

Tomasz Rychlik
Instytut Matematyczny PAN
Warszawa

Warszawa, 18 września 2019 r.

Recenzja rozprawy doktorskiej pani mgr Krystyny Maciąg
pt. *Charakteryzacje rozkładów prawdopodobieństwa przez regresje
niekolejnych uogólnionych statystyk porządkowych
oraz dyskretnych wartości rekordowych*

Najogólniej mówiąc, rozprawa dotyczy charakteryzacji rozkładu wektora lub ciągu niezależnych zmiennych losowych o jednakowym rozkładzie na podstawie pojedynczych regresji funkcji niekolejnych elementów wybranych uporządkowanych ciągów opartych na zmiennych bazowych, gdy elementy te tworzą łańcuch Markowa. Dokładniej, rozważane są dwa modele. Po pierwsze, jest to wektor uogólnionych statystyk pozycyjnych $X_*^{(1)}, \dots, X_*^{(r)}, \dots, X_*^{(r+\ell)}$ o zadanych parametrach $\gamma_1, \dots, \gamma_r, \dots, \gamma_{r+\ell} > 0$ i charakteryzowanej ciągłej dystrybucji bazowej F . Zadanie polega na jednoznacznym wyznaczeniu F w sytuacji, gdy znana jest pojedyncza zależność regresyjna $\xi(x) = \mathbb{E}(h(X_*^{(r+\ell)}) | X_*^{(r)} = x)$ dla zadanych ściśle rosnących funkcji ciągłych ξ i h na przedziałowym nośniku (α, β) dystrybucyjności F oraz liczb naturalnych $r \geq 1$ i $\ell \geq 2$. Po drugie, analogiczne zadanie rozważano dla regresji $\xi(x) = \mathbb{E}(h(W_{r+\ell}) | W_r = x)$ w ciągu słabych rekordów $W_1, \dots, W_r, \dots, W_{r+\ell}$ zdefiniowanych na bazie dyskretnych zmiennych losowych o wartościach całkowitych nieujemnych i skończonym bądź nieskończonym nośniku.

Dla uogólnionych statystyk pozycyjnych zadanie to było rozwiązane w przypadku $\ell = 1$ (p. pozycja [19] bibliografii rozprawy), a dla niekolejnych $\ell \geq 2$ tylko przy ograniczeniu do liniowych funkcji ξ i h (p. również [16]). Ponadto Bieniek [12] pokazał, że pełną charakteryzację rozkładu można uzyskać przy znajomości drugiej funkcji regresji $\xi_1(x) = \mathbb{E}(h(X_*^{(r+\ell)}) | X_*^{(r+1)} = x)$. W przypadku słabych rekordów znane są wyniki charakteryzacyjne dla $\ell = 1$ (p. [53]) oraz dla $\ell = 2$, gdy h jest identycznością, a nośnik obserwacji jest nieskończony (por. [3]). Żadna z metod dowodowych w tych pracach nie da się rozszerzyć na przypadek większych odstępów ℓ między uporządkowanymi obserwacjami i ogólnych funkcji regresji. W pracy doktorskiej uogólnienie to udało się Autorce częściowo uzyskać (z przedostatniego słowa tłumacząc się poniżej).

Rozprawa liczy 90 stron, składa się ze wstępu z krótkim omówieniem jej wartości, 4 rozdziałów, podsumowania wskazującego najistotniejsze osiągnięcia i problemy nie rozwiązane, oraz bibliografii liczącej 56 pozycji literaturowych zacytowanych w pracy. Rozdział 1 zawiera wiadomości wstępne wykorzystane dalej w dowodach głównych wyników. Są tam opisane modele ciągów uporządkowanych zmiennych losowych analizowanych w pracy, własności Markowa tych ciągów, oraz podstawowe własności funkcji regresji w rozważanych modelach. Szczególnie ważne są tu rekurencyjna zależność kolejnych funkcji regresji opisana w lemacie 1.1, uporządkowanie $\xi > h$ oraz zbieżność ich wartości do wspólnej granicy na prawym końcu nośnika β (p. lematy 1.3 i 1.5). Rozdział ten zawiera też krótki opis dotychczasowych wyników w dziedzinie charakteryzacji rozkładów prawdopodobieństwa przez regresje uporządkowanych zmiennych losowych.

Kolejne trzy rozdziały obejmują zasadnicze wyniki rozprawy. Zawartość każdego z nich jest niemal identyczna z zawartościami prac odpowiednio [13], [14] i [15] (przy okazji: czasopisma, w których je opublikowano, są wartościowymi periodykami ze statystyki i analizy). Również ich struktury są do siebie bardzo zbliżone, co ilustrują też nazwy ich podrozdziałów i podpodrozdziałów. Najpierw zawierają ogólne kryteria charakteryzacji dla elementów ciągów o odstępach $\ell = 2$, ich uogólnienia na przypadki $\ell \geq 2$, a następnie dokładniejsze wyniki dla $\ell = 2$: opisy rozwiązań odpowiednich równań różniczkowych, całkowych i różnicowych i sprawdzalne warunki jednoznacznej charakteryzacji w przypadkach skończoności i nieskończoności funkcji h na końcu nośnika. Każdy z rozdziałów jest uzupełniony przykładami nowych charakteryzacji, uzyskanych za pomocą opisanej tam teorii.

Wszystkie rozdziały łączy też wspólna idea rozwiązywania problemów charakteryzacji. Wprowadza się dodatkowo pośrednie funkcje rosnące $h < \varphi_1 < \varphi_{\ell-1} < \xi$. Jeżeli są to funkcje regresji kolejnych elementów ciągu Markowa $Y_r, \dots, Y_{r+\ell}$ (w pracy Y_i to $X_*^{(i)}$ lub W_i), tzn.

$$\varphi_i(x) = \mathbb{E}(\varphi_{i+1}(Y_{r+i+1}) | Y_{r+i} = x), \quad (1)$$

i dodatkowo

$$\xi(x) = \mathbb{E}(\varphi_{r+1}(Y_{r+i+1}) | Y_r = x), \quad (2)$$

$$\varphi_{\ell-1}(x) = \mathbb{E}(h(Y_{r+\ell}) | Y_{\ell-1} = x), \quad (3)$$

dla dowolnego argumentu x z nośnika, to również $\xi(x) = \mathbb{E}(h(Y_{r+\ell}) | Y_r = x)$. Warunki (1)–(3) determinują pewne relacje między funkcjami φ_i , $i = 1, \dots, \ell -$

1, oraz h i ξ . Jeśli więc te relacje opisują jednoznacznie uporządkowany ciąg funkcji $h = \varphi_0 < \varphi_1 < \dots < \varphi_{\ell-1} < \varphi_\ell = \xi$, to odpowiednia ich kombinacja liniowa jest funkcją intensywności awarii rozkładu generującego uporządkowany ciąg Markowa.

Rozdział 2 dotyczy charakteryzacji rozkładów bazowych o ciągłych gęstościach za pomocą regresji niekolejnych uogólnionych statystyk pozycyjnych. Przy takim założeniu funkcje regresji mają ciągłe pochodne, a wiążące je równania różniczkowe mają postać

$$\varphi'_i(x) = \frac{\gamma_{r+i+1}}{\gamma_{r+i}} \frac{\varphi_i(x) - \varphi_{i+1}(x)}{\xi(x) - \varphi_1(x)} \xi'(x), \quad i = 1, \dots, \ell - 1. \quad (4)$$

Kłopot stanowi fakt, że równania te są uzupełnione nietypowymi nierównościami warunkami brzegowymi $h(x) < \varphi_1(x) < \dots < \varphi_{\ell-1}(x) < \xi(x)$, dla których nie są znane metody jednoznacznego rozwiązywania równań różniczkowych. Pewne wyniki dotyczące jednoznaczności udało się jedynie uzyskać w przypadku odstepu $\ell = 2$. Zostały one udowodnione, gdy wartości funkcji h są ograniczone z góry, a bez tego założenia, kiedy $\lim_{x \rightarrow \beta} [\xi(x) - h(x)] = 0$, a w szczególności, gdy h jest funkcją identycznościową, a intensywność awarii rozkładu brzegowego ucieka do nieskończoności.

Rozdział 3 zawiera uogólnienie wyników rozdziału 2 na przypadek rozkładów ciągłych (niekoniecznie absolutnie). Wówczas równania różniczkowe (4) zastępuje się analogicznymi równaniami całkowymi

$$\varphi_i(x) - \varphi_i(\alpha) = \frac{\gamma_{r+i+1}}{\gamma_{r+i}} \int_{\alpha}^x \frac{\varphi_i(t) - \varphi_{i+1}(t)}{\xi(t) - \varphi_1(t)} d\xi(x), \quad i = 1, \dots, \ell - 1.$$

Istotnym osiągnięciem rozdziału 3 (i pracy [15]) jest zastąpienie zbieżności $\xi(x) - h(x)$ do zera na prawym końcu nośnika warunkiem $\lim_{x \rightarrow \beta} [\xi(x) - ah(x) - b] = 0$ dla pewnych $a \geq 1$ i $b \in \mathbb{R}$. Nowe przykłady charakteryzacji (oprócz prostego przykładu nieciągłej gęstości) oparte były na asymptotycznej liniowej zależności ξ i h . Można było istotnie skrócić pracę wycinając rozdział 2, a jego wyniki podać jako wnioski z analogicznych twierdzeń rozdziału 3.

Rozdział 4 dotyczy słabych rekordów o wartościach dyskretnych całkowitych. Zależności regresyjne są wtedy opisane równaniami różnicowymi

$$\varphi_i(j+1) - \varphi_i(j) = \frac{\varphi_i(j) - \varphi_{i+1}(j)}{\xi(j) - \varphi_1(j)} [\xi(j+1) - \xi(j)], \quad i = 1, \dots, \ell - 1.$$

W przypadku $\ell = 2$ ich rozwiązania jednoznacznie charakteryzują rozkład, gdy h jest funkcją ograniczoną lub asymptotycznie liniową względem ξ (por. rezultaty

rozdziału 3 i ich dowody). Sformułowano też podobne twierdzenie charakterystyczne dla zwykłych rekordów. W przypadku $\ell = 2$ warunki charakteryzujące ogon rozkładu wyglądają następująco

$$\varphi(j+1) - \varphi(j) = \frac{\varphi(j+1) - h(j+1)}{\xi(j+1) - \varphi(j+1)} [\xi(j+1) - \xi(j)], \quad i = 1, \dots, \ell - 1,$$

oraz $h(j+2) < \varphi(j+1) < \xi(j)$ dla $j \geq r$. Jak można się domyślić, przesunięcia argumentów zwiększają trudność problemu.

Praca doktorska zawiera nowe ogólne podejście do zagadnień charakteryzacyjnych. Jej słabszym punktem jest jednak fakt, że nie istnieją narzędzia analityczne, które pozwoliłyby w pełni wykorzystać nową metodę. Nie są dostępne techniki rozwiązywania układów równań różniczkowych i różnicowych, gdy podane są jedynie warunki uporządkowania rozwiązań. Z tego powodu ta nowa metoda nie rozszerza istotnie zakresu w pełni rozwiązanych zadań charakteryzacji. Najwięcej można jeszcze uzyskać w przypadku regresji słabych rekordów o odstępach 2 pochodzących z rozkładów o skończonych nośnikach. Wtedy opisany w podrozdziale 4.4 algorytm pozwala całkowicie rozwiązać numerycznie problem charakteryzacji dla dowolnych funkcji h i ξ . W pewnych zadaniach (np. dla uogólnionych statystyk pozycyjnych z rozkładów normalnych i Gompertza czy dla rekordów z rozkładów Poissona i ujemnego dwumianowego) daje się pokazać jednoznaczność charakteryzacji, ale nie można analitycznie opisać funkcji regresji nawet samej nieprzekształconej statystyki $X_*^{(r+2)}$ czy W_{r+2} . Niektóre przykłady były konstruowane "od tyłu": dla danego rozkładu liczone regresję i pokazywano jednoznaczność charakteryzacji. Szczególnie spodobały mi się przykłady 3.4 i 3.5. W pierwszym wykorzystano najsłabsze założenia gwarantujące jednoznaczność charakteryzacji z twierdzenia 3.5. W drugim podano przykład rozkładu wykraczającego poza zakres jego działania. Dodam jeszcze, że metody rozprawy pozwalają jedynie odzyskać liniowe charakteryzacje regresyjne, znane wcześniej dla dowolnego odstępów, jedynie w przypadku $\ell = 2$. Mimo wszystko mam nadzieję, że w najbliższych latach lub stuleciach nastąpi postęp w analizie, szczególnie teorii równań różniczkowych, który pozwoli w pełni wykorzystać rezultaty pani Maciąg i jej promotora.

Rozprawa została napisana klarownie, a starannie dobrane przykłady dobrze ilustrują teorię. Najbardziej przeszkadza mi w niej nieprecyzyjne zastosowanie twierdzenia Stolza-Cesàro w rozdziale 4. Głosi ono, że jeśli iloraz $\frac{b_n - b_{n-1}}{a_n - a_{n-1}}$ ma granicę (być może nieskończoną), to tę samą granicę ma ciąg $\frac{b_n}{a_n}$. Istotnym założeniem jest tu monotoniczność i zbieżność do nieskończoności ciągu a_n . Na str. 78 twierdzenie zostało zastosowane do ciągów q_k i $\sum_{i>k} q_i$, ale rozbieżność tego ostatniego

nie została zweryfikowana. Na str. 72 wnioskowano z niego zbieżność $\frac{\Delta y(j)}{\Delta \xi(j)}$ mimo, że implikacja przeciwna w twierdzeniu Stolza $\lim \frac{b_n}{a_n} = c \Rightarrow \lim \frac{b_n - b_{n-1}}{a_n - a_{n-1}} = c$ nie jest w ogólności prawdziwa. Mam nadzieję, że moje wątpliwości zostaną rozwiązane w trakcie obrony. Poza tym zauważyłem następujące niewielkie usterki.

Str. 9, 10: w definicjach 1.4 i 1.5 należy podkreślić, że $U_*^{(1)}, \dots, U_*^{(n)}$ są jednostajnymi uogólnionymi statystykami pozycyjnymi z parametrami $\gamma_1, \dots, \gamma_n$.

Str. 11: w definicji sekwencyjnych statystyk pozycyjnych, działające elementy nie muszą być zastąpione przez nowe, z innymi intensywnościami awarii, ale ich intensywności awarii mogą się zmienić, ponieważ ich obciążenie zwiększa się w związku z ubytkiem działającego elementu.

Str. 28: w twierdzeniu 2.1 (również 2.2 i 2.3) brak jest założenia, że gęstość jest ciągła (por. lemat 1.3(iii)).

Str. 31 (tak samo str. 50): monotoniczność y nie zależy tylko od warunków $h < y < \xi$, ale też od równości z (2.22) (odpowiednio (3.24)).

Str. 39: w rozważaniach dotyczących parametrów a, b, c, d należy też wziąć pod uwagę β . Jeżeli $a < 1$, to w świetle $d < b$ zachodzi $\beta = \frac{b}{1-a} < \infty$. W przeciwnym przypadku $\beta = +\infty$.

Str. 40, l. 3: przy założeniu $a \neq 1$ nachylenie c leży między a i 1 (różnie uporządkowanych).

Str. 51: Warunek Lipschitza dla K wymaga dokładniejszego uzasadnienia, bo y_1 i y_2 występują w mianowniku stałej.

Str. 54, l. -3: Jeśli $C = b - d$, to $\Psi(x) \rightarrow \frac{b}{0}$. Nie zauważyłem odwołania do konwencji, że $\frac{b}{0} = +\infty$.

Str. 80, l. 1: czy jest sens zakładać, że $N = \frac{b}{1-a}$ musi być liczbą naturalną?

Str. 82, l. 5: należy odnieść się do wzoru (1.5), a nie (1.4).

Powyższe drobne usterki nie wskazują na brak kompetencji Kandydatki, ale świadczą jedynie o tym, że recenzent starał się dokładnie przeczytać pracę. Wnioskuje o dopuszczenie pani mgr Krystyny Maciąg do dalszych etapów przewodu doktorskiego.

Tomasz K. K.