

Warszawa, 3 września 2019 r.

dr hab. Anna Dembińska
 Wydział Matematyki i Nauk Informacyjnych
 Politechnika Warszawska

RECENZJA

rozprawy doktorskiej magister Krystyny Maciąg

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

Probabilistyczne problemy charakteryzacyjne dotyczą identyfikacji rozkładu prawdopodobieństwa na podstawie pewnych własności tego rozkładu. Za szczególnie wartościowe uznawane są charakteryzacje, w których prosto formułujące się własności prowadzą do identyfikacji jednej klasy rozkładów posiadających te własności. Takie charakteryzacje są przydatne podczas modelowania statystycznego, w teorii procesów stochastycznych czy w statystyce matematycznej do budowania testów zgodności. Mogą być także użyte do dowodów wyników związanych z rozkładami prawdopodobieństwa - np. opierając się na charakteryzacji rozkładu normalnego, można udowodnić centralne twierdzenie graniczne.

Z drugiej strony problemy charakteryzacyjne często prowadzą do ciekawej, eleganckiej i nietrywialnej matematyki. Nierzadko wymagają licznych klasycznych narzędzi analizy matematycznej, jak równania różniczkowe czy różnicowe, teoria funkcji zmiennych zespolonych czy teoria równań funkcyjnych. Ze względu na tę swoją matematyczną atrakcyjność przyciągały największe umysły - zajmowali się nimi, między innymi, P. Lévy, S. Bernstein, J.V. Linnik, H. Cramer, G. Polya, C.R. Rao czy J. Prochorow.

Obszerna literatura charakteryzacyjna zawiera także liczne wyniki wykorzystujące własności statystyk porządkowych, rekordowych i uogólnionych statystyk porządkowych. Wśród nich są twierdzenia oparte na zagadnieniach regresyjnych. Prezentowana rozprawa doktorska wpisuje się właśnie w ten nurt badań. Rozwija i uzupełnia znane w literaturze wyniki dotyczące charakteryzacji rozkładów prawdopodobieństwa opartych na własnościach regresyjnych uogólnionych statystyk porządkowych i słabych wartości rekordowych. Dokładniej, zajmuje się odpowiedzią na pytanie czy funkcja regresji dana jednym z poniższych wzorów

$$\xi(x) = E(h(X_{\star}^{(r+\ell)} | X_{\star}^{(r)} = x)), \quad (R1)$$

lub

$$\xi(x) = E(h(W_{r+\ell} | W_r = x)), \quad (R2)$$

gdzie $r \geq 1$ i $\ell \geq 1$ są ustalone, h to pewna ściśle rosnąca funkcja, $X_{\star}^{(r)}$, $r \geq 1$, to uogólnione statystyki porządkowe z rozkładu o dystrybuancie F , zaś $(W_r, r \geq 1)$ to ciąg słabych rekordów odpowiadający ciągowi niezależnych zmiennych losowych o dystrybuancie F , jednoznacznie określa rozkład wyjściowych zmiennych losowych.

W rozdziale 1 krótko przypomniano podstawowe fakty dotyczące teorii uogólnionych statystyk porządkowych i dyskretnych słabych i zwykłych rekordów. Tam też wyprowadzono wniosek opisujący rekurencyjną strukturę regresji danych wzorami (R1) i (R2). Ów wniosek jest punktem wyjścia do rozważań przeprowadzanych w kolejnych rozdziałach. Te kolejne rozdziały, o numerach 2-4, powstały na podstawie 3 wspólnych, już opublikowanych artykułów Doktorantki i jej Promotora, dr hab. Mariusza Bieńka, i zawierają główne wyniki rozprawy.

W rozdziale 2 skoncentrowano się na charakteryzacjach rozkładów z ciągłą gęstością poprzez regresje uogólnionych statystyk porządkowych postaci (R1). Podano ogólne warunki gwarantujące że regresja (R1) jednoznacznie wyznacza wyjściowy rozkład. Warunki te wymagają, między innymi, by pewien układ równań różniczkowych pierwszego rzędu miał dokładnie jedno rozwiązanie spełniające odpowiednie ograniczenia brzegowe. W przypadku regresji o odstępzie dwa, czyli dla $\ell = 2$, udało się ten wynik wykorzystać do wyprowadzenia charakteryzacji z pewnymi szczególnymi funkcjami h i ξ . W ten sposób, obok nowych dowodów znanych już wcześniej twierdzeń charakteryzacyjnych, uzyskano także nowe charakteryzacje.

Analogiczne rozważania dotyczące rozkładów ciągłych przeprowadzono w rozdziale 3. Teraz, zamiast układu równań różniczkowych pierwszego rzędu, otrzymano układ równań całkowych. Podobnie jak w rozdziale 2, używając tego wyniku, podano nowe dowody charakteryzacji znanych w literaturze a także przykłady nowych charakteryzacji rozkładów ciągłych.

Ostatni, czwarty rozdział poświęcony jest charakteryzacji dyskretnych rozkładów prawdopodobieństwa przez regresje słabych rekordów postaci (R2). Główny wynik tego rozdziału to twierdzenie podające warunki dostateczne i konieczne na to aby regresja (R2) jednoznacznie wyznaczała rozkład dyskretny. Jako zastosowanie tego wyniku rozpatrzono przypadek z $\ell = 2$ i wyprowadzono nowe charakteryzacje rozkładów dyskretnych jak i podano nowe dowody znanych już wcześniej charakteryzacji.

Niestety praca zawiera pewne nieścisłości i usterki. Poniżej przedstawiam listę najistotniejszych z nich.

1. **Twierdzenie 4.1 nie jest prawdziwe w takiej wersji, w jakiej zostało podane.**

Można pokazać (por. przykład 4.1 (a)), że gdy $S = \{0, 1, 2\}$, $h(j) = j, j \in S$ oraz $\xi(0) = 5/4$, $\xi(1) = 7/4$ i $\xi(2) = 2$, to funkcja regresji

$$\xi(j) = E(W_{r+2}|W_r = j), \quad j \in S, \quad (R)$$

gdzie $r \geq 1$ jest ustalone, jednoznacznie określa wyjściowy rozkład. Pomimo to równanie różnicowe (4.12) ma więcej niż jedno rozwiązanie $y = \varphi$ spełniające warunek (4.13). Takimi rozwiązaniami są np.

$$\left\{ \begin{array}{l} \varphi(0) = 3/4 \\ \varphi(1) = 3/2 \\ \varphi(2) = 2 \end{array} \right. \quad \text{ i } \quad \left\{ \begin{array}{l} \varphi(0) = 5/8 \\ \varphi(1) = 9/8 \\ \varphi(2) = 47/40 \end{array} \right. .$$

Błąd w dowodzie tkwi w następującym miejscu.

Każde rozwiązanie problemu różnicowego (4.12), które spełnia warunki (4.13) oraz (4.14) definiuje właściwą funkcję rozkładu prawdopodobieństwa $p_k, k \in S$, określoną

przez (4.20). Jednak skąd pewność, że dla każdej z takich funkcji rozkładu prawdopodobieństwa zachodzi warunek (R)? Rozumowanie przeprowadzone w dowodzie twierdzenia 4.1 pokazuje, że warunek (R) pociąga za sobą (4.12), (4.13) i (4.14). Natomiast wcale nie wykazuje równoważności, tzn. tego, że (4.12), (4.13) i (4.14) pociągają za sobą (R).

W szczególności (4.16) nie jest równoważne pierwszej linijce (4.18). Aby uzyskać równoważność można np. pierwszą linijkę (4.18) uzupełnić o

$$\begin{cases} \varphi(N) = h(N) & \text{gdy } N < \infty \\ \lim_{j \rightarrow \infty} \varphi(j)q_j = 0 & \text{gdy } N = \infty \text{ i } \{h(j)\} \text{ nie jest zbieżny.} \end{cases} \quad (\text{pW})$$

Reasumując, twierdzenie 4.1 stanie się prawdziwe jeśli równoważność zastąpimy w nim wynikaniem z dołu w górę; innymi słowy jeśli *wtedy i tylko wtedy, gdy* zastąpimy przez *jeśli*. Ponadto do warunku (4.13) warto dodać warunek

$$\varphi(N) = h(N) \text{ gdy } N < \infty, \quad (\text{W})$$

bo wtedy otrzymamy mocniejszą wersję twierdzenia która się przyda w dowodzie twierdzenia 4.3 oraz w przykładzie 4.1.

Oczywiście można się starać napisać wersję twierdzenia 4.1 i jego poprawnego dowodu w ogólniejszej wersji z *wtedy i tylko wtedy, gdy*. Wymagać to będzie dokładnego przeanalizowania (i uzasadnienia) wynikania (R) z (4.12)-(4.14) i (pW), ewentualnie dodania kolejnych warunków owe wynikanie gwarantujących. Jednak takie starania nie wydają mi się kluczowe - twierdzenie 4.1 okrojone do wersji, w której *wtedy i tylko wtedy, gdy* jest zastąpione przez *jeśli* i do której jest dodany warunek (W) w zupełności wystarcza do rozważań w podrozdziałach 4.3-4.5.

Z powodów przedstawionych powyżej, także twierdzenie 4.2 nie jest prawdziwe w wersji, w której zostało podane. Analogiczne uwagi dotyczą też twierdzeń 2.1, 2.2, 3.1, 3.2 i 4.6.

2. Dowód twierdzenia 4.3 nie działa w sytuacji, gdy $N < \infty$. W tej sytuacji bowiem nie możemy rozważać granicy $\lim_{j \rightarrow N} z(j)$. Ponadto wtedy fakty, że z jest funkcją ściśle rosnącą na $\bar{S}$ i $0 \leq z(j) \leq (\xi(j) - h(j))^2$, $j \in \bar{S}$, nie prowadzą do sprzeczności.

Nieścisłość tą daje się jednak szybko naprawić korzystając z poprawionej wersji twierdzenia 4.1 (wersji uzupełnionej o warunek (W)). Skoro musimy mieć $\varphi(N) = y(N) = h(N)$, to z (4.31) otrzymamy $\varphi(N-1) = y(N-1)$ i dalej $\varphi(N-2) = y(N-2)$, ..., $\varphi(0) = y(0)$ czyli $\varphi = y$.

3. W uwadze 4.5 jest założone, że $\lim_{j \rightarrow \infty} \lambda(j) = \infty$. Jednak założenie to nigdy nie jest spełnione, bo w przypadku rozkładów dyskretnych mamy

$$\lambda(j) = \frac{p_j}{q_j} = \frac{P(X = j)}{P(X \geq j)} \leq 1.$$

4. Rozdział 4.4:

Jednym z założeń twierdzenia 4.3 jest to, że funkcja ξ jest funkcją regresji, tzn. spełnia

$$\xi(j) = E(W_{r+2} | W_r = j), \quad j \in S, \quad (\text{R})$$

dla pewnego ustalonego $r \geq 1$ i pewnej dystrybuanty F .

W przykładzie 4.1 jest używane twierdzenie 4.3, ale nie zostało wyjaśnione skąd wiadomo, że powyższe założenie jest spełnione, tzn. skąd wiadomo, że funkcja ξ zdefiniowana w (4.41) oraz funkcja ξ z przykładu 4.1 (b) są takie, że zachodzi (R).

Analogiczne pytanie dotyczy funkcji ξ z trzeciej linii strony 77.

5. W przykładzie 2.2 i uwadze 2.4 należałoby jeszcze uzasadnić, że odpowiednie funkcje regresji w ogóle istnieją (uzasadnić, że $E|R_{r+2}| < \infty$).
6. Jeśli $(X_n, n \geq 1)$ jest ciągiem niezależnych zmiennych losowych o tym samym rozkładzie dyskretnym o nośniku $\{0, 1, 2, \dots\}$ zaś $(R_n, n \geq 1)$ - odpowiadającym mu ciągiem (zwykłych) wartości rekordowych, to nośnikiem R_r jest $\{r-1, r, \dots\}$ a nie $\{r, r+1, \dots\}$. Pociąga to za sobą konieczność zmian w (1.4), lemacie 1.7 i dalej w podrozdziale 4.6.

Drobniejsze uwagi:

1. Str. 10, linia 14: $F^{-1}(0)$ nie zostało zdefiniowane.
2. Str. 13, linia 8: $F(x) = x$ dla jakich x ?
3. Str. 13, wzór (1.8): brakuje warunku $x < \beta$.
4. Str. 18, linia 10 i dalej w pracy: rozważając dowolną funkcję h (bez warunku, że jest ona ściśle rosnąca) i $E|h(Y_{r+1})|$, trzeba założyć, że h jest mierzalna.
5. Str. 21, lemat 1.2 (iv): Jeśli F ma nośnik (α, β) , to F nie może być stała na przedziale $I \subset (\alpha, \beta)$.
6. Str. 23, linie 11 i 12: zamiast Y powinny być X .
7. Str. 31, linia 14: zamiast y powinno być $y(x)$
8. Str. 31, linia 18: należy wyjaśnić co rozumiemy przez $h(\beta)$.
9. Str. 31, linia 19: czy tu zakładamy, że $\beta = \infty$?
10. Str. 35, linia 1: jeśli $h(x) = x$, to wcale nie musimy mieć $h(\beta) := \lim_{x \rightarrow \beta^-} h(x) = \infty$. W szczególności jeśli $\beta < \infty$, to $h(\beta) = \beta < \infty$.
11. Str. 35, lemat 2.2: do założeń trzeba dodać $\beta = \infty$.
12. Str. 35, wniosek 2.1: brakuje założenia, że F to dystrybuenta rozkładu absolutnie ciągłego z ciągłą gęstością.
13. Str. 38, linia 2 od dołu: dlaczego bez straty ogólności możemy założyć, że $\alpha = 0$ (czyli dlaczego z góry możemy wykluczyć przypadki z $\alpha = -\infty$)?
14. Str. 54, linia 5 i 18: podane nierówności zachodzą jedynie dla dostatecznie dużych x (ale to do dowodu wystarcza). Analogiczny komentarz dotyczy nierówności ze strony 72 podanych w liniach 5 i 16.

15. Str. 66, linia 5: $\xi > \varphi > h$ trzeba zastąpić przez $\xi(j) > \varphi(j) > h(j)$, $j \in \bar{S}$.
16. Str. 70, linie 8 i 9: $w(0) < w < 1$ trzeba zastąpić przez $w(0) \leq w < 1$.
17. Str. 71, twierdzenie 4.3: założenie $E|h(W_{r+2})| < \infty$ jest tu zbędne, bo wynika ono z pozostałych.

Podsumowując, rozprawa zawiera zarówno nowe dowody twierdzeń znanych już wcześniej w literaturze jak i nowe wyniki. Prezentowane rezultaty otrzymano korzystając z nowatorskiej metody polegającej na sprowadzaniu danego problemu charakteryzacyjnego do problemu istnienia rozwiązań pewnego układu równań różniczkowych, całkowych lub różnicowych pierwszego rzędu. Dla paru klas problemów regresyjnych z odstępem równym dwa, zaproponowaną metodę udało się wykorzystać do uzyskania twierdzeń charakteryzacyjnych. Zarówno szczegółowe opracowanie samej metody, jak i jej użycie w dowodach charakteryzacji rozkładów prawdopodobieństwa, wymagało od Autorki dobrej znajomości licznych narzędzi matematycznych, teorii uogólnionych statystyk porządkowych i dyskretnych rekordów a także twórczego myślenia. Niestety rozprawa zawiera pewne błędy i nieścisłości. W szczególności twierdzenia 4.1 i 4.2, podawane przez Autorkę jako jedne z głównych wyników pracy, nie są prawdziwe w wersji, w której zostały zaprezentowane. Pozostałe główne wyniki pracy tzn. twierdzenia 2.1, 2.2, 3.1 i 3.2 zawierają luki w dowodach. i stąd brak mojej pewności, że są prawdziwe. W tej sytuacji **uważam, że, przed dopuszczeniem Pani magister Krystyny Maciąg do dalszych etapów przewodu doktorskiego, rozprawa powinna być poprawiona z uwzględnieniem uwag przedstawionych w recenzji.**

Anno Dembinska