x)<\xi(x),\quad x\in(\alpha,\beta), \tag{2.10}$$

$$\int_\alpha^x \frac{\xi'(t)}{\xi(t)-\tau(t)}\,\mathrm{d}t<\infty,\quad x\in(\alpha,\beta), \tag{2.11}$$

$$\int_\alpha^\beta \frac{\xi'(t)}{\xi(t)-\tau(t)}\,\mathrm{d}t=\infty \tag{2.12}$$

oraz

$$\lim_{x\to\beta^-}\xi(x)\exp\left(-\int_\alpha^x \frac{\xi'(t)}{\xi(t)-\tau(t)}\,\mathrm{d}t\right)=0. \tag{2.13}$$

Wtedy wzór

$$G(x)=1-\exp\left(-\int_\alpha^x \frac{\eta'(t)}{\xi(t)-h(t)}\,\mathrm{d}t\right),\quad x\in(\alpha,\beta), \tag{2.14}$$

gdzie $\eta(t)=\frac{\xi(t)}{\gamma_{r+1}}+\frac{\tau(t)}{\gamma_{r+2}}$, określa absolutnie ciągłą dystrybuantę z ciągłą funkcją gęstości i taką, że dla uogólnionych statystyk porządkowych $Y_*^{(r)},Y_*^{(r+2)}$ z parametrami $\gamma_1,\ldots,\gamma_{r+2}$ z rozkładu G zachodzi równość

$$\xi(x)=E\left(h\big(Y_*^{(r+2)}\big)\;\middle|\;Y_*^{(r)}=x\right),\quad x\in(\alpha,\beta). \tag{2.15}$$

Uwaga 2.1. Jeżeli $\xi(\beta) = \lim_{x\to\beta^-} \xi(x) < \infty$, to przy spełnionym warunku (2.12) warunek (2.13) jest automatycznie spełniony.

Dowód Lematu 2.1. Niech τ będzie dowolnym rozwiązaniem równania różniczkowego (2.9), które spełnia warunki (2.10)–(2.13). Pokażemy najpierw, że funkcja G określona wzorem (2.14) jest dystrybuantą pewnego absolutnie ciągłego rozkładu o ciągłej funkcji gęstości.

Na mocy równania (2.9) można łatwo sprawdzić, że funkcja G określona wzorem (2.14) może być zapisana w następujących równoważnych postaciach

$$G(x) = 1 - \exp\left(-\frac{1}{\gamma_{r+1}}\int_\alpha^x \frac{\xi'(t)}{\xi(t)-\tau(t)}\,\mathrm{d}t\right) \tag{2.16}$$

bądź

$$G(x) = 1 - \exp\left(-\frac{1}{\gamma_{r+2}}\int_\alpha^x \frac{\tau'(t)}{\tau(t)-h(t)}\,\mathrm{d}t\right). \tag{2.17}$$

Wystarczy zatem pokazać, że wzór (2.16) bądź (2.17) określa dystrybuantę G o żądanych własnościach. Zauważmy najpierw, że skoro funkcje h, ξ są ściśle rosnące, to biorąc pod uwagę warunek (2.10) z równania różniczkowego (2.9) dla dodatnich parametrów $\gamma_{r+1}, \gamma_{r+2}$ wynika, że rozwiązanie τ tego równania jest również funkcją ściśle rosnącą na (α, β). Ponadto, ponieważ funkcja h jest ciągła a ξ różniczkowalna w sposób ciągły, to z równania (2.9) wynika również, że funkcja τ jest różniczkowalna w sposób ciągły na (α, β). Tak więc funkcja podcałkowa w równaniu (2.16) jest dodatnia i ciągła. Stąd funkcja G określona tym równaniem jest funkcją rosnącą, a na mocy podstawowego twierdzenia rachunku całkowego jest ona również różniczkowalna w sposób ciągły. Co więcej, gdy $x \to \alpha^+$, to $G(x) \to 0$, a gdy $x \to \beta^-$, to z warunku (2.12) otrzymujemy, że $G(x) \to 1$. Tak więc funkcja G jest dystrybuantą o żądanych własnościach.

Niech $Y_*^{(r)}$, $Y_*^{(r+1)}$ oraz $Y_*^{(r+2)}$ oznaczają uogólnione statystki porządkowe z rozkładu G. Pokażemy najpierw, że dla $x \in (\alpha, \beta)$ zachodzi równość

$$\tau(x) = E\left(h\big(Y_*^{(r+2)}\big) \mid Y_*^{(r+1)} = x\right). \tag{2.18}$$

Korzystając ze wzoru (2.17) otrzymujemy funkcję gęstości rozkładu G w postaci

$$g(x) = \frac{\bar{G}(x)}{\gamma_{r+2}}\frac{\tau'(x)}{\tau(x)-h(x)}.$$

Mnożąc stronami przez $\bar{G}(x)^{\gamma_{r+2}-1}$, po prostych przekształceniach sprowadzamy to równanie do postaci

$$\gamma_{r+2}h(x)\bar{G}(x)^{\gamma_{r+2}-1}g(x) = \gamma_{r+2}\tau(x)\bar{G}(x)^{\gamma_{r+2}-1}g(x) - \bar{G}(x)^{\gamma_{r+2}}\tau'(x).$$

Całkując stronami powyższą równość po dowolnie ustalonym przedziale $[x,y] \subset (\alpha, \beta)$, gdzie

$\alpha < x < y < \beta$, dostajemy

$$\begin{aligned}\gamma_{r+2}&\int_x^y h(t)\bar{G}(t)^{\gamma_{r+2}-1}g(t)\,\mathrm{d}t\\&=\gamma_{r+2}\int_x^y \tau(t)\bar{G}(t)^{\gamma_{r+2}-1}g(t)\,\mathrm{d}t-\int_x^y \bar{G}(t)^{\gamma_{r+2}}\tau'(t)\,\mathrm{d}t\\&=\tau(x)\bar{G}(x)^{\gamma_{r+2}}-\tau(y)\bar{G}(y)^{\gamma_{r+2}}.\end{aligned}\tag{2.19}$$

Ostatnią równość otrzymujemy stosując do ostatniej z całek wzór na całkowanie przez części.

Pokażemy teraz, że $\lim_{y\to\beta^-}\tau(y)\bar{G}(y)^{\gamma_{r+2}}=0$. Jeżeli $\lim_{y\to\beta^-}\tau(y)<\infty$, to przy spełnionym warunku (2.12) oczywiście $\lim_{y\to\beta^-}\tau(y)\bar{G}(y)^{\gamma_{r+2}}=0$. Jeżeli $\lim_{y\to\beta^-}\tau(y)=\infty$, to biorąc pod uwagę warunek (2.10) oraz założenie $0<\gamma_{r+1}\leq\gamma_{r+2}$, dla dostatecznie dużych y mamy

$$0\leq\tau(y)\bar{G}(y)^{\gamma_{r+2}}\leq\xi(y)\bar{G}(y)^{\gamma_{r+1}}=\xi(y)\exp\left(-\int_\alpha^y\frac{\xi'(t)}{\xi(t)-\tau(t)}\,\mathrm{d}t\right).$$

Korzystając zatem z warunku (2.13), otrzymujemy również w tym przypadku, że $\lim_{y\to\beta^-}\tau(y)\bar{G}(y)^{\gamma_{r+2}}=0$. Przechodząc w (2.19) do granicy, gdy y dąży do β otrzymujemy równość

$$\gamma_{r+2}\int_x^\beta h(t)\bar{G}(t)^{\gamma_{r+2}-1}g(t)\,\mathrm{d}t=\tau(x)\bar{G}(x)^{\gamma_{r+2}}.$$

Stąd i z (2.3) dostajemy, że funkcja τ spełnia równanie regresji (2.18).

Analogiczne obliczenia przy zastosowaniu (2.16) zamiast (2.17) oraz bezpośrednio warunku (2.13) prowadzą z kolei do równości

$$\xi(x)=E\left(\tau\big(Y_*^{(r+1)}\big)\;\middle|\;Y_*^{(r)}=x\right),\quad x\in(\alpha,\beta).\tag{2.20}$$

Łącząc teraz równości (2.18) i (2.20), na mocy Lematu 1.1, dostajemy równanie regresji (2.15) dla uogólnionych statystyk porządkowych $Y_*^{(r)}$, $Y_*^{(r+2)}$ z rozkładu G z odstępem 2. □

Uwaga 2.2. Zgodnie z Lematem 1.3(v) wiemy, że jeśli $h(\beta)<\infty$ oraz ξ jest zadana przez (2.15), to $\xi(\beta)<\infty$, a więc również $\tau(\beta)<\infty$ oraz $\xi(\beta)=\tau(\beta)=h(\beta)$. Jednakże, w założeniach Lematu 2.1 warunek ten nie występuje. Stąd powstaje pytanie, czy warunek ten jest również potrzebny. Pokażemy, że z założeń Lematu 2.1 wynika, że

$$\text{jeśli } h(\beta)<\infty, \text{ to } \xi(\beta)=\tau(\beta)=h(\beta).\tag{2.21}$$

Zauważmy, że jeśli $h(\beta)<\infty$, to na mocy (2.19) otrzymujemy dla $\alpha<x<y<\beta$

$$\begin{aligned}\tau(x)\bar{G}(x)^{\gamma_{r+2}}-\tau(y)\bar{G}(y)^{\gamma_{r+2}}&\leq h(\beta)\gamma_{r+2}\int_x^y\bar{G}(t)^{\gamma_{r+2}-1}g(t)\,\mathrm{d}t\\&=h(\beta)[\bar{G}(x)^{\gamma_{r+2}}-\bar{G}(y)^{\gamma_{r+2}}],\end{aligned}$$

gdzie oczywiście G jest dana wzorem (2.16). Skoro $\lim_{y\to\beta^-}\tau(y)\bar{G}(y)^{\gamma_{r+2}}=0$, to z powyższej nierówności wynika, że dla $\alpha<x<\beta$

$$\tau(x)\bar{G}(x)^{\gamma_{r+2}}\leq h(\beta)\bar{G}(x)^{\gamma_{r+2}}$$

(bo $\bar{G}(y)^{\gamma_{r+2}} \to 0$, gdy $y \to \beta^-$). Z definicji α i β mamy $\bar{G}(x) \in (0,1)$ dla $x \in (\alpha,\beta)$, a więc dzieląc stronami otrzymamy

$$\tau(x) \leq h(\beta), \quad \text{dla } x \in (\alpha,\beta).$$

Zatem $\tau(\beta) < \infty$, a więc $\tau(\beta) \leq h(\beta)$. Z drugiej strony, na mocy własności $\tau > h$ na (α,β), otrzymujemy $\tau(\beta) \geq h(\beta)$. Zatem, jeśli $h(\beta) < \infty$, to $\tau(\beta) = h(\beta)$.

Analogicznie, z dowodu równości (2.20) wynika, że jeśli $\tau(\beta) < \infty$, to dla $\alpha < x < y < \beta$ mamy

$$\xi(x)\bar{G}(x)^{\gamma_{r+1}} - \xi(y)\bar{G}(y)^{\gamma_{r+1}} \leq \tau(\beta)[\bar{G}(x)^{\gamma_{r+1}} - \bar{G}(y)^{\gamma_{r+1}}].$$

Zatem, jeśli $\tau(\beta) < \infty$, to $\xi(\beta) = \tau(\beta)$. To dowodzi prawdziwości implikacji (2.21).

Przejdźmy teraz do sformułowania i dowodu pierwszego głównego wyniku tego rozdziału. Lemat 2.1 pokazuje, że każde rozwiązanie τ opisane w punkcie (b) tego lematu generuje pewien rozkład prawdopodobieństwa G (zależny od τ), dla którego odpowiednia regresja o odstępie 2 jest równa danej funkcji ξ, niezależnie od wyboru τ. Fakt ten wykorzystamy w dowodzie poniższego twierdzenia.

Twierdzenie 2.1. *Dla ustalonej liczby naturalnej $r \geq 1$, niech $X_*^{(r)}, X_*^{(r+2)}$ będą uogólnionymi statystykami porządkowymi z absolutnie ciągłego rozkładu $F\colon (\alpha,\beta) \to (0,1)$ o ciągłej funkcji gęstości. Załóżmy, że $h\colon (\alpha,\beta) \to \mathbb{R}$ jest funkcją ściśle rosnącą, ciągłą i taką, że $E\left|h\big(X_*^{(r+2)}\big)\right| < \infty$. Wówczas różniczkowalna w sposób ciągły funkcja regresji $\xi\colon (\alpha,\beta) \to \mathbb{R}$ określona wzorem*

$$\xi(x) = E\left(h\big(X_*^{(r+2)}\big) \mid X_*^{(r)} = x\right) \tag{2.22}$$

jednoznacznie charakteryzuje rozkład F mający ciągłą funkcję gęstości wtedy i tylko wtedy, gdy równanie różniczkowe (2.9) *ma dokładnie jedno rozwiązanie $y = \varphi(x)$ spełniające warunki* (2.10)–(2.13) *z τ zastąpionym przez φ. Ponadto funkcja rozkładu F określona jest wówczas wzorem*

$$F(x) = 1 - \exp\left(-\int_\alpha^x \frac{\eta'(t)}{\xi(t) - h(t)}\,\mathrm{d}t\right), \tag{2.23}$$

gdzie $\eta(t) = \frac{\xi(t)}{\gamma_{r+1}} + \frac{\varphi(t)}{\gamma_{r+2}}$.

Dowód. W dowodzie twierdzenia, bez zmniejszenia ogólności, przyjmujemy, że $\gamma_{r+1} \leq \gamma_{r+2}$, gdyż wartości funkcji (2.22) nie zależą od kolejności parametrów $\gamma_{r+1}, \gamma_{r+2}$ (zob. Uwaga 1.2).

Załóżmy, że dana jest regresja (2.22) i zdefiniujmy

$$\varphi(x) = E\left(h\big(X_*^{(r+2)}\big) \mid X_*^{(r+1)} = x\right). \tag{2.24}$$

Wtedy na mocy Wniosku 1.1 dla $\ell = 2$ (patrz (1.20) z $\varphi_2 = h$) otrzymujemy

$$\xi(x) = E\left(\varphi\big(X_*^{(r+1)}\big) \mid X_*^{(r)} = x\right). \tag{2.25}$$

Ponieważ funkcja φ określona wzorem (2.24) jest funkcją regresji kolejnych uogólnionych statystyk porządkowych, to spełnia ona warunki (2.11)–(2.13) będące odpowiednikami warunków (2.6)–(2.8). Co więcej, stosując dwukrotnie Lemat 1.2(iii) dostajemy $\xi > \varphi > h$ na przedziale (α, β), a więc spełniony jest warunek (2.10). Ponieważ funkcje φ i ξ są dwiema funkcjami regresji kolejnych uogólnionych statystyk porządkowych z rozkładu określonego przez dystrybuantę F z ciągłą funkcją gęstości, to obie są różniczkowalne w sposób ciągły. Stosując tożsamość (2.4) do (2.24) oraz (2.25) otrzymujemy dwa równania

$$\begin{cases} \varphi'(x) = \gamma_{r+2}[\varphi(x) - h(x)]H'(x), \\ \xi'(x) = \gamma_{r+1}[\xi(x) - \varphi(x)]H'(x), \end{cases} \qquad x \in (\alpha, \beta). \tag{2.26}$$

Zdefiniujmy pomocniczą funkcję $\eta : (\alpha, \beta) \to \mathbb{R}$ następująco

$$\eta(x) = \frac{\xi(x)}{\gamma_{r+1}} + \frac{\varphi(x)}{\gamma_{r+2}}. \tag{2.27}$$

Dzieląc pierwsze równanie w (2.26) przez γ_{r+2}, a drugie przez γ_{r+1}, a następnie sumując stronami równania dostajemy $\eta'(x) = [\xi(x) - h(x)]H'(x)$. Teraz podobnie jak dla przypadku $\ell = 1$, dzieląc obie strony przez $\xi - h$, a następnie całkując stronami względem x wyznaczamy funkcję H

$$H(x) = \int_{\alpha}^{x} \frac{\eta'(t)}{\xi(t) - h(t)}\,\mathrm{d}t, \quad x \in (\alpha, \beta).$$

Powyższa równość implikuje wzór (2.23).

Potraktujmy teraz (2.26) jako układ dwóch równań z dwiema nieznanymi funkcjami φ i H. Eliminując H z układu równań otrzymujemy

$$\frac{\varphi'(x)}{\gamma_{r+2}[\varphi(x) - h(x)]} = H'(x) = \frac{\xi'(x)}{\gamma_{r+1}[\xi(x) - \varphi(x)]},$$

a stąd wynika, że funkcja φ określona przez (2.24) spełnia równanie różniczkowe (2.9). Ponadto, z powyższych rozważań wynika, że funkcja φ spełnia również warunki (2.10)–(2.13). Zatem, jeżeli ξ jest funkcją regresji daną wzorem (2.22), to φ jest rozwiązaniem równania różniczkowego (2.9) spełniającym warunki (2.10)–(2.13).

Z drugiej strony, na mocy Lematu 2.1, każde rozwiązanie problemu różniczkowego (2.9)–(2.13) definiuje przy użyciu wzoru (2.23) pewien ciągły rozkład prawdopodobieństwa F z ciągłą funkcją gęstości, dla którego regresja uogólnionych statystyk porządkowych jest określona wzorem (2.22). Tak więc liczba rozwiązań problemu różniczkowego pokrywa się z liczbą absolutnie ciągłych rozkładów prawdopodobieństwa charakteryzowanych przez regresję ξ określoną wzorem (2.22). Jeżeli zatem problem (2.9) z warunkami (2.10)–(2.13) ma jednoznaczne rozwiązanie τ, to τ pokrywa się z φ danym przez (2.24) i F jest jednoznacznie określona przez (2.23). W przeciwnym razie istnieją co najmniej dwa różne rozkłady, dla których zachodzi (2.22). To kończy dowód. $\square$

Twierdzenie 2.1 daje się uogólnić na przypadek regresji z dowolnym odstępem $\ell \geq 2$.

Twierdzenie 2.2. *Dla ustalonych liczb naturalnych $r \geq 1$, $\ell \geq 2$, niech $X_*^{(r)}, X_*^{(r+\ell)}$ będą uogólnionymi statystykami porządkowymi z absolutnie ciągłego rozkładu $F\colon (\alpha,\beta) \to (0,1)$ o ciągłej funkcji gęstości. Załóżmy, że $h\colon (\alpha,\beta) \to \mathbb{R}$ jest funkcją ściśle rosnącą, ciągłą i taką, że $E\left|h\left(X_*^{(r+\ell)}\right)\right| < \infty$. Wówczas różniczkowalna w sposób ciągły funkcja regresji $\xi\colon (\alpha,\beta) \to \mathbb{R}$ określona wzorem* (1.11) *jednoznacznie charakteryzuje rozkład F mający ciągłą funkcję gęstości wtedy i tylko wtedy, gdy problem (układ $\ell - 1$ równań różniczkowych)*

$$y_i' = \frac{\gamma_{r+i+1}}{\gamma_{r+1}} \frac{y_i - y_{i+1}}{\xi(x) - y_1} \xi'(x), \quad 1 \leq i \leq \ell - 1, \tag{2.28}$$

gdzie $y_\ell = h$ ma dokładnie jedno rozwiązanie $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_{\ell-1})$ spełniające warunki

$$h(x) < \varphi_{\ell-1}(x) < \ldots < \varphi_1(x) < \xi(x), \quad x \in (\alpha,\beta), \tag{2.29}$$

$$\int_\alpha^x \frac{\xi'(t)}{\xi(t) - \varphi_1(t)} \,\mathrm{d}t < \infty, \quad x \in (\alpha,\beta), \tag{2.30}$$

$$\int_\alpha^\beta \frac{\xi'(t)}{\xi(t) - \varphi_1(t)} \,\mathrm{d}t = \infty, \tag{2.31}$$

oraz

$$\lim_{x \to \beta^-} \xi(x) \exp\left(-\int_\alpha^x \frac{\xi'(t)}{\xi(t) - \varphi_1(t)} \,\mathrm{d}t \right) = 0. \tag{2.32}$$

Ponadto funkcja rozkładu F określona jest wówczas wzorem

$$F(x) = 1 - \exp\left(-\int_\alpha^x \frac{\eta'(t)}{\xi(t) - h(t)} \,\mathrm{d}t \right), \tag{2.33}$$

gdzie

$$\eta(x) = \frac{\xi(x)}{\gamma_{r+1}} + \sum_{i=1}^{\ell-1} \frac{\varphi_i(x)}{\gamma_{r+i+1}}. \tag{2.34}$$

W dowodzie twierdzenia pomocny jest następujący lemat będący uogólnieniem Lematu 2.1

Lemat 2.2. *Dla ustalonych liczb naturalnych $r \geq 1$, $\ell \geq 2$, niech $\gamma_1, \ldots, \gamma_r > 0$ będą dowolnymi parametrami oraz $0 < \gamma_{r+1} \leq \gamma_{r+i}$, $2 \leq i \leq \ell$. Przypuśćmy, że*

(a) $h, \xi\colon (\alpha,\beta) \to \mathbb{R}$ są funkcjami ściśle rosnącymi, ciągłymi i takimi, że $h(x) < \xi(x)$ dla $x \in (\alpha,\beta)$, a ponadto funkcja ξ jest różniczkowalna w sposób ciągły,

(b) $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_{\ell-1})$ jest dowolnym rozwiązaniem układu równań różniczkowych (2.28) *spełniającym warunki* (2.29), (2.30), (2.31) *oraz* (2.32) *z $\boldsymbol{\varphi}$ zastąpionym przez $\boldsymbol{\tau}$.*

Wtedy wzór

$$G(x) = 1 - \exp\left(-\int_\alpha^x \frac{\eta'(t)}{\xi(t) - h(t)} \,\mathrm{d}t \right), \quad x \in (\alpha,\beta),$$

z funkcją η określoną wzorem (2.34) *(z φ zastąpionym przez τ), określa absolutnie ciągłą dystrybuantę z ciągłą funkcją gęstości taką, że dla uogólnionych statystyk porządkowych $Y_*^{(r)}, Y_*^{(r+\ell)}$ z parametrami $\gamma_1, \ldots, \gamma_{r+\ell}$ z rozkładu G zachodzi równość*

$$\xi(x) = E\left(h\left(Y_*^{(r+\ell)}\right) \mid Y_*^{(r)} = x \right), \quad x \in (\alpha,\beta).$$

Dowód lematu przeprowadza się analogicznie jak dowód Lematu 2.1, który jest jego szczególnym przypadkiem dla $\ell = 2$. W tym celu wystarczy dowodzić kolejno równości

$$\tau_i(x) = E\left(\tau_{i+1}\big(Y_*^{(r+i+1)}\big) \mid Y_*^{(r+i)} = x\right),$$

gdzie $\tau_\ell = h$, $\tau_0 = \xi$, począwszy od $i = \ell - 1$ do $i = 1$. Równoważnie

$$\tau_i(x)\bar{G}(x)^{\gamma_{r+i+1}} = \gamma_{r+i+1}\int_x^\beta \tau_{i+1}(t)\bar{G}(t)^{\gamma_{r+i+1}-1}g(t)\,\mathrm{d}t,$$

dla $x \in (\alpha, \beta)$. W dowodach tych równości korzystamy z faktu, że

$$\lim_{x\to\beta^-} \tau_i(x)\exp\left(-\int_\alpha^x \frac{\tau_i'(t)}{\tau_i(t)-\tau_{i+1}(t)}\,\mathrm{d}t\right) = 0,$$

co wynika z założenia (2.32) oraz nierówności $0 < \gamma_{r+1} \le \gamma_{r+i}$, dla $2 \le i \le \ell$. Istotnie, dla $i = 0, 1, \ldots, \ell - 1$, mamy na mocy równań (2.28)

$$\bar{G}(x) = \exp\left(-\frac{1}{\gamma_{r+i+1}}\int_\alpha^x \frac{\tau_i'(t)}{\tau_i(t)-\tau_{i+1}(t)}\,\mathrm{d}t\right).$$

Zatem dla dostatecznie dużych x

$$\begin{aligned} 0 \le \tau_i(x)\exp\left(-\int_\alpha^x \frac{\tau_i'(t)}{\tau_i(t)-\tau_{i+1}(t)}\,\mathrm{d}t\right) &= \tau_i(x)\bar{G}(x)^{\gamma_{r+i+1}} \\ \le \xi(x)\bar{G}(x)^{\gamma_{r+1}} = \xi(x)\exp\left(-\int_\alpha^x \frac{\xi'(t)}{\xi(t)-\tau_1(t)}\,\mathrm{d}t\right)&, \end{aligned}$$

co dąży do 0 na mocy (2.32).

Dowód Twierdzenia 2.2. Załóżmy, że zachodzi równość regresji (1.11) i oznaczmy $\varphi_\ell = h$. Następnie rozważmy funkcje φ_i, $0 \le i \le \ell - 1$, określone równością (1.19). Wówczas z Wniosku 1.1 otrzymujemy $\varphi_0 = \xi$. Co więcej, ponieważ każda funkcja φ_i jest funkcją regresji kolejnych uogólnionych statystyk porządkowych, ściśle mówiąc regresją $\varphi_{i+1}\big(X_*^{(r+i+1)}\big)$ względem $X_*^{(r+i)}$, to każda funkcja φ_i jest różniczkowalna w sposób ciągły. Każda z tych funkcji spełnia warunki analogiczne do (2.6)–(2.8), w szczególności spełnione są warunki (2.30)–(2.32). Ponadto na mocy Lematu 1.2(iii) mamy $\varphi_{i+1} < \varphi_i$ na przedziale (α, β) dla $0 \le i \le \ell - 1$. Zatem zachodzi nierówność (2.29). Dalej funkcje te spełniają ℓ równań różniczkowych (patrz równanie (2.4) dla $\ell = 1$)

$$\varphi_i'(x) = \gamma_{r+i+1}\left[\varphi_i(x) - \varphi_{i+1}(x)\right]H'(x), \quad 0 \le i \le \ell - 1. \tag{2.35}$$

Podobnie do (2.27) określmy teraz funkcję $\eta\colon (\alpha, \beta) \to \mathbb{R}$ wzorem (2.34). Dzieląc i-te równanie w (2.35) przez γ_{r+i+1}, a następnie sumując wszystkie równania otrzymujemy $\eta'(x) = [\xi(x) - h(x)]H'(x)$, a więc ponownie funkcję rozkładu F można wyrazić w postaci (2.33).

Zauważmy, że (2.35) jest układem ℓ równań z ℓ nieznanymi funkcjami $\varphi_1,\dots,\varphi_{\ell-1}$ oraz H (przypomnijmy, że $\varphi_0=\xi$ oraz $\varphi_\ell=h$ są znane). Aby wyeliminować H wyznaczmy tę funkcję z każdego równania otrzymując

$$H'(x)=\frac{\varphi_i'(x)}{\gamma_{r+i+1}[\varphi_i(x)-\varphi_{i+1}(x)]},\quad 1\le i\le \ell-1, \tag{2.36}$$

oraz

$$H'(x)=\frac{\xi'(x)}{\gamma_{r+1}[\xi(x)-\varphi_1(x)]}.$$

Porównując teraz prawą stronę ostatniego równania z prawą stroną równań (2.36) widzimy, że funkcje $\varphi_1,\dots,\varphi_{\ell-1}$ spełniają układ $\ell-1$ równań

$$\begin{cases} \varphi_i'(x)=\dfrac{\gamma_{r+i+1}}{\gamma_{r+1}}\dfrac{\varphi_i(x)-\varphi_{i+1}(x)}{\xi(x)-\varphi_1(x)}\xi'(x), & \text{dla } 1\le i\le \ell-2,\\ \varphi_{\ell-1}'(x)=\dfrac{\gamma_{r+\ell}}{\gamma_{r+1}}\dfrac{\varphi_{\ell-1}(x)-h(x)}{\xi(x)-\varphi_1(x)}\xi'(x). \end{cases}$$

Z powyższych rozważań funkcje te spełniają również warunki (2.29)–(2.32). W drugą stronę dowód wynika z Lematu 2.2, tak samo jak w dowodzie Twierdzenia 2.1. □

Uwaga 2.3. Oczywiście dla $\ell=2$ układ (2.28) redukuje się do jednego równania różniczkowego (2.9). Dlatego też Twierdzenie 2.1 jest szczególnym przypadkiem Twierdzenia 2.2.

2.2 Jednoznaczność charakteryzacji – szczególne przypadki dla $\ell=2$

W poprzednim podrozdziale sformułowaliśmy kryterium jednoznaczności charakteryzacji absolutnie ciągłych rozkładów prawdopodobieństwa w terminach równań różniczkowych. Mimo że otrzymane kryterium jest zaskakująco proste, okazuje się trudne w zastosowaniu nawet dla szczególnych funkcji h i ξ. Największa trudność wynika z nałożonych ograniczeń na rozwiązania nieliniowych równań różniczkowych. Warunki ograniczające nie mają klasycznej postaci typu $y(x_0)=y_0$, jak w problemie Cauchy'ego, którego jednoznaczność rozwiązania jest dobrze znana, ale na przykład dla $\ell=2$ mamy warunek ograniczający w postaci nierówności $h(x)<y(x)<\xi(x)$ dla $x\in(\alpha,\beta)$.

Do końca tego rozdziału zajmujemy się problemem charakteryzacji z wykorzystaniem nowego kryterium dla regresji z odstępem $\ell=2$. Wykażemy jednoznaczność charakteryzacji w następujących szczególnych przypadkach:

(i) h jest funkcją ograniczoną z góry, a więc $h(\beta)=\lim_{x\to\beta^-}h(x)<\infty$;

(ii) $h(x)=x$ oraz $\xi(x)-h(x)\to 0$, gdy $x\to\infty$;

(iii) $h(x) = x$ oraz $\xi(x) = ax + b$ dla pewnego $a > 0$.

W dowodach przedstawionych w niniejszym podrozdziale będziemy stosować następujące rozumowanie: funkcja φ określona wzorem (2.24) spełnia wszystkie warunki (2.9), (2.10), (2.11), (2.12) i (2.13), a ponadto w każdym z rozważanych przypadków szczególnych nie istnieją żadne inne rozwiązania równania różniczkowego (2.9) spełniające warunek (2.10). Zatem φ jest jedyną funkcją mającą żądane własności oraz na mocy Twierdzenia 2.1 rozkład F jest wówczas wyznaczony jednoznacznie wzorem (2.23).

2.2.1 Własności rozwiązań pomocniczego problemu różniczkowego

W niniejszym podrozdziale zbadamy własności rozwiązań następującego problemu różniczkowego

$$\begin{cases} y' = \gamma \dfrac{y - h(x)}{\xi(x) - y} \xi'(x), \\ h(x) < y(x) < \xi(x), \end{cases} \qquad x \in (\alpha, \beta), \tag{2.37}$$

gdzie γ jest dowolną liczbą dodatnią.

Przy założeniu, że funkcje ξ i h są ściśle rosnące, z warunku $h < y < \xi$ wnioskujemy, że $y'(x) > 0$ dla każdego $x \in (\alpha, \beta)$. Zatem każde rozwiązanie problemu (2.37) jest również funkcją ściśle rosnącą. Poniższy lemat podaje inne istotne w dalszej części pracy własności rozwiązań tego problemu.

Lemat 2.3. *Niech $h, \xi\colon (\alpha, \beta) \to \mathbb{R}$ będą funkcjami ściśle rosnącymi i takimi, że h jest ciągła, ξ jest różniczkowalna oraz $h(x) < \xi(x)$ dla $x \in (\alpha, \beta)$. Załóżmy, że $\gamma > 0$ oraz $\varphi, y\colon (\alpha, \beta) \to \mathbb{R}$ są dwoma różnymi rozwiązaniami problemu różniczkowego* (2.37)*. Wtedy*

(i) $y(x) \neq \varphi(x)$ dla każdego $x \in (\alpha, \beta)$;

(ii) albo $y < \varphi$, albo $y > \varphi$ na przedziale (α, β);

(iii) jeżeli $y > \varphi$, to $y' > \varphi'$ oraz jeżeli $y < \varphi$, to $y' < \varphi'$;

(iv) jeżeli $z = (y - \varphi)^2$, to z jest funkcją ściśle rosnącą na (α, β).

Dowód. Niech $D = \{(x, y)\colon \alpha < x < \beta, h(x) < y < \xi(x)\}$. Dla dowolnie wybranego $x_0 \in (\alpha, \beta)$ oznaczmy $\varphi_0 = \varphi(x_0)$. Wybierzmy $\varepsilon > 0$ tak małe, aby kula $B = B\big((x_0, \varphi_0), \varepsilon\big) \subset D$. Rozważmy funkcję $K\colon D \to (0, \infty)$ zdefiniowaną następująco

$$K(x, y) = \gamma \frac{y - h(x)}{\xi(x) - y} \xi'(x).$$

Wówczas równanie różniczkowe (2.37) można zapisać jako $y' = K(x, y)$. Zauważmy również, że pochodna

$$\frac{\partial K}{\partial y} = \gamma \frac{\xi(x) - h(x)}{(\xi(x) - y)^2} \xi'(x)$$

jest ograniczona w B, a więc K spełnia warunek Lipschitza ze względu na y w zbiorze B. Na mocy twierdzenia Picarda zagadnienie $y' = K(x,y)$, $y(x_0) = \varphi_0$ ma jednoznaczne rozwiązanie w przedziale $(x_0 - \delta, x_0 + \delta)$ dla pewnego $\delta > 0$ wystarczająco małego. Oczywiście rozwiązanie to musi pokrywać się z funkcją φ na tym przedziale. Zatem jeżeli y jest innym rozwiązaniem równania (2.37), to $y(x_0) \neq \varphi(x_0)$, co kończy dowód punktu (i).

Aby wykazać punkt (ii) wystarczy zauważyć, że φ i y są różniczkowalne, a zatem ciągłe. Skoro ich różnica zawsze jest różna od zera, to albo jest dodatnia, albo ujemna na całym przedziale (α, β).

Aby wykazać punkt (iii) obliczymy pochodną różnicy $y - \varphi$

$$y' - \varphi' = \gamma \frac{(y-\varphi)(\xi - h)}{(\xi - y)(\xi - \varphi)} \xi'.$$

Ponieważ $h < \varphi < \xi$ i $h < y < \xi$, to oczywiście $y' > \varphi'$, gdy $y > \varphi$ oraz $y' < \varphi'$, gdy $y < \varphi$.

(iv) Wyznaczając pochodną funkcji $z = (y - \varphi)^2$, a następnie korzystając z punktu (iii) lub nierówności $h < \varphi < \xi$ i $h < y < \xi$ otrzymujemy

$$z' = 2(y-\varphi)(y' - \varphi') = \gamma \frac{(y-\varphi)^2(\xi - h)}{(\xi - y)(\xi - \varphi)} \xi' > 0.$$

Zatem z jest funkcją ściśle rosnąca. $\square$

Na koniec tego paragrafu rozważymy możliwość podania dokładnego jawnego rozwiązania problemu różniczkowego (2.37). Rozpatrzmy równanie Abela drugiego rodzaju, które ma następującą ogólną postać (por. [48], Rozdział 0.1.6)

$$[y + g(x)]\, y' = f_2(x) y^2 + f_1(x) y + f_0(x). \tag{2.38}$$

Przepisując (2.37) w następującej postaci

$$(y - \xi) y' = -\gamma \xi' y + \gamma h \xi' \tag{2.39}$$

widzimy, że jest to równanie Abela drugiego rodzaju z $g = -\xi$, $f_2 = 0$, $f_1 = -\gamma\xi'$ oraz $f_0 = \gamma h \xi'$. Rozwiązania wielu szczególnych przypadków równania (2.38) można znaleźć w Rozdziale 1.3 w wyżej wspomnianej monografii [48]. Podstawiając teraz $w = y - \xi$ możemy przekształcić równanie (2.39) do postaci

$$w w' = F_1(x) w + F_0(x), \tag{2.40}$$

gdzie $F_1(x) = -(\gamma + 1)\xi'$ i $F_0(x) = -\gamma(\xi - h)\xi'$. Również różne przykłady równań w postaci (2.40) są rozważane w [48] (patrz Rozdział 1.3.3). Kolejna zamiana zmiennych

$$z = \int F_1(x)\, \mathrm{d}x = -(\gamma + 1)\xi(x),$$

lub równoważnie $x = \xi^{-1}\left(-\frac{z}{\gamma+1}\right)$, prowadzi do równania różniczkowego

$$w w'_z - w = G(z), \tag{2.41}$$

gdzie w'_z oznacza pochodną funkcji w względem z oraz

$$G(z) = -\frac{\gamma}{(\gamma+1)^2}z - \frac{\gamma}{\gamma+1}h \circ \xi^{-1}\Big(-\frac{z}{\gamma+1}\Big).$$

Rozwiązania szczególnych równań tej postaci są rozważane w Rozdziale 1.3.1 w monografii [48]. Jednakże wszystkie przedstawione rozwiązania równań (2.38), (2.39), jak i (2.41) nie są podane w postaci jawnej, nawet dla szczególnych przypadków. Stąd też trudno zweryfikować, czy spełniają warunek początkowy (2.10) lub zastosować je do wyznaczenia funkcji rozkładu ze wzoru (2.23). Dlatego wydaje się, że wyniki przedstawione w [48] nie są pomocne w naszych rozważaniach.

2.2.2 Jednoznaczność charakteryzacji dla $h(\beta) < \infty$

Korzystając z Lematu 2.3 wykażemy jednoznaczność charakteryzacji absolutnie ciągłych rozkładów prawdopodobieństwa przez regresje uogólnionych statystyk porządkowych w przypadku, gdy $\ell = 2$ oraz funkcja h jest ograniczona z góry, a więc $h(\beta) < \infty$. Przypomnijmy, że na mocy Lematu 1.3(v), gdy $h(\beta) < \infty$, to $\xi(\beta) = h(\beta) < \infty$.

Twierdzenie 2.3. *Dla ustalonej liczby naturalnej $r \geq 1$, niech $X_*^{(r)}, X_*^{(r+2)}$ będą uogólnionymi statystykami porządkowymi z absolutnie ciągłego rozkładu F o ciągłej funkcji gęstości. Załóżmy, że $h\colon (\alpha,\beta) \to \mathbb{R}$ jest funkcją ściśle rosnącą, ciągłą i taką, że $h(\beta) < \infty$ oraz $E\big|h\big(X_*^{(r+2)}\big)\big| < \infty$. Jeżeli znamy regresję*

$$\xi(x) = E\Big(h\big(X_*^{(r+2)}\big) \mid X_*^{(r)} = x\Big), \quad x \in (\alpha,\beta),$$

która ma ciągłą pochodną, to wzór (2.23) *wyznacza jednoznacznie rozkład F z ciągłą gęstością f.*

Uwaga 2.4. Nie zakładamy, że h jest funkcją ograniczoną z dołu i dlatego istnienie $E\big|h\big(X_*^{(r+2)}\big)\big|$ musimy założyć dodatkowo. W Przykładach 3.1 i 3.2 pokażemy, że możliwe jest, że $h(\alpha) = -\infty$.

Dowód. W niniejszym dowodzie stosujemy Lemat 2.3 z $\gamma = \frac{\gamma_{r+2}}{\gamma_{r+1}}$. Wiemy już, że φ określona przez (2.24) jest rozwiązaniem równania różniczkowego (2.9) spełniającym warunki (2.10)–(2.13). Naturalnie φ jest także rozwiązaniem problemu (2.37) z $\gamma = \frac{\gamma_{r+2}}{\gamma_{r+1}}$. Przypuśćmy, że istnieje również inne rozwiązanie y tego problemu. Wówczas dla $z = (y-\varphi)^2$ mamy $0 \leq z(x) \leq [\xi(x) - h(x)]^2$, czyli $\lim_{x\to\beta^-} z(x) = 0$. Z drugiej strony $z(\alpha) > 0$ oraz na podstawie Lematu 2.3(iv), funkcja z jest ściśle rosnąca. To daje sprzeczność. Zatem φ jest jedynym rozwiązaniem równania różniczkowego (2.9) spełniającym żądane warunki. Teza twierdzenia wynika teraz z Twierdzenia 2.1. □

Uwaga 2.5. Przypuszczamy, że prawdziwe jest uogólnienie Twierdzenia 2.3 na przypadek dowolnego $\ell \geq 3$, ale nie znaleźliśmy na to formalnego dowodu. Mianowicie dla dowolnego $\ell \geq 3$, jeżeli $h(\beta) < \infty$, to regresja (1.11) określa jednoznacznie funkcję rozkładu F wzorem (2.33). To przypuszczenie jest poparte faktem, iż w literaturze nie są znane przykłady niejednoznaczności dla takiego przypadku. Aby udowodnić to uogólnione twierdzenie wystarczyłoby uogólnić Lemat 2.3(i) oraz (iv), tj. wykazać, że jeżeli $\mathbf{y} = (y_1, \dots, y_{\ell-1})$ oraz $\boldsymbol{\varphi} = (\varphi_1, \dots, \varphi_{\ell-1})$ są dwoma rozwiązaniami uogólnionego problemu (2.28), to te rozwiązania nie mają punktów wspólnych oraz odległość (euklidesowa) tych rozwiązań jest funkcją rosnącą. Pierwsze zadanie nie jest trudne, ale drugie wydaje się być problematyczne. Przypuśćmy, że $\mathbf{y} = (y_1, \dots, y_{\ell-1})$ oraz $\boldsymbol{\varphi} = (\varphi_1, \dots, \varphi_{\ell-1})$ są dwoma rozwiązaniami problemu (2.28) i określmy

$$z = \|\mathbf{y} - \boldsymbol{\varphi}\|^2 = \sum_{i=1}^{\ell-1} (y_i - \varphi_i)^2 .$$

Zatem pochodna funkcji z jest następująca

$$z' = 2\xi' \sum_{i=1}^{\ell-1} \frac{\gamma_{r+i+1}}{\gamma_{r+1}} (y_i - \varphi_i) \left(\frac{y_i - y_{i+1}}{\xi - y_1} - \frac{\varphi_i - \varphi_{i+1}}{\xi - \varphi_1} \right).$$

Jeżeli $\ell = 2$, to $y_2 = \varphi_2 = h$ oraz powyższy problem sprowadza się do Lematu 2.3(iv), w którym wykazaliśmy, że $z' > 0$. Jednakże dla $\ell \geq 3$ nie znajdujemy argumentów, które świadczyłyby, że z' jest większe od zera, co oznaczałoby, że odległość między $\mathbf{y}$ i $\boldsymbol{\varphi}$ rośnie. Zatem dla $\ell \geq 3$ problem pozostaje wciąż otwarty.

2.2.3 Jednoznaczność charakteryzacji dla $h(\beta) = \infty$

W niniejszym paragrafie rozważamy przypadek, gdy funkcja h jest nieograniczona. Przyjmujemy dla uproszczenia (bez straty ogólności rozważań), że $h(x) = x$, a zatem dla funkcji nieograniczonej mamy $\beta = \infty$ oraz $h(\beta) = \infty$. W takim przypadku nie możemy zastosować Twierdzenia 2.3 w celu wykazania jednoznaczności charakteryzacji rozkładu prawdopodobieństwa przez funkcję regresji ξ daną wzorem (2.22). Jednakże analizując dowód Twierdzenia 2.3 widzimy, że jeśli

$$\lim_{x\to\infty} [\xi(x) - h(x)] = 0, \tag{2.42}$$

to również zachodzi jednoznaczność charakteryzacji. W praktyce trudno jest obliczyć funkcję regresji ξ, a więc nie jest łatwo sprawdzić czy warunek (2.42) jest spełniony. Wykażemy jednakże, że dla dużej klasy rozkładów można ten warunek zweryfikować jedynie na podstawie znajomości funkcji intensywności awarii tych rozkładów, bez konieczności obliczania regresji ξ. Przypomnijmy, że funkcja intensywności awarii λ jest określona wzorem (2.1).

Lemat 2.4. *Jeżeli* $h(x) = x$, $\beta = \infty$ *oraz* $\lim_{x\to\infty} \lambda(x) = \infty$, *to spełniony jest warunek* (2.42).

Dowód. Wychodząc od równości (2.3) warunki (2.24) oraz (2.25) możemy odpowiednio zapisać w postaci

$$\varphi(x) = \frac{\gamma_{r+2}}{\bar{F}(x)^{\gamma_{r+2}}} \int_x^\infty h(t)\bar{F}(t)^{\gamma_{r+2}-1} f(t)\,\mathrm{d}t \tag{2.43}$$

oraz

$$\xi(x) = \frac{\gamma_{r+1}}{\bar{F}(x)^{\gamma_{r+1}}} \int_x^\infty \varphi(t)\bar{F}(t)^{\gamma_{r+1}-1} f(t)\,\mathrm{d}t. \tag{2.44}$$

Całkując przez części prawą stronę równania (2.43) otrzymujemy

$$\varphi(x) = h(x) + \frac{1}{\bar{F}(x)^{\gamma_{r+2}}} \int_x^\infty \bar{F}(t)^{\gamma_{r+2}}\,\mathrm{d}t.$$

Następnie, korzystając z reguły de l'Hospitala obliczamy granicę

$$\lim_{x\to\infty} [\varphi(x) - h(x)] = \frac{1}{\gamma_{r+2}} \lim_{x\to\infty} \frac{1}{\lambda(x)} = 0.$$

Podobnie z (2.44) otrzymujemy

$$\begin{aligned}\xi(x) = h(x) &+ \frac{1}{\bar{F}(x)^{\gamma_{r+1}}} \int_x^\infty \bar{F}(t)^{\gamma_{r+1}}\,\mathrm{d}t \\ &+ \frac{\gamma_{r+1}}{\bar{F}(x)^{\gamma_{r+1}}} \int_x^\infty \left(\int_t^\infty \bar{F}(u)^{\gamma_{r+2}}\,\mathrm{d}u \right) \bar{F}(t)^{\gamma_{r+1}-\gamma_{r+2}-1} f(t)\,\mathrm{d}t\end{aligned}$$

i ponownie stosując regułę de l'Hospitala dostajemy

$$\lim_{x\to\infty} [\xi(x) - h(x)] = \left(\frac{1}{\gamma_{r+1}} + \frac{1}{\gamma_{r+2}} \right) \lim_{x\to\infty} \frac{1}{\lambda(x)} = 0.$$

To kończy dowód. □

Wniosek 2.1. Jeżeli $\lim_{x\to\infty} \lambda(x) = \infty$, to ciągły rozkład F z ciągłą funkcją gęstości jest określony jednoznacznie przez regresję $\xi(x) = E\big(X_*^{(r+2)} \mid X_*^{(r)} = x\big)$.

2.3 Przykłady i zastosowania

2.3.1 Regresje o odstępie dwa

W tym paragrafie wykorzystamy wyniki otrzymane w poprzednich podrozdziałach do otrzymania nowych charakteryzacji rozkładów absolutnie ciągłych. W poniższych przykładach, dla uproszczenia obliczeń, rozważamy funkcje regresji wartości rekordowe R_r, $r \geq 1$, ciągu niezależnych zmiennych losowych o jednakowym rozkładzie określonym przez funkcję gęstości f. Wiemy, że rekordy są uogólnionymi statystykami porządkowymi z parametrami $\gamma_j = 1$, $j \geq 1$ (zob. podrozdział 1.1.3). Korzystając z równości (2.3) w połączeniu z równaniami (2.24) i (2.25) regresję $\xi(x) = E\big(R_{r+2} \mid R_r = x\big)$ możemy przedstawić w postaci

$$\xi(x) = \frac{1}{\bar{F}(x)} \int_x^\beta \varphi(t) f(t)\,\mathrm{d}t, \tag{2.45}$$

gdzie funkcja $\varphi(x) = E\big(R_{r+2} \mid R_{r+1} = x\big)$ jest zdefiniowana następująco

$$\varphi(x) = \frac{1}{\bar{F}(x)} \int_x^{\beta} t f(t)\, \mathrm{d}t. \tag{2.46}$$

W pierwszym przykładzie pokażemy nową charakteryzację pewnego rozkładu beta. Rozkład Beta(a,b) określa funkcja gęstości postaci

$$f(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}, \quad 0 < x < 1, \quad a,b > 0,$$

gdzie B oznacza funkcję beta Eulera.

Przykład 2.1. Dla rozkładu Beta$(2,2)$ z gęstością $f(x) = 6x(1-x)$, $0 < x < 1$, otrzymujemy

$$\varphi(x) = E\big(R_{r+2} \mid R_{r+1} = x\big) = \frac{1+2x+3x^2}{2(1+2x)}.$$

W konsekwencji z (2.45) po żmudnych obliczeniach wykazujemy, że

$$\xi(x) = E\big(R_{r+2} \mid R_n = x\big) = \frac{2(1-x)(26-x+2x^2-18x^3)+27\log\left(\frac{1+2x}{3}\right)}{32(1-x)^2(1+2x)}. \tag{2.47}$$

Oczywiście dla badanego rozkładu mamy $\beta = 1$ i $h(x) = x$, a więc $h(\beta) < \infty$. Zatem, na mocy Twierdzenia 2.3 rozkład Beta$(2,2)$ jest jedynym rozkładem, dla którego zachodzi warunek regresji wartości rekordowych (2.47).

Do końca tego paragrafu będziemy rozważać problem jednoznacznej charakteryzacji popularnych rozkładów prawdopodobieństwa przez odpowiadające im regresje uogólnionych statystyk porządkowych, mimo że nie znamy analitycznej postaci tych funkcji. Istnienie odpowiednich funkcji regresji dla rozważanych poniżej rozkładów prawdopodobieństwa wynika z istnienia momentów tych rozkładów oraz twierdzenia o istnieniu momentów uogólnionych statystyk porządkowych z tych rozkładów (zob. [20], Twierdzenie 2.2).

Kolejny przykład dotyczy charakteryzacji rozkładu normalnego.

Przykład 2.2. Rozważmy rozkład normalny $N(\mu,\sigma^2)$ z parametrami $\mu \in \mathbb{R}$, $\sigma > 0$, określony przez następującą funkcję gęstości

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}, \quad \mu \in \mathbb{R}, \sigma > 0.$$

Funkcja $\bar{F}$ wyraża się następująco

$$\bar{F}(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_x^{\infty} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) \mathrm{d}t.$$

Korzystając z reguły de l'Hospitala łatwo dostajemy $\lambda(x) \to \infty$, gdy $x \to \infty$. Teraz biorąc pod uwagę Wniosek 2.1 stwierdzamy, że rozkład normalny jest jednoznacznie charakteryzowany przez odpowiadającą mu regresję uogólnionych statystyk porządkowych, mimo że nie znamy analitycznej postaci funkcji ξ.

Uwaga 2.6. Można łatwo sprawdzić, że warunek $\lim_{x\to\infty} \lambda(x) = \infty$ zachodzi również m.in. dla rozkładu Weibulla z parametrem kształtu $\delta > 1$ oraz rozkładu Gompertza. Istotnie:

- rozkład Weibulla z parametrami $\delta, \theta > 0$, $\alpha \in \mathbb{R}$, dla którego

  $$f(x) = \delta\theta(x-\alpha)^{\delta-1} \exp\left[-\theta(x-\alpha)^{\delta}\right], \quad \bar{F}(x) = \exp\left[-\theta(x-\alpha)^{\delta}\right], \quad x > \alpha, \tag{2.48}$$

  ma potęgową funkcję intensywności awarii

  $$\lambda(x) = \delta\theta(x-\alpha)^{\delta-1}.$$

  Dla $\delta > 1$ mamy $\lambda(x) \to \infty$, gdy $x \to \infty$, a więc jest to sytuacja analogiczna jak w Przykładzie 2.2. Natomiast dla $0 < \delta < 1$ mamy $\lambda(x) \to 0$ oraz dla $\delta = 1$ (rozkład wykładniczy) mamy $\lambda(x) = \theta$.

- rozkład Gompertza z parametrami $\delta, \theta > 0$, dla którego

  $$f(x) = \delta \exp(\theta x) \exp\left[\frac{\delta}{\theta}\left[1-\exp(\theta x)\right]\right], \quad \bar{F}(x) = \exp\left[\frac{\delta}{\theta}\left[1-\exp(\theta x)\right]\right], \quad x > 0,$$

  ma wykładniczą funkcję intensywności awarii

  $$\lambda(x) = \delta \exp(\theta x).$$

  Oczywiście $\lambda(x) \to \infty$, gdy $x \to \infty$.

Zatem na mocy Wniosku 2.1 rozkłady te również są jednoznacznie charakteryzowane przez odpowiadające im funkcje regresji uogólnionych statystyk porządkowych o ostępie dwa ($\ell = 2$). Jednakże nie udało nam się znaleźć analitycznych wzorów tych regresji.

Z drugiej strony warunek z Wniosku 2.1 nie zachodzi na przykład dla rozkładu gamma, Weibulla z $\delta \leq 1$ (w szczególności dla rozkładu wykładniczego, $\delta = 1$), Pareto oraz rozkładu logarytmiczno-normalnego. Zatem na pytanie, czy te rozkłady są jednoznacznie charakteryzowane przez regresje uogólnionych statystyk porządkowych nie możemy odpowiedzieć korzystając z wyników zaprezentowanych w podrozdziale 2.2. Do tego problemu wrócimy w kolejnym rozdziale.

2.3.2 Regresja liniowa

W tym paragrafie rozpatrzymy klasyczny problem charakteryzacji rozkładów prawdopodobieństwa przez liniowe regresje uogólnionych statystyk porządkowych z perspektywy nowego kryterium jednoznaczności charakteryzacji otrzymanego w podrozdziale 2.1.

Załóżmy najpierw, że $E\left(X_*^{(r+1)} \mid X_*^{(r)} = x\right) = ax + b$ dla $x \in (\alpha, \beta)$. Zatem rozważamy przypadek, gdy $h(x) = x$ i $\xi(x) = ax + b$, gdzie $a > 0$, gdyż z Lematu 1.2 funkcja ξ jest rosnącą. Co więcej, korzystając z (2.5) i postępując analogicznie jak np. w sekcji 4 w pracy [11], otrzymujemy, że możliwe są tylko następujące przypadki:

(i) jeżeli $0<a<1$, to $\beta=\frac{b}{1-a}$ oraz bazowy rozkład jest rozkładem potęgowym określonym na nośniku (α,β), to znaczy $\bar{F}(x)=\left(\frac{\beta-x}{\beta-\alpha}\right)^{\theta}$ dla $x\in(\alpha,\beta)$, gdzie $\theta=\frac{a}{(1-a)\gamma_{r+1}}>0$;

(ii) jeżeli $a=1$, to $\beta=\infty$, $b>0$ oraz bazowy rozkład jest rozkładem wykładniczym o nośniku (α,∞), to znaczy $\bar{F}(x)=\exp[-\theta(x-\alpha)]$ dla $x\geq\alpha$, gdzie $\theta=\frac{1}{b\gamma_{r+1}}$;

(iii) jeżeli $a>1$, to $\alpha>\frac{b}{1-a}$, $\beta=\infty$ oraz bazowy rozkład jest rozkładem Pareto o nośniku (α,∞), to znaczy $\bar{F}(x)=\left(\frac{\delta+\alpha}{\delta+x}\right)^{\theta}$ dla $x\geq\alpha$, gdzie $\theta=\frac{a}{(a-1)\gamma_{r+1}}>0$, $\delta=\frac{b}{a-1}>0$.

Powyższe trzy rodziny rozkładów prawdopodobieństwa otrzymujemy również w przypadku dowolnego $\ell\geq 2$. Jeżeli F jest ciągłą funkcją rozkładu na (α,β) oraz zachodzi relacja $E\big(X_*^{(r+\ell)}\mid X_*^{(r)}\big)=aX_*^{(r)}+b$ dla pewnego $a>0$, to wówczas zachodzą tylko następujące przypadki:

(i) $0<a<1$ i F jest funkcją rozkładu potęgowego;

(ii) $a=1$ i F jest funkcją rozkładu wykładniczego;

(iii) $a>1$ i F jest funkcją rozkładu Pareto.

W każdym przypadku parametry charakteryzujące powyższe rozkłady są jednoznacznie wyznaczone przez α, β, a, b oraz parametry uogólnionych statystyk porządkowych $\gamma_{r+1},\ldots,\gamma_{r+\ell}$ (zob. [16], Twierdzenie 1 oraz [19], Twierdzenie 5.1). Znany dowód tego wyniku został oparty na pomyśle zastosowanym dla regresji statystyk porządkowych i wartości rekordowych w [25, 26] i bazuje na rozwiązaniu scałkowanego równania funkcyjnego Cauchy'ego [49]. Poniżej korzystając z Twierdzenia 2.1 podamy inny niemal elementarny dowód tego wyniku w przypadku, gdy $\ell=2$.

Załóżmy, że dla pewnego rozkładu F zachodzi równość

$$E\left(X_*^{(r+2)}\mid X_*^{(r)}=x\right)=ax+b,\quad x\in(\alpha,\beta),\tag{2.49}$$

dla pewnych stałych $a,b\in\mathbb{R}$ oraz $-\infty\leq\alpha<\beta\leq\infty$. Ponadto, dla uproszczenia zapisu przyjmijmy, że $\gamma=\frac{\gamma_{r+2}}{\gamma_{r+1}}$. Pokażemy najpierw, że $\alpha>-\infty$. Oczywiście $\xi(x)=ax+b$ jest funkcją różniczkowalną, a więc na mocy Twierdzenia 2.1 może ona charakteryzować jedynie rozkłady absolutnie ciągłe z gęstością ciągłą na (α,β). Dalej zgodnie z Twierdzeniem 2.1 szukamy rozwiązań równania

$$\varphi'(x)=a\gamma\frac{\varphi(x)-x}{ax+b-\varphi(x)},\quad x\in(\alpha,\beta)\tag{2.50}$$

takich, że w szczególności

$$x<\varphi(x)<ax+b,\quad x\in(\alpha,\beta)\tag{2.51}$$

oraz

$$\int_{\alpha}^{x}\frac{a}{at+b-\varphi(t)}\,\mathrm{d}t<\infty,\quad x\in(\alpha,\beta).\tag{2.52}$$

Jeśli $\xi(x) = ax + b$ ma być regresją, to musi być ściśle rosnąca, a więc $a > 0$. Jeśli $a > 1$, to warunek $\xi(x) > x$ zachodzi jedynie dla $x > \frac{b}{1-a}$, a więc w tym przypadku $\alpha \geq \frac{b}{1-a} > -\infty$. Jeśli $a \in (0,1]$ oraz φ spełnia (2.50) i (2.51), to

$$\int_\alpha^x \frac{\mathrm{d}t}{at + b - \varphi(t)} \geq \int_\alpha^x \frac{\mathrm{d}t}{(a-1)t + b}, \quad x \in (\alpha, \beta).$$

Jeśli $\alpha = -\infty$, to ostatnia całka jest rozbieżna do ∞ dla każdego $x \in (\alpha, \beta)$. Zatem gdyby $\alpha = -\infty$, to nie istniałoby żadne rozwiązanie φ równania (2.50) spełniające jednocześnie warunki (2.51) oraz (2.52). Zatem $\alpha > -\infty$.

Załóżmy najpierw, że $\alpha = 0$, a więc $h(x) = x$, $\xi(x) = ax + b$ dla pewnych $a, b > 0$ oraz $x \in (0, \beta)$. Z własności funkcji regresji (Lemat 1.3) otrzymujemy

$$\beta = \inf\{x > 0 : ax + b = x\} = \begin{cases} \frac{b}{1-a}, & \text{gdy} \quad 0 < a < 1, \\ \infty, & \text{gdy} \quad a \geq 1. \end{cases}$$

Nasz pierwszy problem to znaleźć rozwiązanie y takie, że

$$\begin{cases} y' = a\gamma \dfrac{y - x}{ax + b - y}, \\ x < y(x) < ax + b, \end{cases} \qquad x \in (0, \beta). \tag{2.53}$$

Sprawdźmy najpierw, czy istnieją liniowe rozwiązania tego problemu postaci

$$y(x) = cx + d. \tag{2.54}$$

Oczywiście z uwagi na to, że $0 = h(0) < y(0) < \xi(0)$, mamy $d \in (0, b)$. Ponadto jeżeli $0 < a < 1$, to $c < 1$, gdyż $x < cx + d$ oraz $c > a$, gdyż $cx + d < ax + b$ dla $x > 0$. Podobnie jeżeli $a > 1$, to $c > 1$ oraz jeżeli $a = 1$, to $c = 1$. Podsumowując, możliwe są następujące przypadki

$$d \in (0, b) \quad \text{oraz} \quad \begin{cases} c \in (a, 1), & \text{gdy } 0 < a < 1, \\ c = 1, & \text{gdy } a = 1, \\ c \in (1, a), & \text{gdy } a > 1. \end{cases}$$

Wstawiając $a = 1$ do (2.53) otrzymujemy $y(x) = x + d$ oraz $1 = \frac{\gamma d}{b-d}$, a ponieważ $\gamma > 0$, to $d = \frac{b}{1+\gamma} \in (0, b)$. Jeżeli natomiast $a \neq 1$, to

$$c = a\gamma \frac{cx + d - x}{ax + b - (cx + d)}$$

lub równoważnie $(a-c)cx + (b-d)c = a\gamma(c-1)x + ad\gamma$. To z kolei prowadzi do układu równań

$$\begin{cases} (a - c)c = a\gamma(c - 1), \\ (b - d)c = ad\gamma. \end{cases} \tag{2.55}$$

Pierwsze równanie jest równaniem kwadratowym względem parametru c

$$c^2 + a(\gamma - 1)c - a\gamma = 0. \tag{2.56}$$

Zauważmy, że $\Delta = a^2(\gamma-1)^2 + 4a\gamma > 0$, zatem to równanie ma dwa pierwiastki różnych znaków. Zdefiniujmy $f(c) = c^2 + a(\gamma-1)c - a\gamma$. Wtedy jeżeli $a > 1$, to $f(1) = 1 - a < 0$ oraz $f(a) = a\gamma(a-1) > 0$. Zatem równanie kwadratowe (2.56) ma dokładnie jedno rozwiązanie c w przedziale $(1,a)$. Podobnie w przypadku $0 < a < 1$, równanie (2.56) ma dokładnie jedno rozwiązanie c w przedziale $(a,1)$. Co więcej, dla $a \neq 1$ z równań (2.55) otrzymujemy

$$\frac{b-d}{d} = \frac{a\gamma}{c} = \frac{a-c}{c-1},$$

a zatem

$$d = \frac{c-1}{a-1} b \in (0,b). \tag{2.57}$$

Podsumowując, jeżeli $a > 0$ oraz $b > 0$, to problem różniczkowy (2.53) ma dokładnie jedno rozwiązanie liniowe $\varphi(x) = cx + d$ takie, że: $c = 1$ i $d = \frac{b}{1+\gamma}$, gdy $a = 1$, w przeciwnym razie c jest jedynym dodatnim rozwiązaniem (2.56), a d jest określone przez (2.57).

Zauważmy, że rozwiązanie liniowe naszego problemu różniczkowego spełnia również pozostałe warunki z Twierdzenia 2.1, czyli (2.11), (2.12) i (2.13). Mianowicie, jeśli $a = 1$, to $\beta = \infty$ oraz

$$\int_0^x \frac{\xi'(t)}{\xi(t) - \varphi(t)} \,\mathrm{d}t = \frac{1+\gamma}{\gamma} \int_0^x \frac{1}{b} \,\mathrm{d}t = \frac{1+\gamma}{\gamma} \frac{x}{b}.$$

Zatem spełnione są warunki (2.11) oraz (2.12). Ponadto

$$\lim_{x\to\infty} \xi(x) \exp\left(-\int_0^x \frac{\xi'(t)}{\xi(t) - \varphi(t)} \,\mathrm{d}t \right) = \lim_{x\to\infty} (ax+b) \exp\left(-\frac{1+\gamma}{\gamma} \frac{x}{b} \right) = 0,$$

a więc zachodzi warunek (2.13).

Jeśli $a \neq 1$, to $b - d = \frac{(a-c)b}{a-1}$, a zatem

$$\begin{aligned}
\int_0^x \frac{\xi'(t)}{\xi(t) - \varphi(t)} \,\mathrm{d}t &= a \int_0^x \frac{\mathrm{d}t}{(a-c)t + (b-d)} \\
&= \frac{a(a-1)}{a-c} \int_0^x \frac{\mathrm{d}t}{(a-1)t + b} \\
&= \frac{a}{a-c} \log \frac{(a-1)x + b}{b}.
\end{aligned}$$

Jeśli $0 < a < 1$, to $\beta = \frac{b}{1-a}$ oraz $a < c < 1$, a więc oczywiście zachodzą warunki (2.11) oraz (2.12). Ponadto $\lim_{x\to\beta^-} \xi(x) = \frac{b}{1-a}$, czyli (2.13) oczywiście zachodzi. Jeśli $a > 1$, to $\beta = \infty$ oraz $a > c > 1$, a więc oczywiście również zachodzą warunki (2.11) oraz (2.12). Ponadto

$$\xi(x) \exp\left(-\int_0^x \frac{\xi'(t)}{\xi(t) - \varphi(t)} \,\mathrm{d}t \right) = (ax+b) \left(\frac{b}{(a-1)x + b} \right)^{\frac{a}{a-c}} \sim C x^{-\frac{c}{a-c}} \to 0,$$

gdy $x \to \infty$. Zatem i w tym przypadku zachodzi warunek (2.13).

W następnym kroku musimy wykazać, że problem (2.53) nie ma innych rozwiązań (rozwiązań nieliniowych).

Jeżeli $0 < a < 1$, to $\beta = \frac{b}{1-a} < \infty$. W tym przypadku jednoznaczność rozwiązania problemu (2.53) przez funkcję liniową (2.54) wynika z Twierdzenia 2.3. To z kolei implikuje, że

$$E\left(X_*^{(r+2)} \mid X_*^{(r+1)} = x\right) = cx + d, \quad x \in (0, \beta), \tag{2.58}$$

z $c \in (0,1)$, a więc F jest jednoznacznie określoną funkcją rozkładu potęgowego na $(0, \frac{b}{1-a})$.

Jeżeli $a \geq 1$, to mamy $\beta = \infty$ i $h(\beta) = \infty$, tak więc nie możemy zastosować Twierdzenia 2.3. Stąd naszym kolejnym zadaniem jest wykazać jednoznaczność rozwiązania problemu (2.53) przez funkcję liniową (2.54) ze współczynnikiem $c \geq 1$. W tym celu wyznaczmy druga pochodną y, aby wykazać, że y' jest albo funkcją rosnącą, albo funkcją malejącą na przedziale $(0, \infty)$.

Załóżmy najpierw, że $a = 1$. Wtedy mamy szczególne rozwiązanie $\varphi(x) = x + d$ oraz jeżeli y jest innym rozwiązaniem problemu

$$y' = \gamma \frac{y - x}{x + b - y}, \quad x < y < x + b,$$

to y' jest funkcją różniczkowalną i wyznaczając pochodną dostajemy

$$y'' = \frac{\gamma b}{(x + b - y)^2}(y' - 1).$$

Dalej korzystając z Lematu 2.3 mamy do rozważenia następujące przypadki:

- jeżeli $y > x + d$, to $y' > 1$ i $y'' > 0$, zatem y' jest funkcją rosnącą;
- jeżeli $y < x + d$, to $y' < 1$ i $y'' < 0$, zatem y' jest funkcją malejącą.

W obu przypadkach istnieje granica $\lim_{x\to\infty} y'(x) = y'(\infty)$ (być może granica jest niewłaściwa). Z warunku $x < y < x + b$, dostajemy $1 < \frac{y}{x} < 1 + \frac{b}{x}$, a stąd

$$\lim_{x\to\infty} \frac{y(x)}{x} = 1.$$

Z reguły de l'Hospitala mamy również

$$y'(\infty) = \lim_{x\to\infty} y'(x) = \lim_{x\to\infty} \frac{y(x)}{x} = 1.$$

Z drugiej strony, dla każdego rozwiązania y naszego problemu zachodzi tożsamość

$$y'(0) = \frac{\gamma y(0)}{b - y(0)},$$

gdzie $y(0) \in (0, b)$. Zatem $y'(0)$ rośnie od 0 do ∞, gdy $y(0)$ zmienia się od 0 do b. W szczególności $y'(0) = 1$ wtedy i tylko wtedy, gdy $y(0) = \frac{b}{1+\gamma} = d$. Zatem, jeżeli y jest rozwiązaniem takim, że $y(0) > d$, to $y(x) > x + d$, $y'(0) > 1$ oraz y' jest funkcją rosnącą. Wtedy dostajemy sprzeczność

$$1 < y'(0) < y'(\infty) = 1.$$

Podobnie, jeżeli y jest rozwiązaniem takim, że $y(0) < d$, to dostajemy kolejną sprzeczność $1 > y'(0) > y'(\infty) = 1$. Tak więc, w przypadku gdy $a = 1$, nie ma innych rozwiązań naszego problemu różniczkowego niż $\varphi(x) = x + d$, gdzie $d = \frac{b}{1+\gamma}$. To implikuje, że relacja (2.58) zachodzi z $c = 1$, a więc F jest jednoznacznie określoną funkcją rozkładu wykładniczego na $(0, \infty)$.

Załóżmy teraz, że $a > 1$. Wówczas z (2.53) y' jest różniczkowalna oraz

$$y - x = \frac{y'}{a\gamma}(ax + b - y). \tag{2.59}$$

Tak więc z (2.59) druga pochodna y jest następująca

$$\begin{aligned} y'' &= \frac{a\gamma}{(ax+b-y)^2}\left[(y'-1)(ax+b-y) - (a-y')\frac{y'}{a\gamma}(ax+b-y)\right] \\ &= \frac{1}{ax+b-y}\left[(y')^2 + a(\gamma-1)y' - a\gamma\right] = \frac{f(y')}{ax+b-y}, \end{aligned} \tag{2.60}$$

gdzie

$$f(x) = x^2 + a(\gamma - 1)x - a\gamma.$$

Wiemy już, że to równanie ma jeden pierwiastek $c \in (1, a)$, a drugi pierwiastek tego równania jest ujemny. Stąd

$$f(x) \begin{cases} < 0, & \text{dla } x \in (0, c), \\ > 0, & \text{dla } x > c. \end{cases}$$

Mając na uwadze Lemat 2.3 ponownie z (2.60) otrzymujemy

- jeżeli $y > \varphi$, to $y' > c$ i $f(y'(x)) > f(c) = 0$ dla $x > 0$, zatem y' jest funkcją rosnącą;
- jeżeli $y < \varphi$, to $y' < c$ i $f(y'(x)) < f(c) = 0$ dla $x > 0$, zatem y' jest funkcją malejącą.

W obu przypadkach granica $y'(\infty) = \lim_{x\to\infty} y'(x)$ istnieje (być może niewłaściwa). Z reguły de l'Hospitala granica

$$L = \lim_{x\to\infty} \frac{y(x)}{x} = \lim_{x\to\infty} y'(x)$$

również istnieje. Z warunku $x < y(x) < ax + b$ mamy $1 < \frac{y}{x} < a + \frac{b}{x}$, zatem L jest granicą skończoną taką, że $1 \leq L \leq a$. Przepisując teraz (2.53) w postaci

$$y' = a\gamma \frac{\frac{y}{x} - 1}{a + \frac{b}{x} - \frac{y}{x}},$$

a następnie przechodząc do granicy, gdy $x \to \infty$ dostajemy $L = a\gamma\frac{L-1}{a-L}$. To pokazuje, że $L \in (1, a)$, a porównanie z pierwszym równaniem w (2.55) prowadzi do wniosku, że $L = c$.

Zatem, jeżeli y jest rozwiązaniem naszego problemu z $a > 1$, to $y'(\infty) = c$, gdzie c jest dodatnim pierwiastkiem otrzymanym z równania kwadratowego (2.56).

Z drugiej strony mamy

$$y'(0) = a\gamma \frac{y(0)}{b - y(0)}.$$

Stąd gdy $y(0)$ rośnie od 0 do b, to $y'(0)$ rośnie od 0 do ∞. Co więcej, $y'(0) = c$ wtedy i tylko wtedy, gdy $y(0) = d$. Teraz podobne rozumowanie jak dla przypadku $a = 1$ prowadzi do sprzeczności. Tym samym pokazaliśmy, że problem różniczkowy (2.53) ma tylko jedno liniowe rozwiązanie ze współczynnikiem $c > 1$. To implikuje, że regresja określona przez (2.58) zachodzi z parametrem $c > 1$, a więc F jest jednoznacznie określoną funkcją rozkładu Pareto na $(0, \infty)$.

W ogólnym przypadku, gdy $\alpha > -\infty$ jest dowolne, wystarczy zastosować Lemat 1.4 przyjmując $g(x) = x - \alpha$. Otrzymamy wtedy, że równość (2.49) zachodzi wtedy i tylko wtedy, gdy

$$E\left(Y_*^{(r+2)} \mid Y_*^{(r)} = y\right) = ay + (a-1)\alpha + b$$

dla $y \in (0, \beta - \alpha)$, gdzie $Y_*^{(r)}$ oraz $Y_*^{(r+2)}$ są uogólnionymi statystykami porządkowymi z rozkładu $G(y) = F(y + \alpha)$.

Uwaga 2.7. Jak zaznaczyliśmy na początku tego podrozdziału, liniowa regresja $\xi = ax + b$ dla dowolnego $\ell \geq 3$ również charakteryzuje bazowy rozkład prawdopodobieństwa jednoznacznie. Tak więc zgodnie z Twierdzeniem 2.2 problem różniczkowy (2.28) dla $h(x) = x$ oraz $\xi(x) = ax + b$ ma jednoznaczne rozwiązanie $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_{\ell-1})$. Nie jest trudno zauważyć, że wówczas $\varphi_i(x) = c_i x + d_i$, $i = 1, \ldots, \ell - 1$ z odpowiednio dobranymi parametrami $c_1, \ldots, c_{\ell-1}$ oraz $d_1, \ldots, d_{\ell-1}$. Jednakże nie jest łatwo bezpośrednio wykazać jednoznaczność rozwiązania $\boldsymbol{\varphi}$, co stanowiłoby nowy dowód charakteryzacji rozkładów prawdopodobieństwa przez liniową regresję uogólnionych statystyk porządkowych.

2.3.3 Dyskusja możliwości zastosowań praktycznych

Na zakończenie tego rozdziału odniesiemy się do możliwych praktycznych zastosowań otrzymanych wyników. Z teoretycznego punktu widzenia Twierdzenie 2.3 jest bardzo istotne: gdy znamy tylko jedną funkcję regresji uogólnionych statystyk porządkowych, to możemy wyznaczyć bazowy rozkład jednoznacznie. Jednakże praktyczne wykorzystanie tego wyniku może być wątpliwe z uwagi na fakt, że w praktyce funkcje regresji między dwiema zmiennymi losowymi są najczęściej nieznane. Aby zastosować nasze wyniki wymagana jest znajomość analitycznej postaci regresji. Nie mniej jednak wydaje się, że pewne zastosowania można otrzymać korzystając z nieparametrycznych metod estymacji funkcji regresji, które dają zazwyczaj numeryczną aproksymację wartości regresji w ustalonych punktach.

Korzystając z tych metod nasz wynik może zostać zastosowany w następujący sposób. Zacznijmy od tego, że wybieramy konkretny model uogólnionych statystyk porządkowych. Rozważmy przykładowo zwykłe statystyki porządkowe omówione w podrozdziale 1.1.1. W pierwszym kroku obserwujemy N niezależnych prób losowych o tym samym rozmiarze n. Dla ustalonej z góry liczby $r \in \{1, \ldots, n-2\}$ obserwujemy dwie statystyki porządkowe $X_{r:n}$ oraz $X_{r+2:n}$. W ten sposób otrzymujemy N par (x_i, y_i), które mogą posłużyć do nieparametrycznej estymacji funkcji regresji $y = \xi(x) = E\big(X_{r+2:n} \mid X_{r:n} = x\big)$ z wykorzystaniem np. estymatorów jądrowych regresji. Oczywiście im większa liczba prób N tym lepsze wyniki aproksymacji funkcji ξ. Dysponując jedynie estymacją numeryczną funkcji ξ odpowiadające równanie różniczkowe $\varphi' = \gamma \xi' \frac{\varphi - x}{\xi - \varphi}$ dla $\gamma = \frac{n-r-1}{n-r}$ możemy rozwiązać również tylko numerycznie. Zatem znajdujemy tylko numeryczne przybliżenie funkcji φ. W ostatnim kroku korzystając ze wzoru (2.23), całkując numerycznie, znajdujemy nieznaną dystrybuantę F bazowego rozkładu. Tak więc, cała procedura wymaga dużej liczby obserwacji oraz przeprowadzenia wielu obliczeń numerycznych, które prowadzą tylko do przybliżonych wyników. Zatem z powyższych względów praktyczne zastosowanie naszych wyników wydaje się raczej ograniczone.

Rozdział 3

Charakteryzacje rozkładów ciągłych przez regresje uogólnionych statystyk porządkowych

W poprzednim rozdziale przedmiotem rozważań były charakteryzacje rozkładów absolutnie ciągłych o ciągłej gęstości przez regresje uogólnionych statystyk porządkowych. W takim przypadku regresja była funkcją różniczkowalną w sposób ciągły, a kryterium jednoznaczności charakteryzacji wyraziliśmy w terminach równań różniczkowych. Jednakże wyniki te nie obejmują nawet najprostszych rozkładów absolutnie ciągłych o nieciągłej gęstości, jak np. rozkładu o gęstości

$$f(x) = \begin{cases} \frac{1}{3}, & \text{dla } x \in [0,1), \\ \frac{2}{3}, & \text{dla } x \in [1,2], \\ 0, & \text{dla pozostałych } x, \end{cases} \tag{3.1}$$

rozważanego w Przykładzie 3.3 w niniejszym rozdziale. Powstaje więc pytanie, czy wyniki z poprzedniego rozdziału można uogólnić na klasę wszystkich rozkładów absolutnie ciągłych, albo nawet na klasę wszystkich rozkładów ciągłych.

W niniejszym rozdziale zakładać będziemy, że F jest ciągłą dystrybuantą rozkładu prawdopodobieństwa skupionego na nośniku (α, β). Wówczas na mocy Lematu 1.3 funkcja regresji ξ jest również ciągła, ale niekoniecznie różniczkowalna we wszystkich punktach.

W podrozdziale 3.1 podamy uogólnienie kryterium jednoznaczności z rozdziału 2, co stanowi główny wynik tej części pracy. Następnie w podrozdziale 3.2 skupimy się na wykorzystaniu nowego kryterium do badania jednoznaczności charakteryzacji przez regresje uogólnionych statystyk porządkowych o odstępie dwa ($\ell = 2$). Otrzymane wyniki zastosujemy w podrozdziale 3.3 w celu otrzymania nowych charakteryzacji rozkładów ciągłych.

Przypomnijmy, że $\bar{F}(x) = 1 - F(x)$ oraz $H(x) = -\log \bar{F}(x)$.

3.1 Kryteria jednoznaczności charakteryzacji rozkładów ciągłych

Ze wzoru (1.22) widzimy, że funkcje regresji ξ kolejnych uogólnionych statystyk porządkowych ($\ell = 1$) wyrażają się w terminach całki Riemanna-Stieltjesa. Zatem zanim przejdziemy do problemu charakteryzacyjnego przypomnimy najpierw pewne własności całki Riemanna-Stieltjesa względem funkcji monotonicznych lub funkcji o wahaniu skończonym. Własności te będą pomocne w naszych dalszych rozważaniach.

Lemat 3.1. *Niech f i g będą funkcjami ciągłymi na przedziale domkniętym $[a,b]$, zaś A będzie funkcją o wahaniu skończonym na przedziale $[a,b]$. Zdefiniujmy*

$$B(x) = \int_a^x f(t)\,\mathrm{d}A(t), \quad \text{dla } a \le x \le b.$$

Wówczas B jest również funkcją o wahaniu skończonym oraz zachodzi równość całek

$$\int_a^b g(t)\,\mathrm{d}B(t) = \int_a^b g(t)f(t)\,\mathrm{d}A(t).$$

Jest to jedna z podstawowych własności całki Riemanna-Stieltjesa, która w tym rozdziale będzie niejednokrotnie stosowana. Elementarny dowód tej własności oparty na sumach całkowych można znaleźć w monografii [35] (patrz Problemy 1.2.26 i 1.3.3). Ponadto zauważmy, że powyższy lemat jest szczególnym przypadkiem dobrze znanej własności zwanej absolutną ciągłością całki Lebesgue'a względem miary (zob. na przykład [50], Twierdzenie 1.29).

Kolejny lemat można wyprowadzić jako wniosek ze wzoru na całkowanie przez części dla całki Riemanna-Stieltjesa.

Lemat 3.2. *Jeżeli f i g są dowolnymi funkcjami ciągłymi o wahaniu skończonym na przedziale $[a,b]$ oraz $g(t) \ge c > 0$ dla $a \le t \le b$, to*

$$\frac{f(b)}{g(b)} - \frac{f(a)}{g(a)} = \int_a^b \frac{1}{g(t)}\,\mathrm{d}f(t) - \int_a^b \frac{f(t)}{(g(t))^2}\,\mathrm{d}g(t).$$

Podamy teraz lemat kluczowy dla naszych dalszych rozważań. W lemacie tym zostanie podana pewna postać funkcji regresji kolejnych uogólnionych statystyk porządkowych z rozkładu określonego przez ciągłą dystrybuantę F.

Lemat 3.3. *Dla ustalonego $r \in \mathbb{N}$, niech $X_*^{(r)}, X_*^{(r+1)}$ będą uogólnionymi statystykami porządkowymi z rozkładu o ciągłej dystrybuancie $F\colon (\alpha,\beta) \to (0,1)$. Załóżmy, że $h\colon (\alpha,\beta) \to \mathbb{R}$ jest funkcją ściśle rosnącą, ciągłą i taką, że $E\left|h\big(X_*^{(r+1)}\big)\right| < \infty$. Jeżeli*

$$\xi(x) = E\left(h\big(X_*^{(r+1)}\big) \mid X_*^{(r)} = x\right), \quad x \in (\alpha,\beta), \tag{3.2}$$

to

$$\xi(x) = \xi(x_0) + \gamma_{r+1} \int_{x_0}^x [\xi(t) - h(t)]\,\frac{\mathrm{d}F(t)}{\bar{F}(t)}, \quad x \in (\alpha,\beta), \tag{3.3}$$

gdzie x_0 jest dowolnie wybranym punktem z przedziału (α,β).

Dowód. Ustalmy $r \in \mathbb{N}$ i dla uproszczenia zapisu przyjmijmy $\gamma = \gamma_{r+1}$. Wówczas ze wzoru (1.22) mamy

$$\xi(x) = \frac{\gamma}{\bar{F}(x)^\gamma} \int_x^\beta h(t)\bar{F}(t)^{\gamma-1}\,\mathrm{d}F(t).$$

Określmy

$$f(x) = \gamma \int_x^\beta h(t)\bar{F}(t)^{\gamma-1}\,\mathrm{d}F(t), \quad g(x) = \bar{F}(x)^\gamma.$$

Korzystając teraz z Lematu 3.2 dla dowolnego przedziału $[x_0, x] \subset (\alpha, \beta)$ mamy

$$\xi(x) - \xi(x_0) = \int_{x_0}^x \frac{1}{g(t)}\,\mathrm{d}f(t) - \int_{x_0}^x \frac{f(t)}{(g(t))^2}\,\mathrm{d}g(t). \tag{3.4}$$

Dalej z Lematu 3.1 dostajemy

$$\int_{x_0}^x \frac{1}{g(t)}\,\mathrm{d}f(t) = -\gamma \int_{x_0}^x \frac{1}{\bar{F}(t)^\gamma} h(t)\bar{F}(t)^{\gamma-1}\,\mathrm{d}F(t) = -\gamma \int_{x_0}^x h(t)\frac{\mathrm{d}F(t)}{\bar{F}(t)}. \tag{3.5}$$

Ponadto

$$\int_{x_0}^x \frac{f(t)}{(g(t))^2}\,\mathrm{d}g(t) = \int_{x_0}^x \xi(t)\frac{\mathrm{d}g(t)}{g(t)} = -\gamma \int_{x_0}^x \xi(t)\frac{\mathrm{d}F(t)}{\bar{F}(t)}. \tag{3.6}$$

Wstawiając teraz równości (3.5) oraz (3.6) do (3.4) kończymy dowód lematu w przypadku, gdy $x_0 < x$. W przeciwnym przypadku wystarczy zamienić rolami x_0 oraz x, aby otrzymać żądaną równość. □

Uwaga 3.1. W Lemacie 3.3 nie zakładamy, że $h(\alpha) > -\infty$, nawet jeśli $\alpha > -\infty$. W poniższych przykładach pokażemy, że możliwe jest, że $h(\alpha) = \lim_{x\to\alpha^+} h(x) = -\infty$, a nawet $\xi(\alpha) = \lim_{x\to\alpha^+} \xi(x) = -\infty$. W przykładach będziemy rozważać statystyki porządkowe, które są uogólnionymi statystykami porządkowymi z parametrami $\gamma_r = n - r + 1$, $1 \le r \le n$. W szczególności $\gamma_n = 1$.

Przykład 3.1. Rozważmy statystyki porządkowe $X_{r:n}$, $1 \le r \le n$, z absolutnie ciągłego rozkładu F, którego funkcja gęstości f oraz funkcja przeżycia $\bar{F}$ są następujące

$$f(x) = \frac{\cos x}{2}, \quad \bar{F}(x) = \frac{1-\sin x}{2}, \quad x \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right).$$

Załóżmy, że $h(x) = \mathrm{tg}(x)$ dla $x \in (-\pi/2, \pi/2)$. Oczywiście funkcja h spełnia założenia Lematu 3.3 – jest funkcją ściśle rosnącą, ciągłą i taką, że $E\big|h\big(X_{n:n}\big)\big| < \infty$. Wówczas korzystając z równości (1.22) otrzymujemy funkcję regresji

$$\xi(x) = E\left(h(X_{n:n}) \mid X_{n-1:n} = x\right) = \frac{1}{\bar{F}(x)} \int_x^{\pi/2} h(t)f(t)\,\mathrm{d}t = \frac{\cos x}{1-\sin x}, \quad x \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right).$$

Zatem, w tym przykładzie $h(\alpha) = h(-\pi/2) = -\infty$, natomiast $\xi(-\pi/2) = 0$.

Przykład 3.2. Rozważmy statystyki porządkowe $X_{r:n}$, $1 \le r \le n$, z rozkładu jednostajnego na przedziale $(0,1)$. Niech funkcja $h\colon (0,1) \to \mathbb{R}$ będzie następującej postaci

$$h(x) = 2 - \frac{1}{x}, \quad x \in (0,1).$$

Funkcja h spełnia założenia Lematu 3.3 oraz $h(0) = \lim_{x\to 0^+} h(x) = -\infty$. Następnie korzystając ze wzoru (1.22) obliczamy funkcję regresji $h(X_{n:n})$ względem $X_{n-1:n}$

$$\xi(x) = E\left(h(X_{n:n}) \,\middle|\, X_{n-1:n} = x\right) = \frac{1}{\bar{F}(x)} \int_x^1 h(t) f(t)\,\mathrm{d}t = 2 + \frac{\log x}{1-x}, \quad x \in (0,1).$$

Zatem również $\xi(0) = \lim_{x\to 0^+} \xi(x) = -\infty$.

Uwaga 3.2. Zwykle $h(\alpha) > -\infty$, a więc również $\xi(\alpha) > -\infty$ i wtedy we wzorze (3.3) można przyjąć $x_0 = \alpha$. Taka sytuacja ma miejsce we wszystkich następnych przykładach w tym rozdziale. Poniżej sformułowane wyniki (Lemat 3.5, Twierdzenie 3.1, Lemat 3.6 oraz Twierdzenie 3.3) zostaną udowodnione przy milczącym założeniu $h(\alpha) > -\infty$. Są one jednak prawdziwe również, gdy $h(\alpha) = -\infty$, ale przy zmianie α na dowolnie ustalone $x_0 \in (\alpha, \beta)$ w równianach (3.12), (3.23), (3.31) oraz (3.34). Wymaga to odpowiednich oczywistych zmian w dowodach. Zauważmy przy tym, że wybór x_0 nie wpływałby w tym przypadku na reprezentację charakteryzowanego rozkładu F (zob. np. Lemat 3.4).

Następny lemat podaje jawny wzór na dystrybuantę F w terminach funkcji ξ oraz h w przypadku $\ell = 1$ (por. [19], Twierdzenie 4.1). Dla k-tych wartości rekordowych ($\gamma_{r+1} = k$) odpowiednik tego wyniku otrzymali Grudzień i Szynal w pracy [33], natomiast dla statystyk porządkowych ($\gamma_{r+1} = n - r$) Franco i Ruiz (por. [29], Twierdzenie 3.3).

Lemat 3.4. *Przy założeniach Lematu 3.3, jeżeli znana jest funkcja regresji ξ dana wzorem* (3.2)*, to F jest jednoznacznie wyznaczona wzorem*

$$F(x) = 1 - \exp\left(-\frac{1}{\gamma_{r+1}} \int_\alpha^x \frac{\mathrm{d}\xi(t)}{\xi(t) - h(t)}\right), \quad x \in (\alpha, \beta).$$

Dowód. Ustalmy dowolne $x_0 \in (\alpha, \beta)$. Wzór (3.3) możemy zapisać w następującej równoważnej postaci

$$\xi(x) = \xi(x_0) + \gamma_{r+1} \int_{x_0}^x [\xi(t) - h(t)]\,\mathrm{d}H(t), \quad x \in (\alpha, \beta), \tag{3.7}$$

gdzie H oznacza funkcję hazardową rozkładu F. Ze wzoru (3.7) i Lematu 3.1 otrzymujemy równość

$$\frac{1}{\gamma_{r+1}} \int_{x_0}^x \frac{\mathrm{d}\xi(t)}{\xi(t) - h(t)} = \int_{x_0}^x \mathrm{d}H(t) = H(x) - H(x_0).$$

Skoro F jest dystrybuantą ciągłą, to $F(\alpha) = 0$, a więc $H(\alpha) = 0$. Zatem kładąc $x_0 \to \alpha^+$ otrzymujemy równość

$$\frac{1}{\gamma_{r+1}} \int_\alpha^x \frac{\mathrm{d}\xi(t)}{\xi(t) - h(t)} = H(x), \tag{3.8}$$

z której możemy wyznaczyć jawny wzór na dystrybuantę F. $\square$

Uwaga 3.3. Zauważmy, że na mocy wzoru (3.8) dla dowolnego rozkładu ciągłego F, dla którego zachodzi równanie regresji (3.2) mamy

$$\int_\alpha^x \frac{\mathrm{d}\xi(t)}{\xi(t) - h(t)} < \infty, \quad x \in (\alpha, \beta), \tag{3.9}$$

oraz

$$\int_{\alpha}^{\beta} \frac{\mathrm{d}\xi(t)}{\xi(t)-h(t)} = \infty. \tag{3.10}$$

Ponadto

$$\lim_{x\to\beta^-} \xi(x)\exp\left(-\int_{\alpha}^{x} \frac{\mathrm{d}\xi(t)}{\xi(t)-h(t)}\right) = 0. \tag{3.11}$$

Ostatni warunek otrzymujemy w analogiczny sposób jak dla rozkładów absolutnie ciągłych z ciągłą funkcją gęstości (por. podrozdział 2.1) wychodząc od równości (1.22).

W dowodzie głównego twierdzenia w tym rozdziale pomocny będzie jeszcze jeden lemat będący uogólnieniem Lematu 2.2.

Lemat 3.5. *Dla ustalonych liczb naturalnych $r \geq 1$, $\ell \geq 2$, niech $\gamma_1,\ldots,\gamma_r$ będą dowolnymi parametrami oraz $0 < \gamma_{r+1} \leq \gamma_{r+i}$, $2 \leq i \leq \ell$. Przypuśćmy, że*

(a) $h,\xi\colon (\alpha,\beta) \to \mathbb{R}$ są funkcjami ściśle rosnącymi, ciągłymi i takimi, że $h(x) < \xi(x)$ dla $x \in (\alpha,\beta)$,

(b) $\boldsymbol{\tau} = (\tau_1,\ldots,\tau_{\ell-1})$ jest dowolnym rozwiązaniem układu $\ell-1$ równań całkowych

$$\tau_i(x) = \tau_i(\alpha) + \frac{\gamma_{r+i+1}}{\gamma_{r+1}} \int_{\alpha}^{x} \frac{\tau_i(t)-\tau_{i+1}(t)}{\xi(t)-\tau_1(t)}\,\mathrm{d}\xi(t), \quad x \in (\alpha,\beta), \quad 1 \leq i \leq \ell-1, \tag{3.12}$$

gdzie $\tau_\ell = h$, spełniającym warunki

$$h(x) < \tau_{\ell-1}(x) < \ldots < \tau_1(x) < \xi(x), \quad x \in (\alpha,\beta), \tag{3.13}$$

$$\int_{\alpha}^{x} \frac{\mathrm{d}\xi(t)}{\xi(t)-\tau_1(t)} < \infty, \quad x \in (\alpha,\beta), \tag{3.14}$$

$$\int_{\alpha}^{\beta} \frac{\mathrm{d}\xi(t)}{\xi(t)-\tau_1(t)} = \infty, \tag{3.15}$$

oraz

$$\lim_{x\to\beta^-} \xi(x)\exp\left(-\int_{\alpha}^{x} \frac{\mathrm{d}\xi(t)}{\xi(t)-\tau_1(t)}\right) = 0. \tag{3.16}$$

Niech $\eta(x) = \frac{\xi(x)}{\gamma_{r+1}} + \sum_{i=1}^{\ell-1} \frac{\tau_i(x)}{\gamma_{r+i+1}}$ oraz

$$\bar{G}(x) = \exp\left(-\int_{\alpha}^{x} \frac{\mathrm{d}\eta(t)}{\xi(t)-h(t)}\right), \quad x \in (\alpha,\beta). \tag{3.17}$$

Wtedy $G = 1-\bar{G}$ określa ciągłą dystrybuantę taką, że dla uogólnionych statystyk porządkowych $Y_*^{(r)}, Y_*^{(r+\ell)}$ z rozkładu G z parametrami $\gamma_1,\ldots,\gamma_{r+\ell}$ zachodzi równość

$$\xi(x) = E\left(h\big(Y_*^{(r+\ell)}\big) \mid Y_*^{(r)} = x\right), \quad x \in (\alpha,\beta).$$

Dowód. Niech $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_{\ell-1})$ będzie dowolnym rozwiązaniem układu równań całkowych (3.12) spełniającym warunki (3.13)–(3.16). Pokażemy najpierw, że funkcja G określona wzorem (3.17) jest dystrybuantą pewnego ciągłego rozkładu.

Z uwagi na warunek (3.13) funkcje podcałkowe w równaniach całkowych (3.12) są dodatnie. Skoro funkcja ξ jest z założenia ściśle rosnąca, to dla dodatnich parametrów $\gamma_{r+1}, \ldots, \gamma_{r+\ell}$ funkcje τ_i, $1 \le i \le \ell - 1$, są również ściśle rosnące na (α, β). Ponadto z ciągłości funkcji ξ oraz podstawowego twierdzenia rachunku całkowego otrzymujemy, że τ_i, $1 \le i \le \ell - 1$, są również funkcjami ciągłymi.

Z monotoniczności i ciągłości funkcji ξ, τ_i, $1 \le i \le \ell - 1$, oraz warunku (3.13) otrzymujemy natychmiast, że funkcja G jest również ściśle rosnąca i ciągła na (α, β). Oczywiście $\lim_{x\to\alpha^+} G(x) = 0$. Aby wykazać, że $\lim_{x\to\beta^-} G(x) = 1$ skorzystamy z równoważnych postaci funkcji $\bar{G}$. Na mocy Lematu 3.1 i równań całkowych (3.12) funkcja $\bar{G}$ określona wzorem (3.17) może zostać przedstawiona również w następujących równoważnych postaciach

$$\bar{G}(x) = \exp\left(-\frac{1}{\gamma_{r+i+1}} \int_\alpha^x \frac{\mathrm{d}\tau_i(t)}{\tau_i(t) - \tau_{i+1}(t)}\right), \quad x \in (\alpha, \beta), \tag{3.18}$$

dla $1 \le i \le \ell - 1$ oraz

$$\bar{G}(x) = \exp\left(-\frac{1}{\gamma_{r+1}} \int_\alpha^x \frac{\mathrm{d}\xi(t)}{\xi(t) - \tau_1(t)}\right), \quad x \in (\alpha, \beta). \tag{3.19}$$

Biorąc teraz pod uwagę ostatnią równość i warunek (3.15) otrzymujemy, że gdy $x \to \beta^-$, to $G(x) \to 1$, a zatem G jest dystrybuantą pewnego ciągłego rozkładu prawdopodobieństwa.

Niech $Y_*^{(r)}, Y_*^{(r+1)}, \ldots, Y_*^{(r+\ell)}$ oznaczają uogólnione statystyki porządkowe z rozkładu G. Pokażemy teraz, że dla $0 \le i \le \ell - 1$ oraz $x \in (\alpha, \beta)$ mamy

$$\tau_i(x) = E\big(\tau_{i+1}(Y_*^{(r+i+1)}) \mid Y_*^{(r+i)} = x\big), \tag{3.20}$$

gdzie $\tau_0 = \xi$, $\tau_\ell = h$ lub równoważnie

$$\tau_i(x)\bar{G}(x)^{\gamma_{r+i+1}} = \gamma_{r+i+1} \int_x^\beta \tau_{i+1}(t)\bar{G}(t)^{\gamma_{r+i+1}-1}\,\mathrm{d}G(t), \quad x \in (\alpha, \beta).$$

Zauważmy najpierw, że z (3.18) oraz (3.19) funkcja hazardowa rozkładu G może być przedstawiona następująco

$$H(x) = \int_\alpha^x \frac{\mathrm{d}G(t)}{\bar{G}(t)} = \frac{1}{\gamma_{r+i+1}} \int_\alpha^x \frac{\mathrm{d}\tau_i(t)}{\tau_i(t) - \tau_{i+1}(t)}, \quad x \in (\alpha, \beta), \tag{3.21}$$

dla $0 \le i \le \ell - 1$. Wybierając teraz dowolny przedział $[x, y] \subset (\alpha, \beta)$ otrzymujemy

$$\begin{aligned}
&\gamma_{r+i+1} \int_x^y \tau_{i+1}(t)\bar{G}(t)^{\gamma_{r+i+1}-1}\,\mathrm{d}G(t) \\
&\quad = -\gamma_{r+i+1} \int_x^y (\tau_i(t) - \tau_{i+1}(t))\bar{G}(t)^{\gamma_{r+i+1}} \frac{\mathrm{d}G(t)}{\bar{G}(t)} + \gamma_{r+i+1} \int_x^y \tau_i(t)\bar{G}(t)^{\gamma_{r+i+1}-1}\,\mathrm{d}G(t) \\
&\quad = -\int_x^y \bar{G}(t)^{\gamma_{r+i+1}}\,\mathrm{d}\tau_i(t) - \int_x^y \tau_i(t)\,\mathrm{d}\bar{G}(t)^{\gamma_{r+i+1}} \\
&\quad = \tau_i(x)\bar{G}(x)^{\gamma_{r+i+1}} - \tau_i(y)\bar{G}(y)^{\gamma_{r+i+1}},
\end{aligned}$$

gdzie drugą równość otrzymujemy na mocy Lematu 3.1 z (3.21), a ostatnią ze wzoru na całkowanie przez części dla całki Riemanna-Stieltjesa.

Postępując dalej jak w dowodzie Lematu 2.1 i stosując analogiczne argumenty jak w dowodzie Lematu 2.2 można wykazać, że dla $0 \le i \le \ell - 1$

$$\lim_{y \to \beta^-} \tau_i(y) \bar{G}(y)^{\gamma_{r+i+1}} = 0.$$

Tym samym prawdziwe są równości (3.20), a zatem na mocy Lematu 1.1 otrzymujemy żądaną tezę. □

Po tym wprowadzeniu podamy główny wynik tego rozdziału.

Twierdzenie 3.1. *Dla ustalonych liczb naturalnych $r \ge 1$ i $\ell \ge 2$, niech $X_*^{(r)}, X_*^{(r+\ell)}$ będą uogólnionymi statystykami porządkowymi z ciągłego rozkładu o dystrybuancie $F\colon (\alpha, \beta) \to (0,1)$. Załóżmy, że $h\colon (\alpha, \beta) \to \mathbb{R}$ jest funkcją ściśle rosnącą, ciągłą i taką, że $E\big|h\big(X_*^{(r+\ell)}\big)\big| < \infty$. Wówczas regresja*

$$\xi(x) = E\Big(h\big(X_*^{(r+\ell)}\big) \;\Big|\; X_*^{(r)} = x\Big), \quad x \in (\alpha, \beta) \tag{3.22}$$

określa rozkład prawdopodobieństwa jednoznacznie wtedy i tylko wtedy, gdy układ $\ell - 1$ równań całkowych

$$y_i(x) = y_i(\alpha) + \frac{\gamma_{r+i+1}}{\gamma_{r+1}} \int_\alpha^x \frac{y_i(t) - y_{i+1}(t)}{\xi(t) - y_1(t)} \, \mathrm{d}\xi(t), \quad 1 \le i \le \ell - 1, \tag{3.23}$$

gdzie $y_\ell = h$ ma dokładnie jedno rozwiązanie $\boldsymbol{\varphi} = (\varphi_1, \dots, \varphi_{\ell-1})$ spełniające warunki

$$h(x) < \varphi_{\ell-1}(x) < \ldots < \varphi_1(x) < \xi(x), \quad x \in (\alpha, \beta), \tag{3.24}$$

$$\int_\alpha^x \frac{\mathrm{d}\xi(t)}{\xi(t) - \varphi_1(t)} < \infty, \quad x \in (\alpha, \beta), \tag{3.25}$$

$$\int_\alpha^\beta \frac{\mathrm{d}\xi(t)}{\xi(t) - \varphi_1(t)} = \infty, \tag{3.26}$$

oraz

$$\lim_{x \to \beta^-} \xi(x) \exp\left(-\int_\alpha^x \frac{\mathrm{d}\xi(t)}{\xi(t) - \varphi_1(t)}\right) = 0. \tag{3.27}$$

Wówczas dystrybuantę F można wyrazić w postaci

$$F(x) = 1 - \exp\left(-\int_\alpha^x \frac{\mathrm{d}\eta(t)}{\xi(t) - h(t)}\right), \quad x \in (\alpha, \beta), \tag{3.28}$$

gdzie funkcja $\eta\colon (\alpha, \beta) \to \mathbb{R}$ jest określona wzorem

$$\eta(x) = \frac{\xi(x)}{\gamma_{r+1}} + \sum_{i=1}^{\ell-1} \frac{\varphi_i(x)}{\gamma_{r+i+1}}. \tag{3.29}$$

Dowód. Załóżmy, że dana jest regresja (3.22) i oznaczmy $\varphi_\ell = h$. Rozważmy funkcje φ_i, $0 \le i \le \ell - 1$, określone wzorami (1.19). Wówczas na mocy Wniosku 1.1 otrzymujemy $\varphi_0 = \xi$. Ponadto, ponieważ każda funkcja φ_i jest funkcją regresji kolejnych uogólnionych statystyk porządkowych, mianowicie regresją $\varphi_{i+1}(X_*^{(r+i+1)})$ względem $X_*^{(r+i)}$, to na mocy Lematu 1.3(iii) otrzymujemy $\varphi_{i+1} < \varphi_i$ na (α, β), dla $0 \le i \le \ell - 1$. Zatem funkcje φ_i spełniają warunek (3.24). Co więcej, na mocy własności (3.9), (3.10) oraz (3.11) dla regresji kolejnych uogólnionych statystyk porządkowych, spełnione są analogiczne warunki (3.25), (3.26) oraz (3.27).

Stosując teraz Lemat 3.3 do funkcji (1.19) otrzymujemy

$$\varphi_i(x) = \varphi_i(\alpha) + \gamma_{r+i+1} \int_\alpha^x [\varphi_i(t) - \varphi_{i+1}(t)] \frac{\mathrm{d}F(t)}{\bar{F}(t)}, \quad 1 \le i \le \ell - 1$$

oraz

$$\xi(x) = \xi(\alpha) + \gamma_{r+1} \int_\alpha^x [\xi(t) - \varphi_1(t)] \frac{\mathrm{d}F(t)}{\bar{F}(t)}.$$

Zapisując teraz funkcje φ_i, $1 \le i \le \ell - 1$ w postaci

$$\varphi_i(x) = \varphi_i(\alpha) + \frac{\gamma_{r+i+1}}{\gamma_{r+1}} \int_\alpha^x \frac{\varphi_i(t) - \varphi_{i+1}(t)}{\xi(t) - \varphi_1(t)} \gamma_{r+1} [\xi(t) - \varphi_1(t)] \frac{\mathrm{d}F(t)}{\bar{F}(t)},$$

a następnie korzystając z Lematu 3.1 (o zmianie funkcji całkującej) widzimy, że funkcje $\varphi_1, \dots, \varphi_{\ell-1}$ spełniają układ równań całkowych (3.23). Zatem, jeśli ξ jest funkcją regresji daną wzorem (3.22), to układ (3.23) ma co najmniej jedno rozwiązanie spełniające warunki (3.24)–(3.27). Ponadto dla funkcji η określonej przez (3.29) (z $\xi = \varphi_0$) mamy

$$\begin{aligned}
\int_\alpha^x \frac{\mathrm{d}\eta(t)}{\xi(t) - h(t)} &= \sum_{i=0}^{\ell-1} \frac{1}{\gamma_{r+i+1}} \int_\alpha^x \frac{1}{\xi(t) - h(t)} \mathrm{d}\varphi_i(t) \\
&= \sum_{i=0}^{\ell-1} \int_\alpha^x \frac{\varphi_i(t) - \varphi_{i+1}(t)}{\xi(t) - h(t)} \frac{\mathrm{d}F(t)}{\bar{F}(t)} \\
&= \int_\alpha^x \frac{1}{\xi(t) - h(t)} \left(\sum_{i=0}^{\ell-1} [\varphi_i(t) - \varphi_{i+1}(t)] \right) \frac{\mathrm{d}F(t)}{\bar{F}(t)} \\
&= \int_\alpha^x \frac{\mathrm{d}F(t)}{\bar{F}(t)} = H(x),
\end{aligned}$$

gdzie drugą równość dostajemy korzystając ponownie z Lematu 3.1, a ostatnia wynika z własności $\lim_{x \to \alpha^+} H(x) = 0$. Ostatnia równość wprost implikuje wzór (3.28).

Z drugiej strony, na mocy Lematu 3.5 każde rozwiązanie układu (3.23) spełniające warunki (3.24)–(3.27) określa pewien rozkład prawdopodobieństwa, dla którego zachodzi równość regresji (3.22). Jeśli zatem istnieje dokładnie jedno takie rozwiązanie, to musi ono pokrywać się z $\boldsymbol{\varphi} = (\varphi_1, \dots, \varphi_{\ell-1})$ określonym przez (1.19). Zatem F jest wyznaczona jednoznacznie. W przeciwnym razie istnieją co najmniej dwa różne rozkłady, dla których zachodzi (3.22). $\square$

Wniosek 3.1. Ustalmy $r \in \mathbb{N}$ i niech $X_*^{(r)}, X_*^{(r+2)}$ będą uogólnionymi statystykami porządkowymi z ciągłego rozkładu F określonego na przedziale (α, β). Załóżmy, że $h \colon (\alpha, \beta) \to \mathbb{R}$

jest funkcją ściśle rosnącą, ciągłą i taką, że $E\left|h\left(X_*^{(r+2)}\right)\right| < \infty$. Wówczas funkcja regresji $\xi\colon (\alpha,\beta) \to \mathbb{R}$ określona wzorem

$$\xi(x) = E\left(h\left(X_*^{(r+2)}\right) \mid X_*^{(r)} = x\right) \tag{3.30}$$

jednoznacznie charakteryzuje rozkład zadany ciągłą dystrybuantą F wtedy i tylko wtedy, gdy równanie całkowe

$$y(x) = y(\alpha) + \frac{\gamma_{r+2}}{\gamma_{r+1}} \int_\alpha^x \frac{y(t) - h(t)}{\xi(t) - y(t)}\, \mathrm{d}\xi(t) \tag{3.31}$$

ma dokładnie jedno rozwiązanie $y = \varphi(x)$ spełniające warunki

(i) $h(x) < \varphi(x) < \xi(x)$ dla $x \in (\alpha,\beta)$,

(ii) $\displaystyle\int_\alpha^x \frac{\mathrm{d}\xi(t)}{\xi(t) - \varphi(t)} < \infty$ dla $x \in (\alpha,\beta)$,

(iii) $\displaystyle\int_\alpha^\beta \frac{\mathrm{d}\xi(t)}{\xi(t) - \varphi(t)} = \infty$,

(iv) $\displaystyle\lim_{x\to\beta^-} \xi(x) \exp\left(-\int_\alpha^x \frac{\mathrm{d}\xi(t)}{\xi(t) - \varphi(t)}\right) = 0$.

Wówczas dystrybuantę F można wyznaczyć ze wzoru

$$F(x) = 1 - \exp\left(-\int_\alpha^x \frac{\mathrm{d}\eta(t)}{\xi(t) - h(t)}\right), \quad x \in (\alpha,\beta), \tag{3.32}$$

gdzie $\eta(x) = \frac{\xi(x)}{\gamma_{r+1}} + \frac{\varphi(x)}{\gamma_{r+2}}$.

Uwaga 3.4. Jeżeli dodatkowo założymy, że funkcja regresji ξ jest różniczkowalna, to z układu równań całkowych (3.23) otrzymujemy układ $\ell - 1$ równań różniczkowych

$$y_i' = \frac{\gamma_{r+i+1}}{\gamma_{r+1}} \frac{y_i - y_{i+1}}{\xi(x) - y_1} \xi'(x), \quad 1 \le i \le \ell - 1,$$

a z warunków (3.24)–(3.27) otrzymamy warunki (2.29)–(2.32). Zatem Twierdzenie 3.1 jest uogólnieniem Twierdzenia 2.2.

Podsumowując, w niniejszym podrozdziale uogólniliśmy główne wyniki rozdziału 2. Podaliśmy kryterium jednoznacznej charakteryzacji rozkładów ciągłych w terminach równań całkowych. W kolejnym podrozdziale zajmiemy się zastosowaniem tego kryterium w szczególnych przypadkach.

3.2 Jednoznaczność charakteryzacji – szczególne przypadki dla $\ell = 2$

W zastosowaniach Twierdzenia 3.1 napotykamy na podobne trudności jak w przypadku kryterium jednoznaczności charakteryzacji wyrażonego w terminach równań różniczkowych (patrz podrozdział 2.2). Równania całkowe, które otrzymaliśmy są nieliniowe ze względu na nieznaną funkcję, co więcej rozwiązanie układu równań całkowych (3.23) musi spełniać nieklasyczne warunki ograniczające (3.24)–(3.27).

W niniejszym podrozdziale zajmiemy się zagadnieniem jednoznaczności charakteryzacji ciągłych rozkładów prawdopodobieństwa przez regresje uogólnionych statystyk porządkowych o odstępie $\ell = 2$. Rozważymy problem dla dwóch szczególnych przypadków:

(i) h jest funkcją ograniczoną z góry, a więc $h(\beta) < \infty$;

(ii) h jest funkcją nieograniczoną ($h(\beta) = \infty$) oraz regresja ξ jest funkcją asymptotycznie liniową względem h, mianowicie

$$\lim_{x \to \beta^-} [\xi(x) - (ah(x) + b)] = 0, \tag{3.33}$$

dla pewnych współczynników $a \geq 1$, $b \in \mathbb{R}$.

3.2.1 Własności rozwiązań pomocniczego problemu całkowego

W niniejszym paragrafie przeanalizujemy własności rozwiązań następującego problemu całkowego

$$\begin{cases} y(x) = y(\alpha) + \gamma \int_\alpha^x \frac{y(t) - h(t)}{\xi(t) - y(t)} \, \mathrm{d}\xi(t), \\ h(x) < y(x) < \xi(x), \end{cases} \qquad x \in (\alpha, \beta), \tag{3.34}$$

gdzie $\gamma > 0$ oraz ξ, h są funkcjami ściśle rosnącymi i ciągłymi. Zatem, biorąc pod uwagę warunek $h < y < \xi$, otrzymujemy, że funkcja podcałkowa w powyższym równaniu całkowym jest dodatnia. Stąd każde rozwiązanie powyższego problemu jest również funkcją ściśle rosnącą.

Lemat 3.6. *Niech $h, \xi\colon (\alpha, \beta) \to \mathbb{R}$ będą funkcjami ściśle rosnącymi, ciągłymi i takimi, że $h < \xi$ na przedziale (α, β). Załóżmy, że $\gamma > 0$ oraz $\varphi, y\colon (\alpha, \beta) \to \mathbb{R}$ są dwoma różnymi rozwiązaniami problemu całkowego* (3.34). *Wtedy*

(i) $y(x) \neq \varphi(x)$ w każdym punkcie $x \in (\alpha, \beta)$;

(ii) albo $y < \varphi$, albo $y > \varphi$ na całym przedziale (α, β);

(iii) jeżeli $z = |y - \varphi|$, to z jest funkcją ściśle rosnącą na (α, β);

(iv) jeżeli $y < \varphi$ i $w = \frac{\varphi - y}{\varphi - h}$, to $w(\alpha) < w < 1$ oraz w jest funkcją ściśle rosnącą na (α, β);

(v) jeżeli $y > \varphi$, $w = \frac{y - \varphi}{\xi - \varphi}$ i dodatkowo $\gamma \geq 1$, to $w(\alpha) < w < 1$ oraz w jest funkcją ściśle rosnącą na (α, β).

Uwaga 3.5. Dla skrócenia zapisu będziemy pisać $\int_\alpha^x f\,\mathrm{d}g$ zamiast $\int_\alpha^x f(t)\,\mathrm{d}g(t)$, jeżeli tylko nie będzie prowadzić to do niejednoznaczności.

Dowód. (i) Dla dowolnie ustalonego $x_0 \in (\alpha,\beta)$ oznaczmy $\varphi_0 = \varphi(x_0)$. Niech $D = \{(x,y)\colon x \in (\alpha,\beta),\ h(x) < y < \xi(x)\}$. Wybierzmy $\varepsilon > 0$ tak małe, aby kula $B = B\big((x_0,\varphi_0),\varepsilon\big) \subset D$. Rozważmy funkcję $K\colon D \to (0,\infty)$ określoną wzorem

$$K(x,y) = \frac{y-h(x)}{\xi(x)-y},$$

a więc równanie całkowe w (3.34) można przedstawić w postaci

$$y(x) = y(\alpha) + \gamma\int_\alpha^x K(t,y(t))\,\mathrm{d}\xi(t). \tag{3.35}$$

Wówczas dla wszystkich y_1, y_2 takich, że $y_1 \neq y_2$, $(x,y_1) \in B$ oraz $(x,y_2) \in B$ mamy

$$\big|K(x,y_1) - K(x,y_2)\big| = \frac{\xi(x)-h(x)}{(\xi(x)-y_1)(\xi(x)-y_2)}\big|y_1-y_2\big|.$$

Stąd funkcja K spełnia warunek Lipschitza ze względu na y w otoczeniu B punktu (x_0,φ_0). Zatem korzystając z klasycznych metod możemy twierdzić, że równanie całkowe (3.35) z warunkiem początkowym $y(x_0) = \varphi_0$ ma jednoznaczne rozwiązanie w przedziale $(x_0-\delta, x_0+\delta)$ dla pewnego $\delta > 0$ wystarczająco małego. Tak więc, jeżeli y jest innym rozwiązaniem równania (3.34), to $y(x_0) \neq \varphi(x_0)$. To kończy dowód punktu (i).

(ii) Wystarczy zauważyć, że rozwiązania φ oraz y są ciągłe. Skoro ich różnica jest zawsze różna od zera, to albo jest dodatnia, albo ujemna na całym przedziale (α,β).

(iii) Najpierw zauważmy, że jeśli y i φ są rozwiązaniami problemu całkowego (3.34), to

$$y(x) - \varphi(x) = y(\alpha) - \varphi(\alpha) + \gamma\int_\alpha^x \frac{(\xi-h)(y-\varphi)}{(\xi-y)(\xi-\varphi)}\,\mathrm{d}\xi. \tag{3.36}$$

Załóżmy, że $y > \varphi$ i rozważmy $z = y - \varphi$. Ponieważ ξ jest funkcją ściśle rosnącą, a funkcja podcałkowa w (3.36) jest dodatnia na (α,β), to również z jest funkcją ściśle rosnącą na (α,β). Jeżeli $y < \varphi$, to wystarczy przyjąć $z = \varphi - y$, aby wykazać żądaną tezę.

(iv) Niech $h < y < \varphi < \xi$ oraz $w = \frac{\varphi-y}{\varphi-h}$. Kładąc $f = \varphi - y$ i $g = \varphi - h$ w Lemacie 3.2, dla dowolnego $x \in (\alpha,\beta)$ uzyskujemy

$$w(x) - w(\alpha) = \int_\alpha^x \frac{1}{\varphi-h}\,\mathrm{d}(\varphi-y) - \int_\alpha^x \frac{\varphi-y}{(\varphi-h)^2}\,\mathrm{d}(\varphi-h). \tag{3.37}$$

Pierwsza całka w równaniu (3.37) jest równa

$$\begin{aligned}\int_\alpha^x \frac{1}{\varphi-h}\,\mathrm{d}(\varphi-y) &= \gamma\int_\alpha^x \frac{(\xi-h)(\varphi-y)}{(\varphi-h)(\xi-y)(\xi-\varphi)}\,\mathrm{d}\xi \\ &= \int_\alpha^x \frac{(\xi-h)(\varphi-y)}{(\varphi-h)^2(\xi-y)}\,\mathrm{d}\varphi,\end{aligned} \tag{3.38}$$

gdzie pierwsza równość wynika z równości (3.36) zastosowanej do $\varphi - y$, a druga równość wynika z Lematu 3.1. Wstawiając teraz (3.38) do (3.37) uzyskujemy

$$w(x) - w(\alpha) = \int_{\alpha}^{x} \frac{(\varphi - y)(y-h)}{(\varphi - h)^2(\xi - y)}\, \mathrm{d}\varphi + \int_{\alpha}^{x} \frac{\varphi - y}{(\varphi - h)^2}\, \mathrm{d}h.$$

Z założenia $h < y < \varphi$ obie funkcje podcałkowe w ostatnim wyrażeniu są dodatnie. Zatem w jest funkcją ściśle rosnącą, gdyż φ i h są ściśle rosnące na (α, β).

(v) Jeżeli $h < \varphi < y < \xi$, to podobne rozumowanie przeprowadzone dla funkcji $w = \frac{y-\varphi}{\xi-\varphi}$ prowadzi do wniosku, że

$$w(x) - w(\alpha) = \int_{\alpha}^{x} \frac{y - \varphi}{\xi - \varphi}\, \mathrm{d}\psi, \tag{3.39}$$

gdzie $\psi(t) = y(t) + (\gamma - 1)\xi(t) + \varphi(t)$. Ponieważ $\gamma \geq 1$ oraz ξ, y, φ są funkcjami rosnącymi, to ψ jest funkcją rosnącą. Z założenia $\varphi < y < \xi$ wynika, że funkcja podcałkowa w równaniu (3.39) jest dodatnia. Zatem możemy wnioskować, że w jest funkcją ściśle rosnącą. □

3.2.2 Jednoznaczność charakteryzacji dla $h(\beta) < \infty$

Następne twierdzenie orzeka o jednoznaczności charakteryzacji ciągłych rozkładów prawdopodobieństwa przez regresje uogólnionych statystyk porządkowych (1.11) w przypadku, gdy $\ell = 2$ oraz h jest funkcją ograniczoną z góry, tak więc $h(\beta) < \infty$.

Twierdzenie 3.2. *Dla ustalonego $r \in \mathbb{N}$, niech $X_*^{(r)}, X_*^{(r+2)}$ będą uogólnionymi statystykami porządkowymi z rozkładu określonego przez ciągłą dystrybuantę F. Załóżmy, że $h\colon (\alpha, \beta) \to \mathbb{R}$ jest funkcją ściśle rosnącą, ciągłą i taką, że $h(\beta) < \infty$ oraz $E\left|h\left(X_*^{(r+2)}\right)\right| < \infty$. Jeżeli znamy regresję*

$$\xi(x) = E\left(h\left(X_*^{(r+2)}\right) \mid X_*^{(r)} = x\right),$$

to F jest jednoznacznie wyznaczona wzorem (3.32).

Powyższe twierdzenie jest uogólnieniem Twierdzenia 2.3 na przypadek funkcji regresji, które nie są różniczkowalne w każdym punkcie. Jego dowód jest powtórzeniem tego samego rozumowania co w dowodzie Twierdzenia 2.3 z wykorzystaniem własności rozwiązań pomocniczego problemu całkowego udowodnionych w Lemacie 3.6.

Spodziewamy się, że twierdzenie to jest również prawdziwe w przypadku dowolnego $\ell > 2$. Jednakże chcąc wykazać jednoznaczność rozwiązania układu równań całkowych (3.23) napotykamy podobne trudności jak omówione w Uwadze 2.5.

3.2.3 Jednoznaczność charakteryzacji dla $h(\beta) = \infty$

W przypadku, gdy $h(\beta) = \infty$ nie jest tak łatwo wykazać jednoznaczność rozwiązania problemu (3.34), a co za tym idzie jednoznaczność charakteryzacji rozkładu prawdopodobieństwa przez regresję ξ. Poniżej wykażemy, że analogiczny wynik jak w Twierdzeniu 3.2 zachodzi dla $h(\beta)=\infty$ przy założeniu, że funkcja regresji uogólnionych statystyk porządkowych ξ jest funkcją asymptotycznie liniową względem h, ściśle mówiąc zachodzi warunek (3.33). Oczywiście w takim przypadku na mocy założenia $\xi > h$ mamy $a \geq 1$ oraz jeżeli $a=1$, to $b \geq 0$. Ponadto, z rozważań na początku podrozdziału 2.2.3 wynika, że dla $a=1$ możemy założyć, że $b>0$ (por. warunek (2.42)). Rozpoczniemy od następującego lematu.

Lemat 3.7. *Załóżmy, że $\xi, \Phi\colon (\alpha,\beta) \to \mathbb{R}$ są funkcjami ciągłymi i ξ jest funkcją ściśle rosnącą taką, że $\lim_{x\to\beta^-}\xi(x)=\infty$. Zdefiniujmy funkcję $g\colon (\alpha,\beta)\to\mathbb{R}$ wzorem*

$$g(x) = g(\alpha) + \int_{\alpha}^{x} \Phi(t)\, \mathrm{d}\xi(t).$$

Jeżeli $\lim_{x\to\beta^-}\Phi(x) = A \in (0,\infty]$, to $\lim_{x\to\beta^-}\frac{g(x)}{\xi(x)} = A$.

Dowód. Niech $\{x_n\}_{n\geq 1}$ będzie dowolnym rosnącym ciągiem liczb z przedziału (α,β) takim, że $\lim_{n\to\infty} x_n = \beta$. Wtedy dla $n \geq 1$ uzyskujemy

$$g(x_{n+1}) - g(x_n) = \int_{x_n}^{x_{n+1}} \Phi(t)\, \mathrm{d}\xi(t).$$

Z twierdzenia o wartości średniej dla całki Riemanna-Stieltjesa wynika, że dla każdego $n \geq 1$ istnieje punkt $\theta_n \in (x_n, x_{n+1})$ taki, że

$$\frac{g(x_{n+1}) - g(x_n)}{\xi(x_{n+1}) - \xi(x_n)} = \Phi(\theta_n).$$

Oczywiście $\theta_n \to \beta$, gdy $n\to\infty$, tak więc z przyjętego założenia dla funkcji Φ otrzymujemy, że $\Phi(\theta_n)\to A$, gdy $n\to\infty$. Z klasycznego twierdzenia Stolza wynika, że

$$\lim_{n\to\infty}\frac{g(x_n)}{\xi(x_n)} = A,$$

co kończy dowód lematu. $\square$

Twierdzenie 3.3. *Niech $h, \xi\colon (\alpha,\beta)\to\mathbb{R}$ będą ściśle rosnącymi, ciągłymi funkcjami takimi, że $h<\xi$, $h(\beta)=\infty$ oraz ξ spełnia warunek* (3.33) *dla pewnych $a\geq 1$, $b\in\mathbb{R}$. Załóżmy, że $\gamma \geq 1$ oraz problem całkowy* (3.34) *ma rozwiązanie φ takie, że*

$$\lim_{x\to\beta^-}\left[\varphi(x) - (ch(x)+d)\right] = 0, \tag{3.40}$$

dla pewnych $c\geq 1$, $d\in\mathbb{R}$.

(a) Jeżeli $a=1$, to $c=1$, $d=\frac{b}{1+\gamma}>0$, a ponadto φ jest jedynym rozwiązaniem problemu całkowego (3.34).

(b) Jeżeli $a > 1$, to c jest wyznaczone jako jedyne rozwiązanie równania kwadratowego

$$c^2 + a(\gamma - 1)c - a\gamma = 0 \tag{3.41}$$

w przedziale $(1, a)$, a ponadto φ jest jedynym rozwiązaniem problemu całkowego (3.34).

Dowód. (a) Z warunku $h < \varphi < \xi$, dla dostatecznie dużych x, dostajemy

$$1 < \frac{\varphi(x)}{h(x)} < \frac{\xi(x)}{h(x)}.$$

Zatem, jeżeli $a = 1$, to oczywiście $c = 1$. Z równości (3.33) i (3.40) dostajemy $\lim_{x\to\beta^-} \frac{\varphi(x)}{\xi(x)} = 1$ oraz $\lim_{x\to\beta^-} [\xi(x) - \varphi(x)] = b - d > 0$. Kładąc

$$\Phi(x) = \frac{\varphi(x) - h(x)}{\xi(x) - \varphi(x)} \tag{3.42}$$

mamy $\Phi(x) \to \frac{d}{b-d}$, gdy $x \to \beta^-$ oraz

$$\varphi(x) - \varphi(\alpha) = \gamma \int_\alpha^x \Phi(t)\,\mathrm{d}\xi(t).$$

Stosując teraz Lemat 3.7 otrzymujemy

$$\lim_{x\to\beta^-} \frac{\varphi(x)}{\xi(x)} = \gamma \frac{d}{b-d}.$$

Z drugiej strony wiemy, że granica ta jest równa 1, zatem

$$1 = \gamma \frac{d}{b-d}. \tag{3.43}$$

Stąd $d = \frac{b}{1+\gamma}$ jest liczbą dodatnią.

Załóżmy teraz, że y jest innym rozwiązaniem problemu (3.34) spełniającym warunek $h < y < \xi$. Wtedy dla dostatecznie dużych x mamy

$$1 < \frac{y(x)}{h(x)} < \frac{\xi(x)}{h(x)},$$

a zatem również $\lim_{x\to\beta^-} \frac{y(x)}{h(x)} = 1$ oraz $\lim_{x\to\beta^-} \frac{y(x)}{\xi(x)} = 1$. Z Lematu 3.6(ii) wiemy, że albo $y < \varphi$, albo $\varphi < y$. Jeżeli $\varphi < y < \xi$, to rozważmy funkcję pomocniczą $z = y - \varphi$. Z Lematu 3.6(iii) funkcja z rośnie od $z(\alpha) > 0$ do dodatniej stałej $C \in (z(\alpha), b-d]$. Załóżmy najpierw, że $C < b - d$. Definiując

$$\Psi = \frac{z + \varphi - h}{\xi - \varphi - z}, \tag{3.44}$$

a następnie przechodząc do granicy, gdy $x \to \beta^-$ dostajemy $\Psi(x) \to \frac{C+d}{b-d-C}$. Stosując ponownie Lemat 3.7 do całki

$$y(x) - y(\alpha) = \gamma \int_\alpha^x \Psi(t)\,\mathrm{d}\xi(t),$$

oraz uwzględniając fakt, że $\frac{y(x)}{\xi(x)} \to 1$, gdy $x \to \beta^-$ otrzymujemy

$$1 = \gamma \frac{C+d}{b-d-C}.$$

Porównując powyższą równość z (3.43) dostajemy $C = 0$, co daje sprzeczność. Gdyby natomiast $C = b-d$, to $\Psi(x) \to \infty$, gdy $x \to \beta^-$. Z Lematu 3.7 otrzymamy $\frac{y(x)}{\xi(x)} \to \infty$, co przeczy temu, że $y < \xi$ na (α, β).

Jeżeli $h < y < \varphi$, to wystarczy rozważyć funkcję $z = \varphi - y$, aby dojść do kolejnej sprzeczności. Zatem funkcja φ jest jedynym rozwiązaniem problemu (3.34).

(b) Ogólna idea dowodu dla $a > 1$ jest taka sama jak dla $a = 1$. Poniżej przedstawiamy szkic tego dowodu. Zauważmy, że $\lim_{x\to\beta^-} \frac{\xi(x)}{h(x)} = a$ oraz $\lim_{x\to\beta^-} \frac{\varphi(x)}{h(x)} = c$. Stąd

$$\lim_{x\to\beta^-} \frac{\varphi(x)}{\xi(x)} = \frac{c}{a}. \tag{3.45}$$

Oczywiście $1 \le c \le a$. Pokażemy najpierw, że $c \in (1, a)$. Przedstawmy funkcję Φ określoną przez (3.42) w następującej postaci

$$\Phi(x) = \frac{\frac{\varphi(x)}{h(x)} - 1}{\frac{\xi(x)}{h(x)} - \frac{\varphi(x)}{h(x)}}.$$

Jeśli $c = 1$, to $\lim_{x\to\beta^-} \Phi(x) = 0$, a gdyby $c = a$, to $\lim_{x\to\beta^-} \Phi(x) = \infty$. W obu przypadkach, na mocy Lematu 3.7, dochodzimy do sprzeczności z równością (3.45). Zatem $c \in (1, a)$ i gdy $x \to \beta^-$, to $\Phi(x) \to \frac{c-1}{a-c}$. Korzystając po raz kolejny z Lematu 3.7 dostajemy $\frac{\varphi(x)}{\xi(x)} \to \gamma \frac{c-1}{a-c}$, co w zestawieniu z warunkiem (3.45) prowadzi do równania kwadratowego (3.41). Zauważmy, że równanie to ma dwa pierwiastki różnych znaków i oznaczmy jego lewą stronę przez $f(c)$. Ponieważ $a > 1$, to $f(1) = 1 - a < 0$ oraz $f(a) = a\gamma(a-1) > 0$. Zatem równanie (3.41) ma dokładnie jedno rozwiązanie w przedziale $(1, a)$.

Załóżmy teraz, że y jest innym rozwiązaniem problemu (3.34) oraz, że istnieje granica $\lim_{x\to\beta^-} \frac{y(x)}{h(x)}$. Oznaczmy tę granicę przez $L(y)$. Wówczas ponownie korzystając z Lematu 3.7 dostajemy, że $L(y) \in (1, a)$ oraz $\frac{L(y)}{a} = \gamma \frac{L(y)-1}{a-L(y)}$. Ale to równanie określa jednoznacznie c, zatem $L(y) = c$ dla dowolnego y spełniającego warunek (3.34).

Dla przypadku $\varphi < y < \xi$ rozważmy funkcję $w = \frac{y-\varphi}{\xi-\varphi}$. Z Lematu 3.6(v) funkcja w rośnie od $w(\alpha) > 0$ do $C \in (w(\alpha), 1]$. Zauważmy, że Ψ określona przez (3.44) może zostać przedstawiona w postaci

$$\Psi = \frac{\frac{y-\varphi}{\xi-\varphi} + \Phi}{1 - \frac{y-\varphi}{\xi-\varphi}} = \frac{w + \Phi}{1 - w}.$$

Załóżmy najpierw, że $C < 1$. Wtedy granica funkcji Ψ w punkcie β istnieje i jest równa $\frac{C + \frac{c}{\gamma a}}{1-C}$. Z drugiej strony

$$y(x) = y(\alpha) + \gamma \int_\alpha^x \Psi(t)\,\mathrm{d}\xi(t),$$

zatem korzystając z Lematu 3.7 otrzymujemy istnienie granicy $\lim_{x\to\beta^-}\frac{y(x)}{\xi(x)}=\gamma\frac{C+\frac{c}{\gamma a}}{1-C}$. Z przedstawienia $\frac{y}{\xi}=\frac{y}{h}\cdot\frac{h}{\xi}$ otrzymujemy istnienie granicy $\lim_{x\to\beta^-}\frac{y(x)}{h(x)}$, co implikuje

$$\frac{c}{a}=\gamma\frac{C+\frac{c}{\gamma a}}{1-C}.$$

Ale powyższa równość zachodzi tylko dla $C=0$, co daje sprzeczność. Gdyby natomiast $C=1$, to $\Psi(x)\to\infty$. Wówczas na mocy Lematu 3.7 otrzymamy $\frac{y(x)}{\xi(x)}\to\infty$, gdy $x\to\beta^-$. To przeczy temu, że $y<\xi$ na (α,β). Stąd dla przypadku $\varphi<y$, wykazaliśmy, że φ jest jedynym rozwiązaniem problemu (3.34).

Dla przypadku $h<y<\varphi$ wystarczy przeprowadzić analogiczne rozważania wykorzystując własności funkcji $w=\frac{\varphi-y}{\varphi-h}$, aby dojść do kolejnej sprzeczności. □

Uwaga 3.6. Zaznaczmy, że warunek $\gamma\geq 1$ jest wymagany tylko w dowodzie punktu (b) powyższego twierdzenia. W kolejnym podrozdziale będziemy stosować to twierdzenie z $\gamma=\frac{\gamma_{r+2}}{\gamma_{r+1}}$. Zgodnie z Uwagą 1.2 nie zmniejsza to ogólności rozważań.

Z Twierdzeń 3.1 i 3.3 otrzymujemy jednoznaczną charakteryzację rozkładu prawdopodobieństwa przez asymptotycznie liniowe regresje uogólnionych statystyk porządkowych o odstępie $\ell=2$.

Twierdzenie 3.4. *Dla ustalonego $r\in\mathbb{N}$, niech $X_*^{(r)},X_*^{(r+2)}$ będą uogólnionymi statystykami porządkowymi z rozkładu określonego przez ciągłą dystrybuantę F. Załóżmy, że $h\colon(\alpha,\beta)\to\mathbb{R}$ jest funkcją ściśle rosnącą, ciągłą i taką, że $E\left|h\big(X_*^{(r+2)}\big)\right|<\infty$ oraz $h(\beta)=\infty$. Jeżeli regresja ξ określona wzorem* (3.30) *spełnia warunek* (3.33) *dla pewnych $a\geq 1$, $b\in\mathbb{R}$ oraz problem całkowy* (3.34) *ma rozwiązanie φ, które spełnia warunek* (3.40)*, to funkcja regresji ξ jednoznacznie wyznacza ciągły rozkład prawdopodobieństwa, którego dystrybuanta F dana jest wzorem* (3.32).

3.3 Przykłady nowych charakteryzacji rozkładów ciągłych

W niniejszym podrozdziale podamy nowe charakteryzacje ciągłych rozkładów przez pojedynczą funkcję regresji wartości rekordowych o odstępie dwa ($\ell=2$). Jako narzędzi do ich otrzymania użyjemy wyników z poprzednich podrozdziałów.

Podobnie jak w podrozdziale 2.3, dla uproszczenia zakładać będziemy, że $h(x)=x$ (por. Lemat 1.4) oraz rozważać będziemy przypadek wartości rekordowych. Zatem, dla ustalonego naturalnego $r\geq 1$, rozważamy następującą funkcję regresji

$$\xi(x)=E\big(R_{r+2}\mid R_r=x\big).$$

Wychodząc od równania (1.22) dla $\gamma_{r+1} = 1$ otrzymujemy następującą postać regresji kolejnych wartości rekordowych

$$\varphi(x) = E\big(R_{r+2} \mid R_{r+1} = x\big) = \frac{1}{\bar{F}(x)} \int_x^\infty t \,\mathrm{d}F(t). \tag{3.46}$$

Stosując Wniosek 1.1 dla $\ell = 2$ dostajemy

$$\xi(x) = E\big(\varphi(R_{r+1}) \mid R_r = x\big) = \frac{1}{\bar{F}(x)} \int_x^\infty \varphi(t) \,\mathrm{d}F(t). \tag{3.47}$$

Odpowiedniki tych wzorów dla różniczkowalnych funkcji stanowiły podstawę obliczeń w podrozdziale 2.3 (por. wzory (2.45) i (2.46)).

W pierwszym przykładzie pokażemy, że dla rozkładu z ciągłą dystrybuantą F i nieciągłą funkcją gęstości f, funkcja regresji ξ w niektórych punktach nie musi być różniczkowalna, ale zachodzi jednoznaczność charakteryzacji.

Przykład 3.3. Rozważmy rozkład, którego funkcją gęstości jest określona wzorem (3.1) a odpowiadająca jej dystrybuanta jest równa

$$F(x) = \begin{cases} \frac{x}{3}, & \text{dla } x \in [0,1), \\ \frac{2x-1}{3}, & \text{dla } x \in [1,2]. \end{cases} \tag{3.48}$$

Korzystając ze wzoru (3.46) i wykonując elementarne obliczenia otrzymujemy

$$\varphi(x) = \begin{cases} \frac{x^2-7}{2(x-3)}, & \text{dla } x \in [0,1), \\ \frac{x}{2} + 1, & \text{dla } x \in [1,2]. \end{cases}$$

Następnie ze wzoru (3.47) dostajemy

$$\xi(x) = \begin{cases} \frac{1}{4(x-3)} \left(x^2 + 6x - 21 + 4\log\left(\frac{3-x}{2}\right)\right), & \text{dla } x \in [0,1), \\ \frac{x+6}{4}, & \text{dla } x \in [1,2]. \end{cases} \tag{3.49}$$

Zauważmy, że funkcja ξ jest ciągła na całym przedziale $x \in (0,2)$, ale nie jest różniczkowalna w punkcie $x = 1$. W przykładzie tym mamy $\beta = 2$, co implikuje $h(\beta) = \xi(\beta) < \infty$. Zatem jednoznaczność charakteryzacji ciągłego rozkładu (3.48) przez funkcję regresji (3.49) wynika wprost z Twierdzenia 3.2. Jednakże z uwagi na nieciągłość funkcji gęstości, jednoznaczność ta nie może być wywnioskowana z Twierdzenia 2.3.

W kolejnym przykładzie podamy nową charakteryzację pewnego rozkładu gamma przez regresję wartości rekordowych. Przypomnijmy, że rozkład Gamma(k,θ) z parametrami $k,\theta > 0$ ma funkcję gęstości daną wzorem

$$f(x) = \frac{\theta^k}{\Gamma(k)} x^{k-1} \exp(-\theta x), \quad x > 0,$$

gdzie Γ oznacza funkcje gamma Eulera.

Przykład 3.4. Rozważmy rozkład Gamma$(2,1)$, którego funkcja gęstości i funkcja przeżycia są następujące $f(x) = x\exp(-x)$, $\bar{F}(x) = (1+x)\exp(-x)$ dla $x > 0$. Wprowadźmy funkcję pomocniczą

$$E(x) = \frac{\exp(x)}{x} \int_x^\infty \frac{\exp(-t)}{t}\,\mathrm{d}t, \quad x > 0.$$

Wykonując elementarne obliczenia otrzymujemy równości

$$E'(x) = \left(1 - \frac{1}{x}\right) E(x) - \frac{1}{x^2}$$

oraz $\lim_{x\to\infty} E(x) = \lim_{x\to\infty} E'(x) = 0$.

Wychodząc teraz od równości (3.46) dostajemy $\varphi(x) = x + 1 + \frac{1}{x+1}$, a następnie ze wzoru (3.47) mamy

$$\xi(x) = x + 2 + \frac{2}{x+1} - E(x+1), \quad x > 0.$$

Teraz chcemy wykazać, że powyższa funkcja regresji charakteryzuje rozkład Gamma$(2,1)$ jednoznacznie. Ponieważ $h(\beta) = \infty$, to nie możemy jak w poprzednim przykładzie zastosować Twierdzenia 3.2. Można łatwo sprawdzić, że $\varphi(x) = x + 1 + \frac{1}{x+1}$ jest rozwiązaniem problemu

$$\varphi'(x) = \frac{\varphi(x) - x}{\xi(x) - \varphi(x)}\,\xi'(x) \tag{3.50}$$

takim, że $x < \varphi(x) < \xi(x)$ dla $x > 0$. Ponadto

$$\lim_{x\to\infty} [\xi(x) - (x+2)] = 0, \quad \lim_{x\to\infty} [\varphi(x) - (x+1)] = 0.$$

Zatem na mocy Twierdzenia 3.3(a) dostajemy, że φ jest jedynym rozwiązaniem problemu (3.50). Ostatecznie z Twierdzenia 3.4 otrzymujemy, że rozkład Gamma$(2,1)$ jest jedynym rozkładem, dla którego spełniony jest warunek

$$E\big(R_{r+2} \mid R_r = x\big) = x + 2 + \frac{2}{x+1} - E(x+1), \quad x > 0.$$

W następnym przykładzie uogólnimy wnioski z Przykładu 3.4. Pokażemy, że każdy rozkład Gamma(k,θ) jest jednoznacznie określony przez odpowiadającą mu regresję uogólnionych statystyk porządkowych $\xi(x) = E\big(X_*^{(r+2)} \mid X_*^{(r)} = x\big)$. Jednakże, w przeciwieństwie do poprzedniego przykładu, z uwagi na skomplikowane obliczenia, nie podajemy jawnych postaci funkcji ξ oraz $\varphi(x) = E\big(X_*^{(r+2)} \mid X_*^{(r+1)} = x\big)$.

Przykład 3.5. Przypomnijmy, że dla dystrybuanty F z gęstością f

$$\lambda(x) = \frac{f(x)}{\bar{F}(x)}, \quad x \in (\alpha, \beta)$$

oznacza intensywność awarii rozkładu F. Dla rozkładu Gamma(k,θ) można łatwo sprawdzić, że dla dowolnego $k > 0$ mamy

$$\lim_{x\to\infty} \lambda(x) = \theta.$$

Przeprowadzając dalej rozumowanie jak w dowodzie Lematu 2.4 dostajemy, że

$$\lim_{x\to\infty}[\varphi(x)-x] = \frac{1}{\gamma_{r+2}}\lim_{x\to\infty}\frac{1}{\lambda(x)} = \frac{1}{\theta\gamma_{r+2}}$$

oraz podobnie

$$\lim_{x\to\infty}[\xi(x)-x] = \left(\frac{1}{\gamma_{r+1}}+\frac{1}{\gamma_{r+2}}\right)\lim_{x\to\infty}\frac{1}{\lambda(x)} = \frac{1}{\theta}\left(\frac{1}{\gamma_{r+1}}+\frac{1}{\gamma_{r+2}}\right).$$

Zatem dla dowolnego rozkładu gamma odpowiadające mu regresje uogólnionych statystyk porządkowych ξ oraz φ spełniają odpowiednio warunki (3.33) oraz (3.40) z $a=c=1$. Co więcej, z dowodu Twierdzenia 3.1 wiemy, że φ jest rozwiązaniem problemu (3.34). Z Twierdzenia 3.4 dostajemy, że jest to jedyne rozwiązanie problemu (3.34), a zatem rozkład Gamma(k,θ) jest jednoznacznie charakteryzowany przez odpowiadającą mu funkcję regresji ξ.

Uwaga 3.7. Podobna sytuacja jak w Przykładzie 3.5 dla rozkładu gamma ma miejsce również dla innych rozkładów. Poniżej podamy kilka dobrze znanych rozkładów, dla których funkcja intensywności awarii ma skończoną dodatnią granicę. Zatem w analogiczny sposób jak w Przykładzie 3.5 możemy wykazać jednoznaczność charakteryzacji tych rozkładów przez odpowiadające im regresje ξ.

- Rozkład wykładniczy ze stałą funkcją intensywności awarii $\lambda(x)=\theta$. Jest on szczególnym przypadkiem rozkładu gamma z parametrem $k=1$, jak również szczególnym przypadkiem rozkładu Weibulla z parametrem $\delta=1$ (por. wzór (2.48)).

- Rozkład Gumbela z parametrami $\theta>0$, $\mu\in\mathbb{R}$ określony przez

  $$f(x)=\theta G(x)\exp[-G(x)], \quad \bar{F}(x)=1-\exp[-G(x)], \quad x\in\mathbb{R},$$

  gdzie $G(x)=\exp[-\theta(x-\mu)]$, dla którego

  $$\lambda(x)=\frac{\theta G(x)}{\exp[G(x)]-1} \quad \text{oraz} \quad \lim_{x\to\infty}\lambda(x)=\theta.$$

- Rozkład logistyczny z parametrami $\theta>0$, $\mu\in\mathbb{R}$ określony przez

  $$f(x)=\theta\frac{G(x)}{[1+G(x)]^2}, \quad \bar{F}(x)=\frac{G(x)}{1+G(x)}, \quad x\in\mathbb{R},$$

  gdzie funkcja G jest określona jak w poprzednim punkcie, dla którego

  $$\lambda(x)=\frac{\theta}{1+G(x)} \quad \text{oraz} \quad \lim_{x\to\infty}\lambda(x)=\theta.$$

Z drugiej strony dla wielu rozkładów prawdopodobieństwa mamy $\lim_{x\to\infty}\lambda(x)=0$, a więc $\lim_{x\to\infty}[\xi(x)-x]=\infty$. Tak jest m.in. dla rozkładu Weibulla z $\delta<1$, Pareto czy rozkładu logarytmiczno-normalnego. W ogólności, pytanie o charakteryzacje tych rozkładów przez regresje uogólnionych statystyk porządkowych pozostaje otwarte.

Zauważmy, że w Przykładach 3.4 i 3.5 funkcje regresji uogólnionych statystyk porządkowych spełniały warunek asymptotycznej liniowości (3.33) z parametrem $a = 1$. W kolejnym przykładzie rozważymy rozkład prawdopodobieństwa, dla którego w warunku (3.33) mamy współczynnik $a > 1$.

Przykład 3.6. Oznaczmy $D(x) = x^2+2x+2$ i rozważmy rozkład określony następującą funkcją gęstości $f(x) = 2\sqrt{2}(1+4x+2x^2)D(x)^{-\frac{5}{2}}$ dla $x > 0$. Funkcja przeżycia $\bar{F}$ dana jest wzorem $\bar{F}(x) = 2\sqrt{2}(1+x)D(x)^{-\frac{3}{2}}$ dla $x > 0$.

Ponownie ze wzorów (3.46) i (3.47) wyznaczamy regresje

$$\varphi(x) = 2x+1+\frac{1}{x+1}$$

oraz

$$\xi(x) = \frac{1}{x+1}\left(8+9x+5x^2+D(x)^{\frac{3}{2}}\log\left[\frac{1+x}{1+\sqrt{D(x)}}\right]\right). \tag{3.51}$$

Elementarne, ale żmudne obliczenia pokazują, że $\lim_{x\to\infty}[\xi(x)-(4x+3)] = 0$ oraz $\lim_{x\to\infty}[\varphi(x)-(2x+1)] = 0$. Zatem na mocy Twierdzenia 3.4 otrzymujemy, że F jest jedynym rozkładem, dla którego zachodzi tożsamość (3.51).

W ostatnim przykładzie pokażemy, że w ogólnym przypadku regresja ξ nie musi być asymptotyczne liniowa względem h, a zatem problem jednoznacznej charakteryzacji rozkładów przez dowolną funkcję (3.30) pozostaje otwarty.

Przykład 3.7. Rozważmy rozkład prawdopodobieństwa określony przed dystrybuantę F daną wzorem

$$F(x) = 1-\frac{\mathrm{e}}{x(\log x)^{1+a}}, \quad x \geq \mathrm{e},$$

gdzie $a \geq 0$. W tym przypadku, korzystając z wyniku Nagaraja [44] otrzymujemy, że $E(R_n) < \infty$ wtedy i tylko wtedy, gdy $n < a+1$. Stąd $E(R_3) < \infty$, o ile $a > 2$. Zatem dla $h(x) = x$ otrzymujemy kolejno

$$\varphi(x) = E\big(R_3 \mid R_2 = x\big) = x\left(1+\frac{1}{a}\log x\right)$$

oraz

$$\xi(x) = E\big(R_3 \mid R_1 = x\big) = x\left(1+\frac{1+2a}{a^2}\log x+\frac{1}{a(a-1)}(\log x)^2\right).$$

Można sprawdzić, że φ jest rozwiązaniem problemu (3.34) takim, że $x < \varphi(x) < \xi(x)$ dla $x > \mathrm{e}$, ale nie mamy podstaw aby twierdzić, że jest to jedyne rozwiązanie tego problemu. W tym przykładzie ξ nie spełnia warunku (3.33), zatem nie możemy zastosować Twierdzenia 3.4, aby wykazać, że powyższa funkcja regresji charakteryzuje rozkład F jednoznacznie.

Na koniec tego podrozdziału podamy jeszcze jedno zastosowanie wyników z podrozdziału 3.2.

Uwaga 3.8. Wiemy, że w przypadku gdy regresja ξ jest liniowa (2.49), to problem różniczkowy (2.53) ma rozwiązanie liniowe (2.54). Zamieniając problem (2.53) na odpowiadający problem całkowy widzimy, że spełnione są założenia Twierdzenia 3.4. Stąd ξ jednoznacznie wyznacza rozkład prawdopodobieństwa, którego dystrybuanta F dana jest wzorem (3.32). Powyższe rozumowanie stanowi kolejny dowód jednoznacznej charakteryzacji rozkładów prawdopodobieństwa przez liniową regresję uogólnionych statystyk porządkowych o odstępie dwa.

Rozdział 4

Charakteryzacje dyskretnych rozkładów prawdopodobieństwa przez regresje słabych wartości rekordowych

W niniejszym rozdziale rozważać będziemy problem charakteryzacji dyskretnych rozkładów prawdopodobieństwa przez regresje słabych wartości rekordowych. Skupimy się głównie na regresji słabych wartości rekordowych, gdyż jak wiemy są one poprawnie określone również dla ciągów zmiennych losowych o skończonych nośnikach. Jak już wyjaśniliśmy w podrozdziale 1.3 jest to również spowodowane faktem, że uogólnione statystyki porządkowe z rozkładów dyskretnych nie mają własności Markowa.

Na początku tego rozdziału przyjrzymy się wybranym funkcjom, które charakteryzują dyskretne rozkłady prawdopodobieństwa. Następnie podamy warunek konieczny i dostateczny dla jednoznacznej charakteryzacji dyskretnego rozkładu prawdopodobieństwa przez regresje słabych rekordów określonych wzorem (1.13), co stanowi główny wynik tego rozdziału. W kolejnych podrozdziałach zastosujemy uzyskane wyniki do otrzymania nowych charakteryzacji rozkładów dyskretnych, jak i wyprowadzenia znanych charakteryzacji przy użyciu nowego podejścia przedstawionego w niniejszym rozdziale. W ostatnim podrozdziale odniesiemy się do problemu charakteryzacji rozkładów przez regresje dyskretnych rekordów, które jak wiemy charakteryzują wyłącznie ogony rozkładów o nieskończonym nośniku.

W dalszym ciągu (por. podrozdział 1.4.2) zakładamy, że $\{X_n, n \geq 1\}$ jest ciągiem niezależnych zmiennych losowych o jednakowym dyskretnym rozkładzie określonym przez niezdegenerowaną dystrybuantę F o nośniku $S = \{0, 1, \ldots, N\}$, gdzie $N \leq \infty$. Ponadto dla $N < \infty$ przyjmujemy $\bar{S} = S \setminus \{N\}$, a dla $N = \infty$ przyjmujemy konwencję $\bar{S} = S$ oraz $N - k = \infty$, $k = 0, 1, \ldots$.

Niech $p_k = P(X_1 = k)$ oraz $q_k = P(X_1 \geq k) = \sum_{j=k}^{N} p_j$ dla $k \in S$ określa rozkład prawdopodobieństwa dyskretnej zmiennej losowej o wartościach w zbiorze S.

4.1 Uwagi wstępne

W tym podrozdziale przedstawimy wybrane funkcje charakteryzujące dyskretne rozkłady prawdopodobieństwa. Funkcje te i ich własności będą wykorzystywane w następnych podrozdziałach.

Niech $\bar{F} = 1 - F$ oznacza *funkcję przeżycia rozkładu F*

$$\bar{F}(k) = 1 - F(k) = P(X > k), \quad k \in S.$$

Wówczas $p_0 = 1 - \bar{F}(0)$, $p_n = \bar{F}(n-1) - \bar{F}(n)$, dla $n = 1, 2, \ldots, N$. Ponadto $\bar{F}(n) = q_{n+1}$. Oczywiście, gdy $N < \infty$, to $\bar{F}(n) = 0$ dla $n \geq N$. Dlatego, w dalszym ciągu nie będziemy szczegółowo specyfikować F w tym przypadku.

Rozważmy następnie *funkcję intensywności awarii rozkładu F* określoną wzorem

$$\lambda(k) = \frac{p_k}{q_k}, \quad k \in S.$$

Stąd dostajemy

$$p_0 = \lambda(0), \quad p_n = \lambda(n) \prod_{k=0}^{n-1} (1 - \lambda(k)), \quad n = 1, 2, \ldots, N. \tag{4.1}$$

Co więcej, funkcja $\bar{F}$ jest jednoznacznie wyznaczona przez funkcję λ wzorem

$$\bar{F}(n) = \prod_{k=0}^{n} (1 - \lambda(k)), \quad n \in S. \tag{4.2}$$

Zauważmy, że w przypadku, gdy $N < \infty$ mamy $p_N = q_N$, zatem $\lambda(N) = 1$ i $\bar{F}(N) = 0$.

Dalej będziemy używać również funkcji zdefiniowanej następująco

$$\mu(k) = \frac{p_k}{q_{k+1}}, \quad k \in \bar{S}, \tag{4.3}$$

którą nazywać będziemy *przesuniętą intensywnością awarii rozkładu F*. W przypadku, gdy nośnik jest skończony, $N < \infty$, przyjmujemy następującą konwencję $\mu(N) = \infty$ oraz $\frac{\mu(N)}{1+\mu(N)} = 1$. Zatem możemy zapisać, że między funkcjami λ i μ zachodzi następująca relacja

$$\lambda(k) = \frac{\mu(k)}{1 + \mu(k)}, \quad k \in S. \tag{4.4}$$

Stąd oraz z równości (4.2) dostajemy, że funkcja μ jednoznacznie wyznacza $\bar{F}$ wzorem

$$\bar{F}(n) = \prod_{k=0}^{n} (1 + \mu(k))^{-1}, \quad n \in S. \tag{4.5}$$

Ponadto z (4.1) funkcja rozkładu prawdopodobieństwa zmiennej losowej X ma postać

$$p_0 = \frac{\mu(0)}{1 + \mu(0)}, \quad p_n = \frac{\mu(n)}{1 + \mu(n)} \prod_{k=0}^{n-1} (1 + \mu(k))^{-1}, \quad n = 1, 2, \ldots, N. \tag{4.6}$$

Zauważmy, że dla dyskretnego rozkładu o nieskończonym nośniku (tj. $N = \infty$), szereg $\sum_{k=0}^{\infty} \frac{p_k}{q_k}$ jest rozbieżny (zob. [51], Twierdzenie 1). Zatem ze wzoru (4.3) i z kryterium porównawczego zbieżności szeregów dostajemy, że szereg $\sum_{k=0}^{\infty} \frac{p_k}{q_{k+1}}$ jest również rozbieżny.

Poniższy lemat podaje warunek konieczny i wystarczający na to, aby pewna funkcja określona na zbiorze $S = \{0, 1, \dots\}$ była intensywnością awarii pewnego rozkładu prawdopodobieństwa o nośniku S.

Lemat 4.1. *([51, 52]) Niech $S = \{0, 1, \dots\}$. Funkcja $\lambda : S \to [0, 1)$ jest intensywnością awarii pewnego rozkładu prawdopodobieństwa o nośniku S wtedy i tylko wtedy, gdy*

$$\sum_{k=0}^{\infty} \lambda(k) = \infty. \tag{4.7}$$

Uwaga 4.1. Rozbieżność szeregu (4.7) oznacza, że iloczyn nieskończony $\prod_{k=0}^{\infty}(1 - \lambda(k))$ jest rozbieżny do 0. To z kolei zapewnia, że formuła (4.2) poprawnie definiuje funkcję przeżycia rozkładu F.

Wniosek 4.1. Z Lematu 4.1 i równości (4.4) dostajemy, że w przypadku $N = \infty$ dowolna funkcja $\mu : S \to [0, \infty)$ jest przesuniętą intensywnością awarii pewnego rozkładu (określonego przez (4.6)) wtedy i tylko wtedy, gdy $\mu(k) > 0$ dla wszystkich $k = 0, 1, 2, \dots$ oraz szereg $\sum_{k=0}^{\infty} \frac{\mu(k)}{1+\mu(k)}$ jest rozbieżny. Równoważnie, iloczyn $\prod_{k=0}^{\infty}(1 + \mu(k))^{-1}$ jest rozbieżny do 0, co implikuje, że funkcja przeżycia jest poprawnie zdefiniowana wzorem (4.5).

4.2 Kryteria jednoznaczności charakteryzacji rozkładów dyskretnych przez regresje słabych rekordów

W tym rozdziale rozważamy regresje słabych rekordów postaci $\xi(j) = E\big(h(W_{r+\ell}) \mid W_r = j\big)$, $j \in S$, gdzie h jest funkcją ściśle rosnącą. W odróżnieniu od przypadku ciągłego, funkcja h jest ograniczona z dołu, gdyż zbiór S jest dyskretny. Warunkiem istnienia tej regresji jest $E\big|h(W_{r+\ell})\big| < \infty$. Zauważmy, że warunek ten jest automatycznie spełniony jeśli S jest zbiorem skończonym (dla dowolnej funkcji h) lub jeśli S jest zbiorem nieskończonym, ale funkcja h jest ograniczona z góry. W przeciwnym razie istnienie $E\big|h(W_{r+\ell})\big|$ musi być dodatkowo założone.

Zanim przejdziemy do najważniejszych w tym rozdziale rozważań, przypomnimy jakie informacje o rozkładzie prawdopodobieństwa niesie regresja kolejnych słabych rekordów ($\ell = 1$). Symbolem Δ oznaczać będziemy operator różnicowy określony wzorem $\Delta f(n) = f(n+1) - f(n)$. Załóżmy, że dla pewnego naturalnego $r \geq 1$ zachodzi warunek

$$\xi(j) = E\big(h(W_{r+1}) \mid W_r = j\big), \quad j \in S, \tag{4.8}$$

gdzie h jest funkcją ściśle rosnącą i taką, że $E\left|h(W_{r+1})\right| < \infty$. Wychodząc od wzoru (1.26) dla funkcji regresji (4.8) otrzymujemy równość

$$\xi(j)q_j = \sum_{k=j}^{N} p_k h(k), \quad j \in S. \tag{4.9}$$

Następnie wyznaczając różnice pierwszego rzędu powyższego równania, po prostych przekształceniach, otrzymujemy równanie różnicowe

$$[\xi(j) - h(j)]p_j = \Delta\xi(j)q_{j+1}, \quad j \in \bar{S}. \tag{4.10}$$

Z Lematu 1.5(i) dla $j \in \bar{S}$ mamy $\xi(j) > h(j)$. Stąd dzieląc ostatnie równanie przez $\xi(j) - h(j)$ i stosując definicję funkcji pomocniczej (4.3) dostajemy

$$\mu(j) = \frac{\Delta\xi(j)}{\xi(j) - h(j)}, \quad j \in \bar{S}.$$

Korzystając teraz ze wzoru (4.5) możemy zapisać funkcję $\bar{F}$ w postaci

$$\bar{F}(n) = \prod_{j=0}^{n} \left(1 + \frac{\Delta\xi(j)}{\xi(j) - h(j)}\right)^{-1}, \quad n \in \bar{S}, \tag{4.11}$$

lub równoważnie stosując wzór (4.6) możemy wyrazić funkcję rozkładu prawdopodobieństwa zmiennej losowej X_1 w terminach funkcji ξ i h następująco

$$p_0 = \frac{\Delta\xi(0)}{\xi(1) - h(0)}, \quad p_n = \frac{\Delta\xi(n)}{\xi(n+1) - h(n)} \prod_{j=0}^{n-1} \frac{\xi(j) - h(j)}{\xi(j+1) - h(j)}, \quad n \geq 1, \tag{4.12}$$

przy czym, gdy $N < \infty$, to $p_N = \prod_{j=0}^{N-1} \frac{\xi(j)-h(j)}{\xi(j+1)-h(j)}$.

Zauważmy ponadto, że funkcja regresji (4.8) spełnia następujące warunki:

- w przypadku $N < \infty$ mamy $p_N = q_N$, i wówczas ze wzoru (4.9) otrzymujemy równość (por. Lemat 1.5(iii))

$$\xi(N) = h(N); \tag{4.13}$$

- w przypadku $N = \infty$, iloczyn $\prod_{j=0}^{\infty} \left(1 + \frac{\Delta\xi(j)}{\xi(j)-h(j)}\right)^{-1}$ jest rozbieżny do 0 lub równoważnie

$$\sum_{j=0}^{\infty} \frac{\Delta\xi(j)}{\xi(j+1) - h(j)} = \infty, \tag{4.14}$$

  gdyż $\bar{F}(n) \to 0$ przy $n \to \infty$ (por. Wniosek 4.1). Ponadto z tożsamości (4.9) oraz wzoru (4.11) wynika warunek

$$\lim_{n\to\infty} \xi(n) \prod_{j=0}^{n-1} \left(1 + \frac{\Delta\xi(j)}{\xi(j) - h(j)}\right)^{-1} = \lim_{n\to\infty} \xi(n)q_n = 0. \tag{4.15}$$

W dowodzie głównego twierdzenia w tym rozdziale będziemy korzystać z następującego lematu. Jest to odpowiednik Lematów 2.2 oraz 3.5.

Lemat 4.2. *Załóżmy, że* $S = \{0, 1, \ldots, N\}$*, gdzie* $N \le \infty$*. Ponadto*

(a) $h, \xi : S \to \mathbb{R}$ *są funkcjami ściśle rosnącymi i takimi, że*

- $h(j) < \xi(j)$*,* $j \in \bar{S}$*,*
- *jeżeli* $N < \infty$*, to* $\xi(N) = h(N)$*.*

(b) $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_{\ell-1})$ *jest dowolnym rozwiązaniem układu* $\ell - 1$ *równań różnicowych* ($\tau_\ell = h$)

$$\Delta\tau_i(j) = \frac{\tau_i(j) - \tau_{i+1}(j)}{\xi(j) - \tau_1(j)} \Delta\xi(j), \quad j \in \bar{S}, \quad 1 \le i \le \ell - 1, \tag{4.16}$$

takim, że

$$h(j) < \tau_{\ell-1}(j) < \ldots < \tau_1(j) < \xi(j), \quad j \in \bar{S}, \tag{4.17}$$

przy czym:

- *jeśli* $N < \infty$*, to*

$$\tau_i(N) = h(N) = \xi(N), \quad 1 \le i \le \ell - 1, \tag{4.18}$$

- *jeśli* $N = \infty$*, to*

$$\sum_{j=0}^{\infty} \frac{\Delta\xi(j)}{\xi(j+1) - \tau_1(j)} = \infty \tag{4.19}$$

  oraz

$$\lim_{n\to\infty} \xi(n) \prod_{j=0}^{n-1} \left(1 + \frac{\Delta\xi(j)}{\xi(j) - \tau_1(j)}\right)^{-1} = 0. \tag{4.20}$$

Niech $\eta(j) = \xi(j) + \sum_{i=1}^{\ell-1} \tau_i(j)$*,* $j \in S$*, oraz*

$$\bar{G}(n) = \prod_{j=0}^{n} \left(1 + \frac{\Delta\eta(j)}{\xi(j) - h(j)}\right)^{-1}, \quad n \in \bar{S},$$

a dodatkowo $\bar{G}(N) = 0$*, jeśli* $N < \infty$*. Wtedy* $G = 1 - \bar{G}$ *jest rozkładem prawdopodobieństwa na* S *takim, że dla słabych rekordów* V_r *oraz* $V_{r+\ell}$ *z rozkładu* G *mamy*

$$\xi(j) = E\big(h(V_{r+\ell}) \mid V_r = j\big), \quad j \in S.$$

Dowód. Mnożąc równania (4.16) przez $\xi(j) - \tau_1(j)$, a następnie sumując wszystkie równania stronami otrzymujemy

$$\frac{\Delta\eta(j)}{\xi(j) - h(j)} = \frac{\Delta\xi(j)}{\xi(j) - \tau_1(j)}, \quad j \in \bar{S}. \tag{4.21}$$

Zatem

$$\lambda(j) = 1 - \frac{\bar{G}(j)}{\bar{G}(j-1)} = \frac{\Delta\xi(j)}{\xi(j+1) - \tau_1(j)} < 1, \quad j \in \bar{S}.$$

Jeśli teraz $N < \infty$, to $\bar{G}(N) = 0$, a więc $\lambda(N) = 1$ i G jest rozkładem prawdopodobieństwa na $\{0, 1, \ldots, N\}$. Jeśli natomiast $N = \infty$, to na mocy Lematu 4.1 oraz warunku (4.19) funkcja G określa jednoznacznie właściwy rozkład prawdopodobieństwa na $\{0, 1, \ldots\}$.

Niech $V_1, V_2, \ldots$ oznaczają słabe wartości rekordowe z rozkładu G. Pokażemy teraz, że dla $0 \leq i \leq \ell - 1$ oraz $j \in S$ mamy

$$\tau_i(j) = E\big(\tau_{i+1}(V_{r+i+1}) \mid V_{r+i} = j\big), \tag{4.22}$$

gdzie $\tau_0 = \xi$ oraz $\tau_\ell = h$. Zauważmy najpierw, że dla rozkładu G mamy

$$\mu(j) = \frac{\Delta\xi(j)}{\xi(j) - \tau_1(j)},$$

(por. (4.5), (4.21) oraz definicja $\bar{G}$). Ponadto, z (4.16) mamy również

$$\mu(j) = \frac{\Delta\tau_i(j)}{\tau_i(j) - \tau_{i+1}(j)}, \quad 1 \leq i \leq \ell - 1,$$

a więc na mocy definicji funkcji μ (patrz (4.3))

$$p_j = \frac{\Delta\tau_i(j)}{\tau_i(j) - \tau_{i+1}(j)} q_{j+1}, \quad j \in \bar{S}. \tag{4.23}$$

Aby udowodnić równość (4.22), wystarczy pokazać, że dla $j \in S$ zachodzi

$$\sum_{k=j}^{N} \tau_{i+1}(k) p_k = \tau_i(j) q_j, \quad j \in S. \tag{4.24}$$

Z równania (4.23) otrzymujemy po prostych przekształceniach następujące równanie

$$\tau_{i+1}(k) p_k = \tau_i(k) p_k - \Delta\tau_i(k) q_{k+1} = \tau_i(k) q_k - \tau_i(k+1) q_{k+1},$$

dla $k \in \bar{S}$. Ustalmy $0 \leq j \leq L < N$. Sumując ostatnie równanie po wszystkich k takich, że $j \leq k \leq L$ otrzymujemy

$$\sum_{k=j}^{L} \tau_{i+1}(k) p_k = \tau_i(j) q_j - \tau_i(L+1) q_{L+1}.$$

Jeśli $N < \infty$, to $p_N = q_N$ oraz na mocy (4.18) mamy $\tau_{i+1}(N) = \tau_i(N)$. Zatem (4.24) zachodzi w sposób oczywisty dla $j = N$. Dla $j < N$ przyjmujemy w ostatniej równości $L = N - 1$. To daje

$$\sum_{k=j}^{N-1} \tau_{i+1}(k) p_k = \tau_i(j) q_j - \tau_i(N) q_N.$$

Przenosząc ostatni składnik na lewą stronę otrzymujemy żądaną równość (4.24).

Jeżeli natomiast $N = \infty$, to udowodnimy, że z warunku (4.20) wynika, że

$$\lim_{L \to \infty} \tau_i(L+1) q_{L+1} = 0.$$

Jeśli $\lim_{L\to\infty}\tau_i(L)<\infty$, to powyższa równość jest oczywista, gdyż $q_{L+1}=\bar{G}(L)\to 0$, gdy $L\to\infty$. Załóżmy więc, że $\lim_{L\to\infty}\tau_i(L)=\infty$, a więc $\tau_i(L)>0$ dla dostatecznie dużych L. Wtedy dla $0\le i\le \ell-1$ mamy $\tau_i\le\xi$, a więc

$$0\le \tau_i(L+1)q_{L+1}=\tau_i(L+1)\bar{G}(L)\le\xi(L+1)\bar{G}(L)=\xi(L+1)\prod_{j=0}^{L}\left(1+\frac{\Delta\xi(j)}{\xi(j)-\tau_1(j)}\right)^{-1},$$

co dąży do 0, gdy $L\to\infty$ na mocy (4.20). Udowodniliśmy zatem równość (4.24) dla $i=0,1,\ldots,\ell-1$ we wszystkich przypadkach. To z kolei dowodzi równości (4.22). Wystarczy teraz skorzystać z Lematu 1.1 aby zakończyć dowód lematu. □

Podamy teraz główne wyniki tego podrozdziału. Najpierw zajmiemy się przypadkiem $\ell=2$, a następnie rozważymy ogólną sytuację.

Twierdzenie 4.1. *Dla ustalonego $r\ge 1$, niech W_r, W_{r+2} będą słabymi rekordami z rozkładu o dystrybuancie F określonej na nośniku $S=\{0,1,\ldots,N\}$, gdzie $N\le\infty$. Załóżmy, że $h\colon S\to\mathbb{R}$ jest funkcją ściśle rosnącą i taką, że $E\left|h(W_{r+2})\right|<\infty$. Wówczas funkcja regresji $\xi\colon S\to\mathbb{R}$ określona wzorem*

$$\xi(j)=E\big(h(W_{r+2})\mid W_r=j\big),\quad j\in S, \tag{4.25}$$

jednoznacznie wyznacza dyskretny rozkład F wtedy i tylko wtedy, gdy równanie różnicowe

$$\Delta y(j)=\frac{y(j)-h(j)}{\xi(j)-y(j)}\Delta\xi(j),\quad j\in\bar{S}, \tag{4.26}$$

ma dokładnie jedno rozwiązanie $y=\varphi(j)$ takie, że

$$h(j)<\varphi(j)<\xi(j),\quad j\in\bar{S}, \tag{4.27}$$

przy czym:

(a) *jeśli $N<\infty$, to*

$$\varphi(N)=\xi(N); \tag{4.28}$$

(b) *jeśli $N=\infty$, to*

$$\sum_{j=0}^{\infty}\frac{\Delta\xi(j)}{\xi(j+1)-\varphi(j)}=\infty \tag{4.29}$$

oraz

$$\lim_{n\to\infty}\xi(n)\prod_{j=0}^{n-1}\left(1+\frac{\Delta\xi(j)}{\xi(j)-\varphi(j)}\right)^{-1}=0. \tag{4.30}$$

Wtedy funkcja przeżycia rozkładu F dana jest wzorem

$$\bar{F}(n)=\prod_{j=0}^{n}\left(1+\frac{\Delta\eta(j)}{\xi(j)-h(j)}\right)^{-1},\quad n\in\bar{S}, \tag{4.31}$$

gdzie $\eta(j)=\xi(j)+\varphi(j)$, przy czym, gdy $N<\infty$, to $\bar{F}(N)=0$.

Dowód. Załóżmy, że dana jest regresja (4.25) i określmy

$$\varphi(j) = E\left(h(W_{r+2}) \mid W_{r+1} = j\right), \quad j \in S. \tag{4.32}$$

Wówczas z Lematu 1.1 z $\ell = 2$ otrzymujemy

$$\xi(j) = E\left(\varphi(W_{r+1}) \mid W_r = j\right), \quad j \in S. \tag{4.33}$$

Teraz, skoro funkcja φ, określona wzorem (4.32), jest funkcją regresji kolejnych słabych wartości rekordowych, to na mocy (4.13) spełnia ona warunek (4.28) jeśli $N < \infty$, a na mocy (4.14) i (4.15) spełnione są warunki (4.29) i (4.30) jeśli $N = \infty$. Co więcej, stosując dwukrotnie Lemat 1.5(i), otrzymujemy $\xi(j) > \varphi(j) > h(j)$, dla $j \in \bar{S}$. Następnie wychodząc od zależności (4.10) w połączeniu z warunkami (4.32) oraz (4.33) otrzymujemy układ dwóch równań różnicowych

$$\begin{cases} \Delta\varphi(j) = [\varphi(j) - h(j)]\mu(j), \\ \Delta\xi(j) = [\xi(j) - \varphi(j)]\mu(j), \end{cases} \quad j \in \bar{S}. \tag{4.34}$$

Wyznaczając μ z drugiego równania, a następnie wstawiając do pierwszego, widzimy, że φ dane wzorem (4.32) spełnia również równanie różnicowe (4.26).

Zatem, jeśli ξ jest regresją (4.25), to równanie (4.26) ma co najmniej jedno rozwiązanie spełniające warunki podane w wypowiedzi twierdzenia. Ponadto sumując stronami równania (4.34) otrzymujemy łatwo

$$\mu(j) = \frac{\Delta\eta(j)}{\xi(j) - h(j)}, \quad j \in \bar{S}, \tag{4.35}$$

gdzie $\eta(j) = \xi(j) + \varphi(j)$. Zatem stosując wzór (4.5) otrzymujemy funkcję $\bar{F}$ wyrażoną wzorem (4.31).

Z drugiej strony, na mocy Lematu 4.2 z $\ell = 2$, każde rozwiązanie równania (4.26) spełniające nierówność (4.27) oraz warunki (4.28) lub (4.29) i (4.30) określa pewien rozkład prawdopodobieństwa, dla którego zachodzi równość (4.25). Zatem liczba charakteryzowanych przez tę równość rozkładów pokrywa się z liczbą rozwiązań równania różnicowego (4.26) spełniających (4.27) oraz (4.28) lub (4.29) i (4.30). Jeśli istnieje dokładnie jedno rozwiązanie, to musi ono pokrywać się z funkcją φ daną wzorem (4.32) i rozkład F jest wówczas wyznaczony jednoznacznie. W przeciwnym razie istnieją przynajmniej dwa różne rozkłady, dla których zachodzi (4.25). □

Uwaga 4.2. Z równości (4.6) możemy określić rozkład prawdopodobieństwa przez regresję ξ następująco

$$p_0 = \frac{\Delta\eta(0)}{\Delta\eta(0) + \xi(0) - h(0)}, \quad p_n = \frac{\Delta\eta(n)}{\Delta\eta(n) + \xi(n) - h(n)} \prod_{j=0}^{n-1} \frac{\xi(j) - h(j)}{\Delta\eta(j) + \xi(j) - h(j)}, \quad n \geq 1, \tag{4.36}$$

przy czym, gdy $N < \infty$ to $p_N = \prod_{j=0}^{N-1} \frac{\xi(j)-h(j)}{\Delta\eta(j)+\xi(j)-h(j)}$.

Uwaga 4.3. Zauważmy, że z układu równań (4.34) możemy wyznaczyć funkcję μ na kilka równoważnych sposobów. Mianowicie funkcja μ może być przedstawiona albo w terminach funkcji h i φ, albo za pomocą funkcji ξ i φ, albo jak w równaniu (4.35) za pomocą wszystkich wspomnianych wyżej funkcji. Łącząc ten wynik z (4.5) możemy również wyrazić $\bar{F}$ za pomocą tylko h i φ albo za pomocą ξ i φ.

Zaznaczmy, że podobny problem charakteryzacyjny dla regresji słabych rekordów o odstępie dwa był przedmiotem rozważań w pracy [3]. Twierdzenie 1 w [3] orzeka, że regresja słabych rekordów z $\ell = 2$ określa bazowy rozkład o nieskończonym nośniku jednoznacznie. Jednakże, nasze Twierdzenie 4.1 jest również prawdziwe dla rozkładów określonych na skończonych nośnikach. Ponadto w powyższym twierdzeniu podajemy wzór określający funkcje rozkładu, czego nie było w dotychczasowych wynikach. Zwróćmy również uwagę, że przeciwnie do metody przedstawionej przez Alieva [3], nasze podejście daje się łatwo uogólnić na regresję słabych rekordów o większym odstępie $\ell > 2$.

Przejdziemy teraz do rozszerzenia Twierdzenia 4.1 na przypadek regresji słabych rekordów o dowolnym odstępie większym niż 2.

Twierdzenie 4.2. *Dla ustalonych $r \geq 1$ i $\ell \geq 2$, niech $W_r, W_{r+\ell}$ będą słabymi rekordami z rozkładu o dystrybuancie F określonej na nośniku $S = \{0, 1, \dots, N\}$, gdzie $N \leq \infty$. Załóżmy, że $h \colon S \to \mathbb{R}$ jest funkcją ściśle rosnącą i taką, że $E\big|h(W_{r+\ell})\big| < \infty$. Wówczas regresja* (1.13) *określa rozkład zmiennej losowej X_1 jednoznacznie wtedy i tylko wtedy, gdy układ $\ell - 1$ równań różnicowych*

$$\Delta y_i(j) = \frac{y_i(j) - y_{i+1}(j)}{\xi(j) - y_1(j)} \Delta\xi(j), \quad 1 \leq i \leq \ell - 1, \quad j \in \bar{S}, \tag{4.37}$$

gdzie $y_\ell = h$ ma dokładnie jedno rozwiązanie $\boldsymbol{\varphi} = (\varphi_1, \dots, \varphi_{\ell-1})$, które spełnia warunki (4.17) *oraz* (4.18) *lub* (4.19) *i* (4.20) *z $\boldsymbol{\tau} = (\tau_1, \dots, \tau_{\ell-1})$ zastąpionym przez $\boldsymbol{\varphi}$.*

Wówczas funkcja przeżycia rozkładu F jest określona wzorem

$$\bar{F}(n) = \prod_{j=0}^{n} \left(1 + \frac{\Delta\eta(j)}{\xi(j) - h(j)}\right)^{-1}, \quad n \in \bar{S}, \tag{4.38}$$

gdzie funkcja $\eta \colon S \to \mathbb{R}$ jest następująca

$$\eta(j) = \xi(j) + \sum_{i=1}^{\ell-1} \varphi_i(j). \tag{4.39}$$

Dowód. Załóżmy, że dana jest regresja (1.13) i oznaczmy $\varphi_\ell = h$. Następnie rozważmy funkcje φ_i, $0 \leq i \leq \ell - 1$, określone wzorami (1.15). Wówczas z Lematu 1.1 otrzymujemy $\varphi_0 = \xi$. Ponadto, ponieważ każda funkcja φ_i jest funkcją regresji kolejnych słabych rekordów, mianowicie regresją $\varphi_{i+1}(W_{r+i+1})$ względem W_{r+i}, to z Lematu 1.5(i) otrzymujemy $\varphi_{i+1} < \varphi_i$ na $\bar{S}$ dla $0 \leq i \leq \ell - 1$. Zatem

$$h(j) < \varphi_{\ell-1}(j) < \ldots < \varphi_1(j) < \xi(j), \quad j \in \bar{S}.$$

Co więcej, jeśli $N < \infty$, to na mocy (4.13) funkcje te spełniają warunek (4.18) z τ zastąpionym przez φ. Jeśli natomiast $N = \infty$, to na mocy (4.14) i (4.15) spełnione są warunki (4.19) i (4.20) z τ_1 zastąpionym przez φ_1.

Dalej, funkcje te spełniają ℓ równań różnicowych

$$[\varphi_i(j) - \varphi_{i+1}(j)]p_j = \Delta\varphi_i(j)q_{j+1}, \quad 0 \le i \le \ell - 1, \quad j \in \bar{S}. \tag{4.40}$$

Zauważmy, że (4.40) jest układem ℓ równań ℓ z nieznanymi funkcjami $\varphi_1, \dots, \varphi_{\ell-1}$ oraz μ (przypomnijmy, że $\varphi_0 = \xi$ oraz $\varphi_\ell = h$ są znane). Aby wyeliminować μ wyznaczamy tę funkcję z każdego równania. Zatem

$$\mu(j) = \frac{\Delta\varphi_i(j)}{\varphi_i(j) - \varphi_{i+1}(j)}, \quad 1 \le i \le \ell - 1 \tag{4.41}$$

oraz

$$\mu(j) = \frac{\Delta\xi(j)}{\xi(j) - \varphi_1(j)}.$$

Porównując teraz prawą stronę ostatniego równania z prawą stroną równań (4.41) widzimy, że funkcje $\varphi_1, \dots, \varphi_{\ell-1}$ spełniają układ $\ell - 1$ równań różnicowych

$$\Delta\varphi_i(j) = \frac{\varphi_i(j) - \varphi_{i+1}(j)}{\xi(j) - \varphi_1(j)}\Delta\xi(j), \quad 1 \le i \le \ell - 1.$$

Zatem, jeśli ξ jest dana przez (1.13), to układ (4.37) ma co najmniej jedno rozwiązanie spełniające warunki podane w wypowiedzi twierdzenia. Ponadto, sumując stronami wszystkie równania w (4.40) otrzymujemy $[\xi(j) - h(j)]p_j = \Delta\eta(j)q_{j+1}$, a zatem μ można przedstawić w postaci $\mu(j) = \frac{\Delta\eta(j)}{\xi(j)-h(j)}$. Ponownie jak w Twierdzeniu 4.1 funkcję $\bar{F}$ można przedstawić w postaci (4.38), a odpowiadająca funkcja rozkładu prawdopodobieństwa jest dana wzorem (4.36) z tą różnicą, że funkcja η jest określona teraz przez (4.39).

Z drugiej strony na mocy Lematu 4.2, każde rozwiązanie układu (4.37) spełniające nierówność (4.17) oraz (4.18) lub (4.19) i (4.20) określa pewien rozkład prawdopodobieństwa, dla którego zachodzi równość regresji (1.13). Jeśli zatem istnieje dokładnie jedno takie rozwiązanie, to musi ono pokrywać się z $\boldsymbol{\varphi} = (\varphi_1, \dots, \varphi_{\ell-1})$ określonym przez (1.15). Zatem F jest wyznaczona jednoznacznie. W przeciwnym razie jednoznaczność nie zachodzi. □

Uwaga 4.4. Oczywiście dla $\ell = 2$ układ (4.37) redukuje się do jednego równania różnicowego (4.26). Zatem Twierdzenie 4.1 jest szczególnym przypadkiem Twierdzenia 4.2.

4.3 Jednoznaczność charakteryzacji – szczególne przypadki dla $\ell = 2$

Kryterium jednoznaczności charakteryzacji rozkładów dyskretnych przez regresje słabych wartości rekordowych (1.13) zaprezentowane w podrozdziale 4.2 podobnie jak kryteria jednoznaczności charakteryzacji rozkładów ciągłych przez regresje uogólnionych statystyk porządkowych jest z jednej strony łatwe do wyprowadzenia, ale z drugiej strony trudne w implementacji.

W ogólności, problem jednoznacznej charakteryzacji dyskretnych rozkładów prawdopodobieństwa przez regresje niekolejnych słabych wartości rekordowych sprowadzony został do rozwiązania odpowiedniego równania (gdy $\ell = 2$) lub układu równań różnicowych (gdy $\ell > 2$) z nieklasycznymi warunkami ograniczającymi.

Znalezienie rozwiązania takiego problemu w ogólnym przypadku wydaje się trudne do uzyskania. W tym podrozdziale rozpatrzymy ten problem dla regresji słabych rekordów o odstępie $\ell = 2$ i wykażemy jednoznaczność rozwiązania szczególnego problemu różnicowego, która pociąga za sobą jednoznaczność charakteryzacji rozkładów prawdopodobieństwa przez regresje wartości rekordowych, dla następujących szczególnych przypadków:

- h jest funkcją ograniczoną, a więc $h(N) < \infty$;
- h jest funkcją nieograniczoną ($h(N) = \infty$) oraz ξ jest funkcją asymptotycznie liniową względem funkcji h, mianowicie
$$\lim_{j\to\infty}[\xi(j) - (ah(j) + b)] = 0, \tag{4.42}$$
dla pewnych $a \geq 1, b \in \mathbb{R}$.

Uwaga 4.5. Jeżeli $N < \infty$, to oczywiście mamy $h(N) < \infty$. Jeżeli $N = \infty$, to $h(N)$ jest zdefiniowana jako granica $h(N) = \lim_{j\to\infty} h(j)$, która w ogólności może być zarówno skończona, jak i nieskończona.

4.3.1 Własności rozwiązań pomocniczego problemu różnicowego

Niech będzie dany problem różnicowy następującej postaci
$$\begin{cases} \Delta y(j) = \dfrac{y(j) - h(j)}{\xi(j) - y(j)} \Delta\xi(j), \\ h(j) < y(j) < \xi(j), \end{cases} \qquad j \in \bar{S}, \tag{4.43}$$
gdzie $h, \xi : S \to \mathbb{R}$ są ściśle rosnącymi funkcjami.

Zauważmy, że ze względu na warunek ograniczający $h < y < \xi$, dowolne rozwiązanie y powyższego problemu jest funkcją ściśle rosnącą. Inne własności rozwiązań problemu (4.43) podają kolejne dwa lematy.

Lemat 4.3. *Niech $S = \{0, 1, \ldots, N\}$, gdzie $N \leq \infty$. Załóżmy, że $h, \xi : S \to \mathbb{R}$ są ściśle rosnącymi funkcjami takimi, że $h(j) < \xi(j)$ dla $j \in \bar{S}$. Przypuśćmy, że $\varphi, y\colon S \to \mathbb{R}$ są dwoma różnymi rozwiązaniami problemu* (4.43)*. Wówczas*
$$y(j+1) - \varphi(j+1) = [y(j) - \varphi(j)] \left(1 + \frac{[\xi(j) - h(j)]\Delta\xi(j)}{[\xi(j) - \varphi(j)][\xi(j) - y(j)]}\right), \quad j \in \bar{S}. \tag{4.44}$$

Dowód. Niech $z = y - \varphi$. Wykonując proste przekształcenia otrzymujemy

$$\Delta z(j) = \frac{[y(j)-\varphi(j)][\xi(j)-h(j)]}{[\xi(j)-y(j)][\xi(j)-\varphi(j)]}\Delta\xi(j), \tag{4.45}$$

co prowadzi do tezy lematu. □

Lemat 4.4. *Przy założeniach Lematu 4.3 mamy*

(i) $y(j) \neq \varphi(j)$ dla $j \in \bar{S}$;

(ii) albo $y < \varphi$, albo $y > \varphi$ na $\bar{S}$;

(iii) jeżeli $z = (y-\varphi)^2$, to z jest funkcją ściśle rosnącą na $\bar{S}$;

(iv) jeżeli $h < y < \varphi$ oraz $w = \frac{\varphi - y}{\varphi - h}$, to $w(0) \leq w < 1$ oraz w jest funkcją ściśle rosnącą na $\bar{S}$;

(v) jeżeli $\varphi < y < \xi$ oraz $w = \frac{y-\varphi}{\xi-\varphi}$, to $w(0) \leq w < 1$ oraz w jest funkcją ściśle rosnącą na $\bar{S}$.

Dowód. Zauważmy, że wyrażenie w nawiasach po prawej strony równania (4.44) jest zawsze dodatnie. Zatem, jeżeli $y(j) = \varphi(j)$ dla pewnego $j \in \bar{S}$, to $y = \varphi$ na S, co kończy dowód punktu (i). Co więcej, z równości (4.44) widzimy, że różnica $y(j+1) - \varphi(j+1)$ jest dodatnia wtedy i tylko wtedy, gdy wyrażenie $y(j) - \varphi(j)$ jest dodatnie, co dowodzi (ii). Jeżeli $\varphi < y < \xi$, to z równości (4.45), różnica $y - \varphi$ jest rosnąca oraz jeżeli $h < y < \varphi$, to różnica $\varphi - y$ jest rosnąca, do dowodzi (iii). Obliczając różnice pierwszego rzędu funkcji $w = \frac{\varphi-y}{\varphi-h}$ otrzymujemy

$$\begin{aligned}\Delta w(j) &= \frac{\Delta\varphi(j)-\Delta y(j)}{\varphi(j+1)-h(j+1)} + \frac{\varphi(j)-y(j)}{\varphi(j+1)-h(j+1)} - \frac{\varphi(j)-y(j)}{\varphi(j)-h(j)}\\ &= \frac{\varphi(j)-y(j)}{[\varphi(j+1)-h(j+1)][\varphi(j)-h(j)]}\left[\frac{y(j)-h(j)}{\xi(j)-y(j)}\Delta\varphi(j)+\Delta h(j)\right] > 0,\end{aligned}$$

co dowodzi (iv). Podobnie dla $w = \frac{y-\varphi}{\xi-\varphi}$, gdy $y > \varphi$ otrzymujemy

$$\begin{aligned}\Delta w(j) &= \frac{\Delta y(j)-\Delta\varphi(j)}{\xi(j+1)-\varphi(j+1)} + \frac{y(j)-\varphi(j)}{\xi(j+1)-\varphi(j+1)} - \frac{y(j)-\varphi(j)}{\xi(j)-\varphi(j)}\\ &= \frac{y(j)-\varphi(j)}{[\xi(j+1)-\varphi(j+1)][\xi(j)-\varphi(j)]}[\Delta y(j)+\Delta\varphi(j)] > 0,\end{aligned}$$

co kończy dowód punktu (v). □

4.3.2 Jednoznaczność charakteryzacji dla $h(N) < \infty$

Następne twierdzenie orzeka o jednoznaczności charakteryzacji dyskretnych rozkładów prawdopodobieństwa przez regresje słabych wartości rekordowych (1.13) w przypadku gdy $\ell = 2$ oraz h jest funkcją ograniczoną, czyli $h(N) < \infty$. Zauważmy, że funkcja h jest zawsze ograniczona z dołu, a więc $h(N) < \infty$ implikuje, że $E\big|h(W_{r+2})\big| < \infty$.

Twierdzenie 4.3. *Dla ustalonego $r \geq 1$, niech W_r, W_{r+2} będą słabymi rekordami z rozkładu o dystrybuancie F określonej na nośniku $S = \{0, 1, \ldots, N\}$, gdzie $N \leq \infty$. Załóżmy, że $h\colon S \to \mathbb{R}$ jest funkcją ściśle rosnącą. Jeżeli $h(N) < \infty$ (w szczególności, gdy $N < \infty$), to regresja*

$$\xi(j) = E\left(h(W_{r+2}) \mid W_r = j\right), \quad j \in S,$$

jednoznacznie charakteryzuje rozkład zmiennej losowej X_1 określony dystrybuantą F, którą można wyznaczyć ze wzoru (4.31).

Dowód powyższego twierdzenia jest podobny do dowodu Twierdzenia 2.3. Jednakże dla kompletności rozważań przedstawiamy go w całości.

Dowód. Z Lematu 1.6(iii) w przypadku $h(N) < \infty$ mamy $\xi(N) = h(N)$. Wiemy, że funkcja φ określona przez (4.32) jest rozwiązaniem równania różnicowego (4.26) spełniającym nierówność (4.27) oraz warunki (4.28) lub (4.29) i (4.30). Przypuśćmy, że istnieje również inne rozwiązanie tego problemu y. Naturalnie φ i y są także rozwiązaniami problemu (4.43). W przypadku, gdy $N < \infty$ korzystając z warunku (4.28) oraz z równości (4.44) otrzymujemy kolejno $\varphi(N) = y(N) = \xi(N)$, $\varphi(N-1) = y(N-1), \ldots, \varphi(0) = y(0)$, czyli $\varphi = y$. W drugim przypadku, gdy $N = \infty$ rozważając funkcję $z = (y - \varphi)^2$ mamy $0 \leq z(j) \leq (\xi(j) - h(j))^2$, a zatem $z(j) \to 0$, gdy $j \to \infty$. Ale $z(0) > 0$ oraz na podstawie Lematu 4.4(iii), z jest funkcją ściśle rosnącą. To daje sprzeczność, czyli również $\varphi = y$. Pokazaliśmy zatem, że funkcja φ określona wzorem (4.32) jest jedynym rozwiązaniem problemu różnicowego (4.43). Teza twierdzenia wynika teraz z Twierdzenia 4.1. □

4.3.3 Jednoznaczność charakteryzacji dla $h(N) = \infty$

Dowód analogicznego wyniku dla przypadku, gdy h jest funkcją nieograniczoną, tj. $N = \infty$ i $h(N) = \infty$ oraz dowolnej funkcji regresji ξ, nie wydaje się łatwy. W takim przypadku nie możemy twierdzić, że $\xi(j) - h(j)$ zbiega do 0, gdy $j \to \infty$, a to jest kluczowe w dowodzie Twierdzenia 4.3. Rzecz jasna możliwe jest nawet, że funkcja $\xi - h$ ma granicę niewłaściwą ∞. Poniżej udowodnimy jednoznaczność charakteryzacji dla szczególnej postaci regresji ξ, mianowicie gdy jest ona funkcją asymptotycznie liniową względem funkcji h, tj. spełniony jest warunek (4.42). Oczywiście w takim przypadku z uwagi na założenie $\xi > h$ mamy $a \geq 1$ oraz jeżeli $a = 1$, to $b \geq 0$. Zauważmy, że w przypadku gdy $a = 1$ i $b = 0$ mamy przypadek analogiczny jak w dowodzie Twierdzenia 4.3, który implikuje żądaną jednoznaczność charakteryzacji (por. (2.42)). Zatem dla $a = 1$ możemy założyć, że $b > 0$.

Lemat 4.5. *Niech $S = \{0, 1, 2, \ldots\}$ oraz niech $h, \xi\colon S \to \mathbb{R}$ będą ściśle rosnącymi funkcjami takimi, że $h < \xi$. Załóżmy, że $h(\infty) = \infty$ oraz ξ spełnia warunek* (4.42) *dla pewnych $a \geq 1$ i $b \in \mathbb{R}$. Załóżmy, że problem różnicowy* (4.43) *ma rozwiązanie φ takie, że*

$$\lim_{j \to \infty} \left[\varphi(j) - (ch(j) + d)\right] = 0 \tag{4.46}$$

dla pewnych $c \geq 1$ i $d \in \mathbb{R}$.

(a) Jeżeli $a = 1$, to $c = 1$, $d = b/2 > 0$ oraz φ jest jedynym rozwiązaniem problemu (4.43).

(b) Jeżeli $a > 1$, to $c = \sqrt{a} \in (1, a)$ oraz φ jest jedynym rozwiązaniem problemu (4.43).

Idea dowodu powyższego lematu jest podobna do dowodu Twierdzenia 3.3, jednakże dla kompletności rozważań przedstawiamy dowód w całości.

Dowód. (a) Z warunku $h < \varphi < \xi$, dla dostatecznie dużych $j \in S$, dostajemy

$$1 < \frac{\varphi(j)}{h(j)} < \frac{\xi(j)}{h(j)}.$$

Zatem oczywiście jeżeli $a = 1$, to również $c = 1$. Zauważmy również, że z warunków (4.42) oraz (4.46) wynika, że $\lim_{j\to\infty} \frac{\varphi(j)}{\xi(j)} = 1$ oraz $\lim_{j\to\infty} [\xi(j) - \varphi(j)] = b - d > 0$.

Kładąc

$$\Phi(j) = \frac{\Delta\varphi(j)}{\Delta\xi(j)} = \frac{\varphi(j) - h(j)}{\xi(j) - \varphi(j)} \tag{4.47}$$

otrzymujemy $\Phi(j) \to \frac{d}{b-d}$, gdy $j \to \infty$. Stosując teraz twierdzenie Stolza otrzymujemy, że $\frac{\varphi(j)}{\xi(j)} \to \frac{d}{b-d}$, gdy $j \to \infty$. Z drugiej strony wiemy, że granica ta jest równa 1, zatem

$$1 = \frac{d}{b-d}.$$

Stąd $d = b/2 > 0$.

Załóżmy teraz, że y jest innym rozwiązaniem problemu (4.43) spełniającym warunek $h < y < \xi$. Wówczas dla dostatecznie dużych j mamy

$$1 < \frac{y(j)}{h(j)} < \frac{\xi(j)}{h(j)},$$

a skoro $\lim_{j\to\infty} \frac{\xi(j)}{h(j)} = 1$, to również $\lim_{j\to\infty} \frac{y(j)}{h(j)} = 1$ oraz $\lim_{j\to\infty} \frac{y(j)}{\xi(j)} = 1$.

Na podstawie Lematu 4.4(ii) wiemy, że $\varphi < y$ albo $y < \varphi$. Jeżeli $\varphi < y < \xi$, to rozważmy funkcję $z = y - \varphi$. Z Lematu 4.4(iii) funkcja z rośnie od $z(0) > 0$ do stałej $C \in (z(0), b/2]$. Załóżmy najpierw, że $C < b/2$. Definiując

$$\Psi(j) = \frac{\Delta y(j)}{\Delta\xi(j)} = \frac{z(j) + \varphi(j) - h(j)}{\xi(j) - \varphi(j) - z(j)}, \tag{4.48}$$

a następnie przechodząc do granicy, gdy $j \to \infty$ dostajemy $\Psi(j) \to \frac{C+b/2}{b/2-C}$. Ponadto z klasycznego twierdzenia Stolza otrzymujemy równość

$$\lim_{j\to\infty} \frac{y(j)}{\xi(j)} = \lim_{j\to\infty} \frac{\Delta y(j)}{\Delta\xi(j)} = \frac{C + b/2}{b/2 - C}.$$

Z drugiej strony wiemy już, że granica po lewej stronie wynosi 1. Zatem

$$1 = \frac{C + b/2}{b/2 - C},$$

a więc $C = 0$. Sprzeczność. Gdyby natomiast $C = b/2$, to $\Psi(j) \to \infty$, gdy $j \to \infty$. Na mocy twierdzenia Stolza otrzymamy wówczas, że $\frac{y(j)}{\xi(j)} \to \infty$, co przeczy temu, że $y < \xi$ na S.

Analogicznie dla przypadku, gdy $h < y < \varphi$, wystarczy rozważyć funkcję $z = \varphi - y$, aby dojść do kolejnej sprzeczności. Zatem φ jest jedynym rozwiązaniem problemu (4.43).

(b) Dowód lematu dla $a > 1$ jest oparty na tym samym pomyśle co dowód dla punktu (a). Zauważmy, że $\lim_{j\to\infty} \frac{\xi(j)}{h(j)} = a$ oraz $\lim_{j\to\infty} \frac{\varphi(j)}{h(j)} = c$, co implikuje

$$\lim_{j\to\infty} \frac{\varphi(j)}{\xi(j)} = \frac{c}{a}. \tag{4.49}$$

Oczywiście $1 \leq c \leq a$. Przypuśćmy najpierw, że $c \in (1,a)$. Przedstawiając funkcję Φ określoną przez (4.47) następująco

$$\Phi(j) = \frac{\frac{\varphi(j)}{h(j)} - 1}{\frac{\xi(j)}{h(j)} - \frac{\varphi(j)}{h(j)}},$$

a następnie przechodząc do granicy, gdy $j \to \infty$ otrzymujemy $\Phi(j) \to \frac{c-1}{a-c}$. Łącząc ten wynik z warunkiem (4.49) dostajemy równanie kwadratowe $c^2 = a$, które w przedziale $(1,a)$ ma dokładnie jedno rozwiązanie $c = \sqrt{a}$. Gdyby $c = 1$, to $\lim_{j\to\infty} \Phi(j) = 0$, a gdyby $c = a$, to $\lim_{j\to\infty} \Phi(j) = \infty$. W obydwu tych przypadkach, z twierdzenia Stolza otrzymujemy sprzeczność z warunkiem (4.49).

Załóżmy teraz, że y jest innym rozwiązaniem problemu (4.43). Jeżeli granica $L(y) = \lim_{j\to\infty} \frac{y(j)}{h(j)}$ istnieje, to powtarzając rozumowanie z punktu (a) dostajemy, że $L(y) \in (1,a)$ oraz $\frac{L(y)}{a} = \frac{L(y)-1}{a-L(y)}$. Ale to równanie określa jednoznacznie c, zatem $L(y) = c$ dla dowolnego y będącego rozwiązaniem problemu (4.43).

Jeżeli $\varphi < y < \xi$, to rozważmy funkcję $w = \frac{y-\varphi}{\xi-\varphi}$. Z Lematu 4.4(v) funkcja w rośnie od $w(0) > 0$ do $C \in (w(0),1]$. Zauważmy, że funkcja Ψ określona przez (4.48) może zostać przedstawiona w postaci

$$\Psi = \frac{\frac{y-\varphi}{\xi-\varphi} + \Phi}{1 - \frac{y-\varphi}{\xi-\varphi}} = \frac{w + \Phi}{1 - w}.$$

Załóżmy najpierw, że $C < 1$. Wtedy otrzymujemy, że granica funkcji Ψ w nieskończoności istnieje i jest równa $\frac{C+\frac{c}{a}}{1-C}$. Na mocy twierdzenia Stolza otrzymujemy istnienie również granicy $\lim_{j\to\infty} \frac{y(j)}{\xi(j)} = \frac{C+\frac{c}{a}}{1-C}$. Z przedstawienia $\frac{y}{\xi} = \frac{y}{h} \cdot \frac{h}{\xi}$ otrzymujemy istnienie granicy $\lim_{j\to\infty} \frac{y(j)}{h(j)}$, co prowadzi do równości

$$\frac{c}{a} = \frac{C + \frac{c}{a}}{1 - C}.$$

Stąd dostajemy $C = 0$, co daje sprzeczność. Gdyby natomiast $C = 1$, to $\Psi(j) \to \infty$, a z twierdzenia Stolza otrzymamy również $\frac{y(j)}{\xi(j)} \to \infty$, gdy $j \to \infty$. Co z kolei przeczy temu, że $y < \xi$ na S.

Jeżeli $h < y < \varphi$, to wystarczy przeprowadzić podobne rozumowanie z wykorzystaniem funkcji $w = \frac{\varphi - y}{\varphi - h}$, aby uzyskać kolejną sprzeczność. Zatem φ jest jedynym rozwiązaniem problemu (4.43). $\square$

Stosując teraz Lemat 4.5 w połączeniu z Twierdzeniem 4.1 otrzymujemy jednoznaczną charakteryzację dyskretnych rozkładów prawdopodobieństwa przez asymptotycznie liniowe regresje słabych rekordów o odstępie dwa.

Twierdzenie 4.4. *Dla ustalonego $r \geq 1$, niech W_r, W_{r+2} będą słabymi rekordami z rozkładu o dystrybuancie F określonej na nośniku $S = \{0, 1, \dots\}$. Załóżmy, że $h \colon S \to \mathbb{R}$ jest funkcją ściśle rosnącą i taką, że $E\left|h(W_{r+2})\right| < \infty$ oraz $h(\infty) = \infty$. Jeżeli regresja $\xi : S \to \mathbb{R}$ określona wzorem* (4.25) *spełnia warunek* (4.42) *dla pewnych $a \geq 1$, $b \in \mathbb{R}$ oraz problem różnicowy* (4.43) *ma rozwiązanie φ, które spełnia warunek* (4.46)*, to funkcja regresji ξ jednoznacznie wyznacza rozkład prawdopodobieństwa, którego funkcja przeżycia $\bar{F}$ dana jest wzorem* (4.31).

Wiemy już, że warunek (4.42) z $a = 1$ oraz $b = 0$ implikuje jednoznaczność charakteryzacji rozkładu F przez regresję (4.25). Wykażemy teraz, że dla dużej klasy dyskretnych rozkładów prawdopodobieństwa można zweryfikować ten warunek jedynie na podstawie znajomości funkcji intensywności awarii tych rozkładów, bez konieczności wyznaczania funkcji regresji ξ. Analogiczny wynik otrzymaliśmy w Lemacie 2.4 dla rozkładów absolutnie ciągłych.

Lemat 4.6. *Jeżeli $h(j) = j$ oraz $\lim\limits_{j\to\infty} \mu(j) = \infty$, to spełniony jest warunek*

$$\lim_{j\to\infty} [\xi(j) - h(j)] = 0.$$

Dowód. Wychodząc od równości (1.26) warunki (4.32) oraz (4.33) możemy odpowiednio zapisać w postaci

$$\varphi(j) = E\left(h(W_{r+2}) \mid W_{r+1} = j\right) = \frac{1}{q_j} \sum_{k=j}^{\infty} p_k h(k) \tag{4.50}$$

oraz

$$\xi(j) = E\left(\varphi(W_{r+1}) \mid W_r = j\right) = \frac{1}{q_j} \sum_{k=j}^{\infty} p_k \varphi(k). \tag{4.51}$$

Zauważmy, że równanie (4.50) możemy przedstawić w postaci (w przypadku gdy $h(j) = j$)

$$\varphi(j) = h(j) + \frac{1}{q_j} \sum_{k=j+1}^{\infty} (k-j) p_k = h(j) + \frac{1}{q_j} \sum_{k=j+1}^{\infty} q_k.$$

Następnie, korzystając z twierdzenia Stolza dla wyrażenia typu $\frac{0}{0}$ (ciąg $\{q_j\}$ jest ściśle malejący, a $\lim\limits_{j\to\infty} \sum_{k=j+1}^{\infty} q_k = \lim\limits_{j\to\infty} q_j = 0$) obliczamy granicę

$$\lim_{j\to\infty} [\varphi(j) - h(j)] = \lim_{j\to\infty} \frac{1}{\mu(j)} = 0.$$

Podobnie, równanie (4.51) można przekształcić do postaci

$$\begin{aligned}\xi(j) &= h(j) + \frac{1}{q_j}\sum_{k=j+1}^{\infty}(k-j)p_k + \frac{1}{q_j}\sum_{k=j}^{\infty}\frac{p_k}{q_k}\left(\sum_{l=k+1}^{\infty}(l-k)p_l\right)\\ &= h(j) + \frac{1}{q_j}\sum_{k=j+1}^{\infty}q_k + \frac{1}{q_j}\sum_{k=j}^{\infty}\frac{p_k}{q_k}\left(\sum_{l=k+1}^{\infty}q_l\right)\end{aligned}$$

i ponownie korzystając z twierdzenia Stolza dla wyrażenia typu $\frac{0}{0}$ dostajemy

$$\lim_{j\to\infty}[\xi(j)-h(j)] = 2\lim_{j\to\infty}\frac{1}{\mu(j)} = 0.$$

To kończy dowód. $\square$

Wniosek 4.2. Jeżeli $\lim\limits_{j\to\infty}\mu(j)=\infty$, to rozkład F jest jednoznacznie określony przez regresję

$$\xi(j) = E\left(W_{r+2} \mid W_r = j\right). \tag{4.52}$$

Uwaga 4.6. Niestety, podejście przyjęte w tym podrozdziale wydaje się trudne do przeniesienia na ogólny przypadek $\ell \geq 3$. Dlatego też, problem jednoznacznej charakteryzacji rozkładu w takim przypadku jest nadal otwarty.

4.4 Przykłady nowych charakteryzacji rozkładów dyskretnych

W tym podrozdziale zaprezentujemy nowe przykłady charakteryzacji dyskretnych rozkładów prawdopodobieństwa przez funkcję regresji słabych rekordów $\xi(j)=E\big(W_{r+2}\mid W_r=j\big)$, $j\in S$, dla dowolnie ustalonego $r\geq 1$. Zaznaczmy, że we wszystkich prezentowanych w tym podrozdziale przykładach funkcja regresji ξ jest nieliniowa. Przypadek liniowej regresji zostanie przedstawiony w kolejnym podrozdziale.

We wszystkich prezentowanych poniżej przykładach, dla badanych rozkładów prawdopodobieństwa, zostały najpierw wyznaczone funkcje regresji ξ, a następnie zostały zweryfikowane warunki jednoznacznej charakteryzacji tych rozkładów przez wyznaczoną funkcję regresji słabych rekordów.

Pierwszy przykład dotyczy charakteryzacji dyskretnych rozkładów prawdopodobieństwa o skończonym nośniku.

Przykład 4.1. (a) Niech $S=\{0,1,2\}$ oraz $h(j)=j$, $j\in S$. Ponadto przypuśćmy, że

$$\xi(0)=\frac{5}{4},\quad \xi(1)=\frac{7}{4},\quad \xi(2)=2. \tag{4.53}$$

Chcemy znaleźć wszystkie rozkłady prawdopodobieństwa p_0, p_1, p_2 o nośniku S, dla których prawdziwa jest następującą relacja

$$E\left(W_{r+2} \mid W_r\right) = \xi\left(W_r\right)$$

dla dowolnie ustalonego $r \geq 1$. Mając na uwadze wyniki prezentowane w podrozdziałach 4.2 i 4.3 nasze zadanie sprowadza się do wyznaczenia funkcji $\varphi\colon S \to \mathbb{R}$, która jest rozwiązaniem problemu różnicowego (4.43), takim, że $\varphi(2) = 2$. Zatem wystarczy znaleźć $a = \varphi(0)$ oraz $b = \varphi(1)$ takie, że $a < b$, $a \in \left(0, \frac{5}{4}\right)$ oraz $b \in \left(1, \frac{7}{4}\right)$. Z równania różnicowego (4.43) dostajemy układ dwóch równań

$$\begin{cases} b - a = \frac{2a}{5-4a}, \\ 2 - b = \frac{b-1}{7-4b}. \end{cases}$$

Z drugiego równania otrzymujemy równanie kwadratowe $4b^2 - 16b + 15 = 0$, które ma dwa rozwiązania $b_1 = \frac{3}{2}$ oraz $b_2 = \frac{5}{2}$. Zauważmy, że b_2 leży poza przedziałem $\left(1, \frac{7}{4}\right)$, dlatego też $b = \frac{3}{2}$. Wstawiając to rozwiązanie do pierwszego równania otrzymujemy kolejne równanie kwadratowe $8a^2 - 26a + 15 = 0$, które ma dwa rozwiązania $a_1 = \frac{5}{2}$ oraz $a_2 = \frac{3}{4}$. Tylko drugie z tych rozwiązań leży w przedziale $\left(0, \frac{5}{4}\right)$ zatem $a = \frac{3}{4}$. Na mocy Twierdzenia 4.3 regresja (4.53) charakteryzuje bazowy rozkład prawdopodobieństwa jednoznacznie. W tym przykładzie mamy z $\eta(0) = 2$, $\eta(1) = \frac{13}{4}$ i $\eta(2) = 4$, a więc z równania (4.36) otrzymujemy $p_0 = \frac{1}{2}$ i $p_1 = p_2 = \frac{1}{4}$.

(b) Podobnie, jeżeli $S = \{0,1,2,3\}$ oraz $h(j) = j$ dla $j \in S$, to funkcja regresji

$$\xi(0) = \frac{37}{18}, \quad \xi(1) = \frac{29}{12}, \quad \xi(2) = \frac{23}{9}, \quad \xi(3) = 3,$$

jednoznacznie charakteryzuje rozkład prawdopodobieństwa

$$p_0 = p_2 = \frac{1}{3}, \quad p_1 = p_3 = \frac{1}{6}.$$

Aby to wykazać postępujemy analogicznie jak w punkcie (a). Najpierw znajdujemy rozwiązanie problemu różnicowego (4.43), który w tym przykładzie sprowadza się do rozwiązania kolejno trzech równań kwadratowych. Wychodząc od $\varphi(3) = 3$ oraz uwzględniając warunek $h < \varphi < \xi$ znajdujemy kolejno $\varphi(2) = \frac{7}{3}$, $\varphi(1) = 2$ oraz $\varphi(0) = \frac{4}{3}$. Następnie ze wzoru (4.36) wyznaczamy funkcję rozkładu prawdopodobieństwa.

Procedura zilustrowana w powyższym przykładzie może zostać zastosowana do znalezienia rozkładu prawdopodobieństwa o skończonym nośniku S charakteryzowanego przez dowolną funkcję regresji ξ słabych wartości rekordowych. Zasadniczo algorytm składa się z dwóch kroków. W pierwszym rozwiązując układ równań kwadratowych znajdujemy wartości funkcji φ

począwszy od $\varphi(N-1)$ do $\varphi(0)$. W drugim kroku przekształcamy wartości funkcji ξ, h oraz φ na wartości funkcji rozkładu prawdopodobieństwa zgodnie ze wzorem (4.36). Poprawność tego algorytmu wynika z Twierdzenia 4.3. Formalny zapis algorytmu znajduje się poniżej.

Algorytm 1 Wyznaczanie rozkładu prawdopodobieństwa determinowanego przez regresję $h(W_{r+2})$ względem W_r

Require: n-elementowy wektor $\boldsymbol{\xi}$ wartości funkcji regresji ξ oraz n-elementowy wektor $\boldsymbol{h}$ wartości funkcji h

$\boldsymbol{\varphi}[n] \leftarrow \boldsymbol{h}[n]$
for $i \leftarrow n-1$ **downto** 1 **do**
  $\boldsymbol{\varphi}[i] \leftarrow$ SOLVE: $(\boldsymbol{\varphi}[i+1]-x)(\boldsymbol{\xi}[i]-x) = (x-\boldsymbol{h}[i])(\boldsymbol{\xi}[i+1]-\boldsymbol{\xi}[i])$;
    pod warunkiem $\boldsymbol{h}[i] < x < \boldsymbol{\xi}[i]$
end for
$\boldsymbol{\eta} \leftarrow \boldsymbol{\xi} + \boldsymbol{\varphi}$
for $i \leftarrow 1$ **to** $n-1$ **do**
  $\boldsymbol{\mu}[i] \leftarrow \dfrac{\boldsymbol{\eta}[i+1]-\boldsymbol{\eta}[i]}{\boldsymbol{\xi}[i]-\boldsymbol{h}[i]}$
end for
$\boldsymbol{p}[1] \leftarrow \dfrac{\boldsymbol{\mu}[1]}{1+\boldsymbol{\mu}[1]}$
for $i \leftarrow 2$ **to** $n-1$ **do**
  $\boldsymbol{p}[i] \leftarrow \dfrac{\boldsymbol{\mu}[i]}{1+\boldsymbol{\mu}[i]} \prod_{k=1}^{i-1} (1+\boldsymbol{\mu}[k])^{-1}$
end for
$\boldsymbol{p}[n] \leftarrow \prod_{k=1}^{n-1} (1+\boldsymbol{\mu}[k])^{-1}$
return $\boldsymbol{p}$

Zaznaczmy, że z uwagi na ilość wymaganych obliczeń zależnych od liczby N prościej jest znaleźć wartości funkcji φ oraz funkcję rozkładu prawdopodobieństwa numerycznie. Przykładowo, jeżeli $N = 5$, ($S = \{0,1,2,3,4,5\}$), a funkcje ξ, h są określone przez wektory $\boldsymbol{\xi} = (3.4, 3.6, 4.1, 4.4, 4.6, 5)$, $\boldsymbol{h} = (0,1,2,3,4,5)$, to funkcja będąca rozwiązaniem problemu różnicowego (4.43) określona jest przez wektor $\boldsymbol{\varphi} = (2.2156982, 2.5898761, 3.3768469, 3.9480318, 4.3675444, 5)$, a rozkład prawdopodobieństwa determinowany przez relację regresji $\xi(j) = E\big(W_{r+2} \mid W_r = j\big)$ jest określony przez wektor $\boldsymbol{p} = (0.1444772, 0.2832625, 0.1677932, 0.1240757, 0.1773351, 0.1030563)$. Powyższe rozwiązanie zostało wyznaczone numerycznie przy użyciu funkcji zaimplementowanych w środowisku R.

W drugim przykładzie podamy nową charakteryzację pewnego rozkładu dyskretnego o nieskończonym nośniku.

Przykład 4.2. Załóżmy, że $S = \{0,1,2,\dots\}$ oraz $p_k = \frac{k+1}{2^{k+2}}$ dla $k = 0,1,2,\dots$. W przykładzie tym korzystać będziemy ze znanych wzorów dla sum szeregów:

$$\sum_{k=n}^{\infty} x^k = \frac{x^n}{1-x}, \tag{4.54}$$

$$\sum_{k=n}^{\infty} kx^{k+1} = \left(\frac{x}{1-x}\right)^2 \left[nx^{n-1}(1-x)+x^n\right] \tag{4.55}$$

oraz

$$\sum_{k=n}^{\infty} k(k+1)x^{k+2} = \left(\frac{x}{1-x}\right)^3 \left[n(n+1)x^{n-1}(1-x)^2 + 2\left((n+1)x^n - nx^{n+1}\right)\right], \tag{4.56}$$

o ile $|x| < 1$.

Stosując odpowiednio wzory (4.54) i (4.55) łatwo sprawdzić, że

$$q_n = \sum_{k=n}^{\infty} p_k = \frac{n+2}{2^{n+1}}, \quad n \in S.$$

Ustalmy $r \geq 1$ i niech $\varphi(n) = E\big(W_{r+2} \mid W_{r+1} = n\big)$ oraz $\xi(n) = E\big(\varphi(W_{r+1}) \mid W_r = n\big)$ dla $n \in S$. Z równości (1.26) mamy

$$\varphi(n) = \frac{1}{q_n}\sum_{k=n}^{\infty} kp_k = \frac{2^{n+1}}{n+2}\sum_{k=n}^{\infty} k(k+1)2^{-(k+2)}.$$

Dalej, wprost ze wzoru (4.56) dostajemy

$$\varphi(n) = E\big(W_{r+2} \mid W_{r+1} = n\big) = n+1+\frac{2}{n+2}.$$

Oczywiście φ spełnia warunek (4.46) (asymptotycznej liniowości) z parametrami $c = d = 1$. Ponownie korzystając z równości (1.26) dostajemy

$$\xi(n) = \frac{1}{q_n}\sum_{k=n}^{\infty} \varphi(k)p_k = \frac{2^{n+1}}{n+2}\sum_{k=n}^{\infty}\left(k+1+\frac{2}{k+2}\right)(k+1)2^{-(k+2)}.$$

Stosując teraz wzory (4.54)–(4.56) możemy sprowadzić powyższe wyrażenie do następującej postaci

$$\xi(n) = E\left(\varphi(W_{r+1}) \mid W_r = n\right) = n+2+\frac{2}{n+2}+R(n),$$

gdzie

$$R(n) = \frac{2^{n+2}}{n+2}\sum_{k=n}^{\infty}\frac{k+1}{k+2}\cdot 2^{-(k+2)}.$$

Zauważmy, że

$$0 \leq R(n) \leq \frac{2^{n+2}}{n+2}\sum_{k=n}^{\infty} 2^{-(k+2)} = \frac{2}{n+2}.$$

Stąd $R(n) \to 0$, gdy $n \to \infty$, a więc ξ jest funkcją asymptotycznie liniową, spełniającą warunek (4.42) z parametrami $a = 1$ i $b = 2$. Zatem na mocy Twierdzenia 4.4, $p_k = \frac{k+1}{2^{k+2}}$, $k = 0,1,2,\dots$ jest jedynym rozkładem określonym na nośniku S, dla którego zachodzi warunek

$$E\big(W_{r+2} \mid W_r = n\big) = n+2+\frac{2}{n+2}+R(n), \quad n \in S.$$

W kolejnych przykładach wykazujemy jednoznaczność charakteryzacji dwóch bardzo dobrze znanych dyskretnych rozkładów prawdopodobieństwa przez odpowiadające im regresje słabych wartości rekordowych nie znając ich jawnych postaci.

Przykład 4.3. Rozważmy rozkład Poissona określony następującą funkcją rozkładu prawdopodobieństwa

$$p_k = \frac{\theta^k}{k!}\exp(-\theta), \quad k = 0, 1, \ldots, \quad \theta > 0.$$

Korzystając z twierdzenia Stolza łatwo dostajemy, że $\mu(k) \to \infty$, gdy $k \to \infty$. Zatem na mocy Wniosku 4.2 możemy stwierdzić, że rozkład Poissona jest jednoznacznie charakteryzowany przez odpowiadającą mu regresję słabych wartości rekordowych (4.52), mimo że nie znamy analitycznej postaci tej funkcji.

Przykład 4.4. Dla rozkładu ujemnego dwumianowego określonego następującą funkcją rozkładu prawdopodobieństwa

$$p_k = \binom{r+k-1}{k}(1-p)^k p^r, \quad k = 0, 1, \ldots, \quad p \in (0,1),\ r > 0$$

otrzymujemy (po zastosowaniu twierdzenia Stolza) $\mu(k) \to \frac{p}{1-p} > 0$, gdy $k \to \infty$. Przeprowadzając dalej rozumowanie jak w dowodzie Lematu 4.6 dostajemy, że

$$\lim_{k\to\infty}[\varphi(k) - k] = \lim_{k\to\infty}\frac{1}{\mu(k)} = \frac{1-p}{p}$$

oraz

$$\lim_{k\to\infty}[\xi(k) - k] = 2\lim_{k\to\infty}\frac{1}{\mu(k)} = 2\frac{1-p}{p}.$$

Zatem, dla dowolnego rozkładu ujemnego dwumianowego odpowiadające mu regresje słabych wartości rekordowych ξ oraz φ spełniają odpowiednio warunki (4.42) oraz (4.46) z $a = c = 1$. Co więcej, wiemy, że φ jest rozwiązaniem problemu (4.43). Zatem z Twierdzenia 4.4 dostajemy, że rozkład ujemny dwumianowy jest jednoznacznie charakteryzowany przez odpowiadającą mu funkcję regresji ξ.

4.5 Liniowość regresji słabych rekordów o odstępie dwa

W tym podrozdziale wykorzystamy wyniki otrzymane w podrozdziałach 4.2 i 4.3 do przeanalizowania problemu charakteryzacji rozkładów prawdopodobieństwa przez liniowość regresji słabych rekordów o odstępie dwa, to znaczy regresji W_{r+2} względem W_r, dla pewnego naturalnego $r \geq 1$. Pierwsze rozwiązanie tego problemu zostało podane przez Wesołowskiego i Ahsanullaha w pracy [57], powiązane wyniki można znaleźć również w pracy [3].

Rozpoczniemy od podania definicji trzech rozkładów prawdopodobieństwa charakteryzowanych w tym podrozdziale (por. [34])

(1) $nh_I(\alpha,\beta,n)$ – **ujemny rozkład hipergeometryczny pierwszego rodzaju** z funkcją rozkładu prawdopodobieństwa daną wzorem

$$p_k = \frac{\binom{\alpha+k-1}{k}\binom{\beta-\alpha+n-k}{n-k}}{\binom{\beta+n}{n}}, \quad k = 0,1,\ldots,n,$$

gdzie $\alpha,\beta \in \mathbb{R}$ są parametrami takimi, że $\beta+1 > \alpha > 0$, natomiast $n \in \mathbb{N}$;

(2) $nh_{II}(\alpha,\beta,\gamma)$ – **ujemny rozkład hipergeometryczny drugiego rodzaju** z funkcją rozkładu prawdopodobieństwa daną wzorem

$$p_k = \frac{\gamma}{\gamma+k}\frac{\binom{\beta}{\gamma}\binom{\alpha+k-1}{k}}{\binom{\alpha+\beta+k}{\gamma+k}}, \quad k = 0,1,\ldots,$$

gdzie $\alpha,\beta,\gamma \in \mathbb{R}$ są parametrami takimi, że $\alpha > 0$, $\beta+1 > \gamma > 0$;

(3) $ge(p)$ – **rozkład geometryczny** opisany następującą funkcją rozkładu prawdopodobieństwa

$$p_k = p(1-p)^k, \quad k = 0,1,\ldots,$$

gdzie $p \in (0,1)$.

Załóżmy najpierw, że dla ustalonej liczby naturalnej $r \geq 1$ zachodzi liniowa regresja W_{r+1} względem W_r

$$E\left(W_{r+1} \mid W_r = j\right) = aj + b, \quad j \in S,$$

dla pewnych $a,b \in \mathbb{R}$. Wówczas oczywiście $a > 0$ oraz $b > 0$ oraz korzystając z (4.11) lub (4.12) można w prosty sposób wykazać, że możliwe są tylko trzy następujące przypadki:

(i) $0 < a < 1$, $N = b/(1-a) < \infty$ oraz X_1 ma rozkład ujemny hipergeometryczny pierwszego rodzaju $nh_I(1,\beta,n)$ z parametrami $\beta = a/(1-a)$, $n = N$;

(ii) $a = 1$, $N = \infty$ oraz X_1 ma rozkład geometryczny $ge(p)$ z parametrem $p = 1/(1+b)$;

(iii) $a > 1$, $N = \infty$ oraz X_1 ma rozkład ujemny hipergeometryczny drugiego rodzaju $nh_{II}(1,\beta,\gamma)$ z parametrami $\beta = (b+1)/(a-1)$ i $\gamma = b/(a-1)$.

Zatem warunek liniowej regresji kolejnych ($\ell = 1$) słabych rekordów determinuje jednoznacznie trzy rodziny dyskretnych rozkładów prawdopodobieństwa, które są zależne od wartości współczynnika kierunkowego prostej regresji. Powyższy wynik dla nieskończonego nośnika S został podany w pracach [2] i [54]. Kompletną charakteryzację rozkładów prawdopodobieństwa, obejmującą również skończony nośnik S, przez liniowe regresje kolejnych słabych rekordów podali Wesołowski i Ahsanullah w pracy [57]. Udowodnili oni również, że te same trzy rodziny rozkładów prawdopodobieństwa są charakteryzowane, jeśli założymy liniowość regresji słabych rekordów o odstępie $\ell = 2$. Poniżej stosując Twierdzenia 4.3 oraz 4.4, podamy inny prostszy i bardziej bezpośredni dowód tego wyniku.

Załóżmy, że dla ustalonej liczby naturalnej $r \geq 1$ zachodzi liniowość regresji W_{r+2} względem W_r

$$E\left(W_{r+2} \mid W_r = j\right) = aj + b, \quad j \in S, \tag{4.57}$$

dla pewnych $a, b \in \mathbb{R}$. Z własności funkcji regresji (Lemat 1.6) otrzymujemy, że $a > 0$ oraz $b > 0$. Ponadto

$$N = \inf\{j = 0, 1, \ldots : aj + b = j\} = \begin{cases} \frac{b}{1-a} \in \mathbb{N}, & \text{gdy} \quad 0 < a < 1, \\ \infty, & \text{gdy} \quad a \geq 1. \end{cases}$$

Naszym zadaniem jest znalezienie funkcji y, która jest rozwiązaniem następującego problemu różnicowego

$$\begin{cases} \Delta y(j) = a\dfrac{y(j) - j}{aj + b - y(j)}, \\ j < y(j) < aj + b, \end{cases} \qquad j \in \bar{S}. \tag{4.58}$$

Sprawdźmy najpierw, czy istnieją liniowe rozwiązania tego problemu mające postać

$$y(j) = cj + d. \tag{4.59}$$

Oczywiście $c > 0$ oraz $d \in (0, b)$. Podstawiając $a = 1$ do (4.58) otrzymujemy $y(j) = j + d$ oraz $1 = d/(b - d)$, co daje $d = b/2 \in (0, b)$. Jeżeli natomiast $a \neq 1$, to

$$c = a\frac{cj + d - j}{aj + b - (cj + d)}.$$

Stąd dostajemy równanie $(a - c)cj + (b - d)c = a(c - 1)j + ad$, z którego wynika następujący układ równań

$$\begin{cases} (a - c)c = a(c - 1), \\ (b - d)c = ad. \end{cases}$$

Pierwsze równanie jest równaniem kwadratowym $c^2 = a$, które ma tylko jeden dodatni pierwiastek $c = \sqrt{a}$. Natomiast z drugiego równania otrzymujemy

$$d = \frac{b}{1 + \sqrt{a}} \in (0, b). \tag{4.60}$$

Podsumowując, jeżeli $a > 0$ i $b > 0$, to problem różnicowy (4.58) ma dokładnie jedno rozwiązanie liniowe $\varphi(j) = cj + d$ takie, że $c = 1$ oraz $d = b/2$, gdy $a = 1$, w przeciwnym razie $c = \sqrt{a}$, a parametr d jest dany przez (4.60).

W następnym kroku musimy wykazać, że badany problem (4.58) nie ma innych rozwiązań. Jeżeli $0 < a < 1$, to $N = b/(1 - a) < \infty$. W tym przypadku jednoznaczność rozwiązania (4.59) dostajemy z Twierdzenia 4.3. Dla $a \geq 1$ mamy $N = \infty$, tak więc $h(N) = \infty$ i nie możemy zastosować Twierdzenia 4.3 jak wyżej. Jednakże, jednoznaczność rozwiązania problemu (4.58) w tym przypadku wynika z Twierdzenia 4.4. Bezpośrednie rachunki pokazują, że φ spełnia również warunki (4.28), (4.29) oraz (4.30).

W ostatnim kroku wystarczy skorzystać z (4.31) lub (4.36), aby otrzymać trzy rodziny dyskretnych rozkłady prawdopodobieństwa takie same jak dla liniowej regresji kolejnych słabych wartości rekordowych.

Twierdzenie 4.5. *Niech $\{X_n, n \geq 1\}$ będzie ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie z nośnikiem $S = \{0, \ldots, N\}$, gdzie $N \leq \infty$. Załóżmy, że zachodzi liniowość regresji* (4.57). *Wówczas $a > 0$, $b > 0$, i możliwe są tylko trzy następujące przypadki:*

(i) $0 < a < 1$ *oraz X_1 ma rozkład ujemny hipergeometryczny pierwszego rodzaju $nh_I(1, \beta, n)$ z parametrami $\beta = (a + \sqrt{a})/(1 - a)$ oraz $n = b/(1 - a)$;*

(ii) $a = 1$ *oraz X_1 ma rozkład geometryczny $ge(p)$ z parametrem $p = 2/(2 + b)$;*

(iii) $a > 1$ *oraz X_1 ma rozkład ujemny hipergeometryczny drugiego rodzaju $nh_{II}(1, \beta, \gamma)$ z parametrami $\beta = (b + \sqrt{a} + 1)/(a - 1)$ oraz $\gamma = b/(a - 1)$.*

Podkreślmy, że zaprezentowany nowy dowód powyższego twierdzenia bazuje na rozwiązaniu równania różnicowego pierwszego rzędu, podczas gdy dowód podany w pracy [57] wymaga rozwiązania równania różnicowego drugiego rzędu, co jest znacznie trudniejsze.

Na koniec tego podrozdziału zaznaczmy, że w ogólnym przypadku, gdy $\ell > 2$ problem charakteryzacji rozkładu przez regresje słabych rekordów jest tylko częściowo rozwiązany. Problem jest nadal otwarty, gdy $\ell \geq 5$ i $a \geq 1$ [38].

4.6 Charakteryzacje przez regresje zwykłych rekordów

Na zakończenie tego rozdziału odniesiemy się do charakteryzacji rozkładów prawdopodobieństwa przez regresje zwykłych rekordów dyskretnych (1.28). Wiemy, że zwykłe rekordy w przypadku dyskretnym są poprawnie określone tylko dla zmiennych losowych o nieskończonym nośniku $S = \{0, 1, 2, \ldots\}$. Ponadto zwykłe rekordy tworzą ciąg ściśle rosnący, tak więc każdy kolejny rekord przekracza poprzedni o co najmniej 1, zatem $R_r \geq r - 1$ dla $r \geq 1$. Wiemy również, że ciąg dyskretnych rekordów $\{R_r, r \geq 1\}$ tworzy łańcuch Markowa z prawdopodobieństwami przejścia (1.4). Stąd pojawia się naturalne pytanie, czy podobne kryterium jednoznaczności jak dla słabych rekordów można wykazać również dla zwykłych rekordów.

W kontekście charakteryzacji rozkładów prawdopodobieństwa należy zaznaczyć, że warunkowy rozkład prawdopodobieństwa R_{r+1} pod warunkiem R_r, a w konsekwencji regresja

R_{r+1} względem R_r nie zawiera informacji o prawdopodobieństwach $p_0, p_1, \ldots, p_{r-1}$ wystąpienia wartości $0, 1, \ldots, r-1$. Zatem dysponując funkcją regresji zwykłych rekordów nie możemy jednoznacznie wyznaczyć całego rozkładu prawdopodobieństwa. Możliwe jest tylko otrzymanie ich ogonów $p_r, p_{r+1}, \ldots$.

Korzystając z rekurencyjnej struktury wynikającej z własności Markowa oraz z własności regresji wartości rekordowych przedstawionych w Lemacie 1.7 możemy otrzymać następujący wynik, będący odpowiednikiem Twierdzenia 4.1.

Twierdzenie 4.6. *Dla ustalonego $r \geq 1$, niech R_r, R_{r+2} będą rekordami z rozkładu o dystrybuancie F określonej na nośniku $S = \{0, 1, \ldots\}$. Załóżmy, że $h \colon S \to \mathbb{R}$ jest funkcją ściśle rosnącą i taką, że $E\left|h(R_{r+2})\right| < \infty$. Wówczas funkcja regresji $\xi \colon \{r-1, r, \ldots\} \to \mathbb{R}$ określona wzorem*

$$\xi(j) = E\left(h(R_{r+2}) \mid R_r = j\right), \quad j \geq r-1, \tag{4.61}$$

charakteryzuje jednoznacznie ogon rozkładu F wtedy i tylko wtedy, gdy równanie różnicowe

$$\Delta y(j) = \frac{y(j+1) - h(j+1)}{\xi(j+1) - y(j+1)} \Delta \xi(j), \quad j \geq r, \tag{4.62}$$

ma jednoznaczne rozwiązanie $y = \varphi(j)$ na $S \setminus \{0, 1, \ldots, r-1\}$ spełniające warunki

$$h(j+2) < \varphi(j+1) < \xi(j), \quad j \geq r-1,$$

$$\sum_{j=r}^{\infty} \frac{\Delta \xi(j-1)}{\xi(j) - \varphi(j)} = \infty,$$

oraz

$$\lim_{n \to \infty} \xi(n) \prod_{j=r}^{n} \left(1 - \frac{\Delta \xi(j-1)}{\xi(j) - \varphi(j)}\right)^{-1} = 0.$$

Oczywiście, wzory określające funkcję rozkładu F są podobne do przypadku regresji słabych rekordów, aczkolwiek potrzebne są pewne modyfikacje. Mianowicie, jeżeli $p_0, \ldots, p_{r-1}$ są znane to q_r oraz $\bar{F}(r-1) = q_r$ są również znane. Wtedy można pokazać, że dla $n \geq r$

$$\bar{F}(n) = q_r \prod_{j=r}^{n} \left(1 - \frac{\Delta \xi(k-1)}{\xi(j) - \varphi(j)}\right).$$

Kryterium jednoznaczności charakteryzacji rozkładów prawdopodobieństwa określone w Twierdzeniu 4.6 okazuje się trudniejsze w użyciu niż dla przypadku słabych rekordów. Dla równania różnicowego (4.62) nie można przeprowadzić analogicznego rozumowania jak w podrozdziale 4.3.1 dla słabych rekordów. W szczególności próba udowodnienia odpowiednika Lematu 4.3 prowadzi do następującej relacji (por. równanie (4.44))

$$y(j) - \varphi(j) = [y(j+1) - \varphi(j+1)] \left(1 - \frac{[\xi(j+1) - h(j+1)] \Delta \xi(j)}{[\xi(j+1) - \varphi(j+1)][\xi(j+1) - y(j+1)]}\right).$$

Przeciwnie niż to było dla przypadku słabych wartości rekordowych nie znajdujemy argumentów świadczących o tym, że wyrażenie w nawiasach po prawej stronie powyższego wyrażenia jest dodatnie. Z tego powodu nie możemy zastosować analogicznego rozumowania jak w dowodzie Lematu 4.4. Pytanie jak wykazać, że $\varphi(j) = E\left(h(R_{r+2}) \mid R_{r+1} = j\right)$ jest jedynym rozwiązaniem równania różnicowego (4.62), a co za tym idzie, że regresja (4.61) zwykłych rekordów jednoznacznie charakteryzuje ogon dyskretnego rozkładu prawdopodobieństwa pozostaje otwarte.

Podsumowanie

W niniejszej pracy zaprezentowaliśmy całkowicie nowe podejście do znanego problemu charakteryzacji rozkładów prawdopodobieństwa przez regresje modeli uporządkowanych zmiennych losowych. Nasze podejście oparte na własności Markowa rozważnych modeli, polega na przeformułowaniu problemu. Mianowicie wyjściowy problem charakteryzacyjny zostaje zastąpiony nowym problemem albo różniczkowym, albo całkowym, albo różnicowym z niestandardowymi warunkami ograniczającymi. Takie postawienie problemu przesuwa aspekty probabilistyczne i statystyczne problemu na drugi plan. Główny problem znajduje teraz miejsce w teorii równań różniczkowych bądź całkowych, bądź różnicowych. Zaprezentowane wyniki są zupełnie nowe, nawet dla najpopularniejszych modeli uporządkowanych zmiennych losowych tj. statystyk porządkowych i wartości rekordowych.

W rezultacie zastosowania nowego podejścia uzyskaliśmy szereg nowych charakteryzacji rozkładów prawdopodobieństwa. Przykłady 2.1, 2.2, 3.3–3.6 oraz 4.1–4.4 podają nowe wyniki, których nie można byłoby otrzymać bez nowego podejścia przedstawionego w niniejszej pracy. Zaprezentowaliśmy również nową technikę dowodzenia jednoznaczności charakteryzacji bez znajomości jawnej postaci funkcji regresji ξ. W ten sposób wykazaliśmy, że popularne rozkłady takie jak normalny, Weibulla z parametrem kształtu $\delta > 1$, Gompertza, gamma, Gumbela i logistyczny są jednoznacznie charakteryzowane przez odpowiadające im regresje uogólnionych statystyk porządkowych, a rozkłady Poissona i ujemny dwumianowy są jednoznacznie charakteryzowane przez odpowiadające im regresje słabych wartości rekordowych.

Z drugiej strony otrzymane rezultaty są trudne w zastosowaniu, co zostało zobrazowane w podrozdziałach 2.2, 3.2 i 4.3. Wyniki charakteryzacji są nadal częściowe i wiele problemów pozostaje wciąż otwartych. W szczególności są to:

(1) Zastosowanie nowych kryteriów jednoznaczności charakteryzacji rozkładów prawdopodobieństwa dla przypadku $\ell > 2$. Czy można uogólnić wyniki otrzymane dla $\ell = 2$ i szczególnych postaci funkcji h oraz ξ? Między innymi, czy jest możliwe rozszerzenie Twierdzeń 2.3, 3.2 i 4.3 na przypadek $\ell > 2$ (por. Uwaga 2.5);

(2) Uzupełnienie wyników dla $\ell = 2$ oraz gdy h jest funkcja nieograniczoną. Jak pokazuje Przykład 3.7 nie mamy dowodu, że regresje, które nie są asymptotycznie liniowe względem funkcji h jednoznacznie charakteryzują rozkłady prawdopodobieństwa;

(3) Dla ustalonych $r, \ell \geq 1$, parametrów modelu $\gamma_1, \ldots, \gamma_{r+\ell}$ oraz funkcji h określenie warunków koniecznych i dostatecznych, na to aby pewna funkcja ξ była funkcją regresji uogólnionych statystyk porządkowych. Podobny problem dla słabych wartości rekordowych;

(4) Wykazanie jednoznaczności rozwiązania problemu różniczkowego (2.28), gdy $h(x) = x$ oraz $\xi(x) = ax + b$, $a > 0$, w przypadku $\ell > 2$. Stanowiłoby to nowy dowód Twierdzenia 5.1 z pracy [19] (por. Uwaga 2.7);

(5) Analogiczny problem jak wyżej, dla regresji słabych wartości rekordowych, mianowicie wykazanie jednoznaczności rozwiązania problemu różnicowego (4.37), gdy $h(j) = j$ oraz $\xi(j) = aj + b$, $a > 0$, w przypadku $\ell > 2$. Z jednej strony byłby to nowy dowód znanych wyników dla szczególnych przypadków regresji, a z drugiej uzupełnienie problemu jednoznacznej charakteryzacji rozkładów prawdopodobieństwa przez liniową regresję słabych wartości rekordowych w przypadku $\ell \geq 5$ i $a \geq 1$;

(6) Uzupełnienie wyników dla regresji dyskretnych wartości rekordowych (por. podrozdział 4.6).

Kolejnym problemem, który wymaga dalszych analiz jest rozszerzenie wyników prezentowanych w niniejszej rozprawie na przypadek regresji dualnej, to znaczy regresji $X_*^{(r)}$ względem $X_*^{(r+\ell)}$, czy też regresji W_r względem $W_{r+\ell}$. Jednakże w tym przypadku, rachunki są znacznie trudniejsze, gdyż odpowiednie funkcje warunkowych rozkładów wyrażają się bardziej skomplikowanymi wzorami (zob. [11], [22] i referencje tam zawarte).

Bibliografia

[1] M. Ahsanullah, V. B. Nevzorov. *Records via Probability Theory*. Atlantis Studies in Probability and Statistics. Atlantis Press, 2015.

[2] F. A. Aliev. Characterization of distributions through weak records. *J. Appl. Statist. Sci.*, 8(1):13–16, 1998.

[3] F. A. Aliev. New characterization of discrete distributions through weak records. *Theory Probab. Appl. (English translation)*, 44(4):756–761, 2000.

[4] B. C. Arnold, N. Balakrishan, N. H. Nagaraja. *Records*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., New York, 1998.

[5] B. C. Arnold, N. Balakrishan, N. H. Nagaraja. *A first Course in Order Statistics*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2008.

[6] N. Balakrishnan, R. Aggarwala. *Progressive censoring: Theory, Methods, and Applications*. Birkhäuser Boston Inc., Boston, MA, 2000.

[7] N. Balakrishnan, E. Cramer. *The art of progressive censoring: Applications to Reliability and Quality*. Statistics for Industry and Technology. Birkhäuser, New York, 2014.

[8] N. Balakrishnan, E. Cramer, U. Kamps. Bounds for means and variances of progressive type II censored order statistics. *Statist. Probab. Lett.*, 54(3):301–315, 2001.

[9] N. Balakrishnan, A. Dembińska. Ordered random variables from discontinuous distributions. *J. Iranian Statist. Soc.*, 6(1):31–46, 2007.

[10] R. E. Barlow, F. Proschan. *Mathematical theory of reliability*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1996.

[11] M. Bieniek. On characterizations of distributions by regression of adjacent generalized order statistics. *Metrika*, 66(2):233–242, 2007.

[12] M. Bieniek. A note on characterizations of distributions by regressions of non-adjacent generalized order statistics. *Aust. N. Z. J. Stat.*, 51(1):89–99, 2009.

[13] M. Bieniek, K. Maciąg. Uniqueness of characterization of absolutely continuous distributions by regressions of generalized order statistics. *AStA Advances in Statistical Analysis*, 102(3):359–380, 2018.

[14] M. Bieniek, K. Maciąg. On the unique characterization of continuous distributions by single regression of non–adjacent generalized order statistics. *J. Integral Equations and Applications*, 30(4):491–519, 2018.

[15] M. Bieniek, K. Maciąg. On the problem of the unique characterization of discrete distributions by single regression of weak records. *Statistics*, 52(3):533–551, 2018.

[16] M. Bieniek, D. Szynal. Characterizations of distributions via linearity of regression of generalized order statistics. *Metrika*, 58(3):259–271, 2003.

[17] K. N. Chandler. The distribution and frequency of record values. *J. Roy. Statist. Soc. Ser. B*, 14(2):220–228, 1952.

[18] E. Cramer, U. Kamps. Marginal distributions of sequential and generalized order statistics. *Metrika*, 58(3):293–310, 2003.

[19] E. Cramer, U. Kamps, C. Keseling. Characterizations via linear regression of ordered random variables: a unifying approach. *Communications in Statistics - Theory and Methods*, 33(12):2885–2911, 2004.

[20] E. Cramer, U. Kamps, T. Rychlik. On the existence of moments of generalized order statistics. *Statist. Probab. Lett.*, 59:397–404, 2002.

[21] E. Cramer, T. T. H. Tran. Generalized order statistics from arbitrary distributions and the markov chain property. *J. Statist. Plann. Inference*, 139(12):4064–4071, 2009.

[22] K. Danielak, A. Dembińska. On characterizing discrete distributions via conditional expectations of weak record values. *Metrika*, 66(2):129–138, 2007.

[23] H. A. David, N. H. Nagaraja. *Order statistics*. Wiley Series in Probability and Statistics. John Wiley & Sons, New York, 3rd edition, 2003.

[24] A. Dembińska. A review on characterizations of discrete distributions based on records and *k*th records. *Communications in Statistics - Theory and Methods*, 36(7):1381–1387, 2007.

[25] A. Dembińska, J. Wesołowski. Linearity of regression for non-adjacent order statistics. *Metrika*, 48(3):215–222, 1998.

[26] A. Dembińska, J. Wesołowski. Linearity of regression for non-adjacent record values. *J. Statist. Plann. Inference*, 90(2):195–205, 2000.

[27] W. Dziubdziela, B. Kopociński. Limiting properties of the k-th record values. *Zastos. Mat.*, 15(2):187–190, 1976.

[28] T. S. Ferguson. On characterizing distributions by properties of order statistics. *Sankhyā Ser. A*, 29(3):265–278, 1967.

[29] M. Franco, J. M. Ruiz. On characterization of continuous distributions with adjacent order statistics. *Statistics*, 26(4):375–385, 1995.

[30] M. Franco, J. M. Ruiz. On characterization of continuous distributions by conditional expectation of record values. *Sankhyā Ser. A*, 58(1):135–141, 1996.

[31] M. Franco, J. M. Ruiz. Characterization based on conditional expectations of adjacent order statistics: a unified approach. *Proc. Amer. Math. Soc.*, 127(3):861–874, 1999.

[32] J. Galambos, S. Kotz. *Characterizations of probability distributions*. Lecture Notes in Mathematics. Springer, New York, 1978.

[33] Z. Grudzień, D. Szynal. On the expected values of k-th record values and associated characterizations of distributions. *Probability and Statistical Decision Theory*, A:119–127, 1985.

[34] N. L. Johnson, A. W. Kemp, S. Kotz. *Discrete Univariate Distributions*. Wiley, New York, 1992.

[35] W. J. Kaczor, M. T. Nowak. *Problems in mathematical analysis III: Integration*. American Mathematical Society, Providence, RI, 2003.

[36] U. Kamps. A concept of generalized order statistics. *J. Statist. Plann. Inference*, 48(1):1–23, 1995.

[37] U. Kamps. *A concept of generalized order statistics*. B. G. Teubner, Stuttgart, 1995.

[38] R. Karczewski, J. Wesołowski. Linearity of regression for weak records, revisited. *Statistics*, 51(4):878–887, 2017.

[39] C. Keseling. Conditional distributions of generalized order statistics and some characterizations. *Metrika*, 49(1):27–40, 1999.

[40] R. M. Korwar. On characterizing distributions for which the second record value has a linear regression on the first. *Sankhyā Ser. B*, 46(1):108–109, 1984.

[41] F. López-Blázquez. Linear prediction of weak records: the discrete case. *Theory Probab. Appl.*, 48(4):718–723, 2004.

[42] A. M. Mathai. *A handbook of generalized special functions for statistical and physical sciences*. Oxford Science Publications. Clarendon Press, New York, 1993.

[43] H. N. Nagaraja. On a characterization based on record values. *Austral. J. Statist.*, 19(1):70–73, 1977.

[44] H. N. Nagaraja. On the expected values of record values. *Austral. J. Statist.*, 20(2):176–182, 1978.

[45] H. N. Nagaraja. On the non-markovian structure of discrete order statistics. *J. Statist. Plann. Inference*, 7(1):29–33, 1982.

[46] V. B. Nevzorov. *Records: Mathematical Theory*. Translations of mathematical monographs. American Mathematical Society, Providence, RI, 2001.

[47] D. Pfeifer. Characterizations of exponential distributions by independent nonstationary record increments. *J. Appl. Probab.*, 19(1):127–135, 1982.

[48] A. D. Polyanin, V. F. Zaitsev. *Handbook of exact solutions for ordinary differential equations*. Chapman & Hall/CRC Press, Boca Raton, FL, 2nd edition, 2003.

[49] C. R. Rao, D. N. Shanbhag. Recent results on characterization of probability distributions: a unified approach through extensions of Deny's theorem. *Adv. in Appl. Probab.*, 18(3):660–678, 1986.

[50] W. Rudin. *Analiza rzeczywista i zespolona*. Wydawnictwo Naukowe PWN, Warszawa, Wyd. 2, 1998.

[51] A. A. Salvia, R. C. Bollinger. On discrete hazard functions. *IEEE Transactions on Reliability*, 31(5):458–459, 1982.

[52] M. Shaked, J. G. Shanthikumar, J. B. Valdez-Torres. Discrete hazard rate functions. *Computers and Operations Research*, 22(4):391–402, 1995.

[53] R. C. Srivastava. Two characterizations of the geometric distribution by record values. *Sankhyā Ser. B*, 40(3/4):276–278, 1979.

[54] A. V. Stepanov. A characterization theorem for weak records. *Theory Probab. Appl.*, 38(4):762–764, 1994.

[55] T. T. H. Tran. *Discrete Generalized Order Statistics*. Ph.D. Thesis. University of Oldenburg, Germany, 2006.

[56] W. Vervaat. Limit theorems for records from discrete distributions. *Stoch. Proc. Appl.*, 1(4):317–334, 1973.

[57] J. Wesołowski, M. Ahsanullah. Linearity of regression for non-adjacent weak records. *Statistica Sinica*, 11(1):39–52, 2001.