

Toruń, 24 lipca 2023 r.

dr hab. Agnieszka Goroncy, prof. UMK  
Katedra Statystyki Matematycznej i Eksploracji Danych  
Wydział Matematyki i Informatyki  
Uniwersytet Mikołaja Kopernika w Toruniu

**Recenzja rozprawy doktorskiej mgr Luizy Pańczyk**  
**pt. „Wybór optymalnych  $L$ -statystyk w estymacji kwantyli”**

Rozprawa doktorska pani mgr Luizy Pańczyk składa się z sześciu rozdziałów. Wyniki zawarte w rozprawie zostały opisane w dwóch współautorskich artykułach naukowych. Jeden z artykułów jest opublikowany w bardzo dobrym (stupunktowym, zgodnie z aktualnym ministerialnym wykazem czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych) czasopiśmie wydawnictwa Springer z dziedziny statystyki matematycznej, natomiast drugi został przesłany do recenzji.

Po zapoznaniu się z rozprawą doktorską pani mgr Luizy Pańczyk stwierdzam, że spełnia ona wymagania stawiane rozprawom doktorskim. Dalsza część tej recenzji zawiera omówienie zawartości rozprawy oraz uzasadnienie tej oceny.

Recenzowana rozprawa, zgodnie z jej tytułem, dotyczy wyboru optymalnych liniowych kombinacji statystyk pozycyjnych (tzw.  $L$ -statystyk) w zagadnieniu estymacji kwantyla rozkładu zadanego rzędu  $p \in (0, 1)$ . Została ona podzielona na dwie główne części, obejmujące rozdziały od trzeciego do szóstego. Pierwsza część dotyczy wyboru pojedynczej statystyki pozycyjnej, natomiast druga wyboru kombinacji liniowej dwóch statystyk pozycyjnych jako optymalnego estymatora kwantyla nieznanego rozkładu, z którego pobrana została prosta próba losowa  $X_1, \dots, X_n$  o ustalonej liczbie  $n$ . Jak podkreśla Autorka rozprawy, jej podejście do estymacji kwantyli nie jest asymptotyczne i może być stosowane również dla małych prób, co jest dużą zaletą jeżeli chodzi o zastosowanie uzyskanych wyników w praktyce. Drugą ogromną zaletą otrzymanych w rozprawie wyników jest to, że jej Autorka nie ogranicza się do rozkładów z ciągłymi i ściśle rosnącymi dystrybuantami, tzn. obejmują również sytuacje, gdy kwantyl zadanego rzędu nie jest wyznaczony jednoznacznie. W rozprawie rozważana jest klasa  $\mathcal{F}_2$  dystrybuant  $F$  o skończonej wariancji, przedstawionej za pomocą dolnej (lub równoważnie górnej) funkcji kwantylowej.

W rozdziale pierwszym zatytułowanym „Pojęcia wstępne” zaprezentowano podstawowe pojęcia związane z zagadnieniem estymacji kwantyli, w szczególności zdefiniowano pojęcie kwantyla ustalonego rzędu  $p$ , dolną oraz górną funkcję kwantylową, przywołano pojęcie statystyki porządkowej oraz jej rozkładu definiowanego przy użyciu wielomianów Bernsteina, wprowadzono notację dla szczególnie interesujących podzbiorów zbioru  $C_n$ , który jest zbiorem wektorów współczynników  $\mathbf{c} = (c_1, \dots, c_n)$ , odpowiadających  $L$ -statystykom, o nieujemnych współrzędnych sumujących się do jedności. Autorka

przypada również popularne i znane w literaturze estymatory kwantyli oparte na statystykach pozycyjnych i ich kombinacjach. Problem estymacji kwantyli jest od dawna omawiany w literaturze naukowej. W klasycznych rozważaniach estymator kwantyla zadanego rzędu  $p \in (0, 1)$  oparty jest na odwrotności dystrybuanty empirycznej  $F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{(-\infty, x]}(X_i)$ , która jest prawostronnie ciągła, dla ustalonej próby jest funkcją schodkową i bazuje, zgodnie z intuicją, na statystykach porządkowych z indeksami bliskimi wartości  $np$ . Wobec wielu propozycji estymatorów kwantyli znanych w literaturze, zasadnym wydaje się pytanie, który z nich powinien zostać użyty w praktyce. Przedstawiona tu rozprawa jest odpowiedzią na to ważne pytanie.

Drugi rozdział wydaje się kluczowy dla dalszych rozważań i wyników otrzymanych dla estymatorów opartych na statystykach porządkowych. Wprowadza on bowiem nowe kryterium optymalności pozwalające jednoznacznie wskazać  $L$ -statystykę, a dokładniej wektor  $\mathbf{c}$  z nią powiązany tak, aby była ona najlepszym estymatorem kwantyla zadanego rzędu, wyznaczonym na podstawie próby losowej ustalonego rozmiaru. Kryterium, o którym mowa jest oparte na optymalnych górnych i dolnych oszacowaniach obciążenia estymatora  $L$ -statystyki dla ustalonego wektora współczynników  $\mathbf{c}$ , jako estymatora dolnej i górnej funkcji kwantylowej. Jak się okazuje, oszacowania te jedynie dla wektorów  $\mathbf{c}$  spełniających pewne własności mogą być wyrażone w sposób jawny, między innymi wówczas, gdy tylko jeden współczynnik  $c_j = 1$ ,  $j = 1, \dots, n$ , rozważanej  $L$ -statystyki jest niezerowy. Sytuacja ta odpowiada obciążeniu pojedynczej statystyki pozycyjnej  $X_j$  jako estymatora kwantyla. W dalszej części rozdziału drugiego znalazły się lematy opisujące własności oszacowań górnych i dolnych obciążeń estymatorów opartych na  $L$ -statystykach i pojedynczych statystykach porządkowych oraz uzasadnienie wprowadzenia nowego kryterium. Kryterium optymalności według mnie słusznie opiera się na dwóch czynnikach: pierwszym z nich jest miara odpowiadająca stopniowi symetrii oszacowań obciążenia estymatora względem zera, drugim zaś jest czynnik normalizujący, który przeciwdziała przeszacowaniu (i niedoszacowaniu) kwantyla zadanego rzędu, bez względu na rozkład próby losowej. Dalej, Autorka rozprawy uzasadnia własność symetrii oszacowań obciążenia statystyki pozycyjnej  $X_{j:n}$  jako estymatora rozważanego kwantyla, a także udowadnia pewne istotne własności funkcji pomocniczych. Dowody lematów przedstawionych w rozdziale drugim nie są przesadnie skomplikowane, wiele wyników jest rezultatem wcześniej udowodnionych tez.

W kolejnym, trzecim rozdziale rozprawy pani mgr Luiza Pańczyk przedstawia jawne wzory na górne i dolne oszacowania obciążenia pojedynczych statystyk pozycyjnych  $X_j$ ,  $j = 1, \dots, n$  jako estymatorów funkcji kwantylowych. Ma to na celu wybór optymalnej statystyki porządkowej  $X_j$ . Wymaga to dość skrupulatnej analizy angażującej dwa punkty styczności ( $\theta_j$  i  $\xi_j$ ), będące rozwiązaniami odpowiednich równań. Co więcej, Autorka otrzymała poprawną wersję stwierdzenia z pracy Okolewskiego i Rychlika [22]. Główne twierdzenie (3.1) tego rozdziału, umożliwia wyznaczenie górnych i dolnych oszacowań obciążenia pojedynczej statystyki pozycyjnej jako estymatora kwantyla zadanego rzędu oraz własności tych oszacowań. Dowód, którego idea jest oparta na pracy Okolewskiego i Rychlika [22], wymagał od Autorki dokładnej i żmudnej analizy mono-



toniczności i wklęsłości dystrybuanty  $F_{j:n}$  statystyki pozycyjnej  $X_{j:n}$ , dla  $j = 1, \dots, n$ , w celu zastosowania tzw. nierówności Morigutiego. Dokładniej, celem było wyznaczenie prawostronnej pochodnej największej wypukłej minoranty odpowiedniej funkcji powiązanej z  $F_{j:n}$ . Otrzymano również wzory na pewne funkcje pomocnicze, umożliwiające wybór optymalnej statystyki pozycyjnej. Ważnym spostrzeżeniem Autorki jest fakt, że w przypadku rozważań dotyczących oszacowań obciążenia ogólnych  $L$ -statystyk, postępując analogicznie jak w przypadku pojedynczej statystyki porządkowej, nie da się efektywnie wyznaczyć przedziałów wypukłości i wklęsłości odpowiednich funkcji  $F_c$ , dla dowolnego  $c \in \mathcal{C}_n$ . Jest to możliwe jedynie dla wektorów z pewnego szczególnego podzbioru  $\mathcal{C}_n$ , który nakłada odpowiednie warunki na monotoniczność ich współrzędnych. Do analizy przebiegu zmienności odpowiedniej funkcji  $f_c$ , będącej kombinacją liniową gęstości standardowych jednostajnych statystyk pozycyjnych Autorka rozprawy zastosowała przyjazne narzędzie, zwane własnością zmniejszającą zmienność wielomianów Bernsteina (VDP). Niestety, znajomość wypukłości funkcji  $F_c$  nie umożliwia wyznaczenia oszacowania obciążenia rozważanych  $L$ -statystyk z uwagi na niemożność ustalenia położenia zadanego rzędu kwantyla  $p$  względem pomocniczej liczby  $\theta_c$  (będącej odpowiednikiem  $\theta_j$  z rozważań dotyczących pojedynczej statystyki  $X_{j:n}$ ). Okazuje się, że jest to możliwe jednak w szczególnym przypadku, gdy  $L$ -statystyka jest kombinacją wypukłą dwóch sąsiadujących statystyk pozycyjnych.

Kolejnym krokiem poczynionym w rozprawie jest wybór optymalnej w sensie wprowadzonego kryterium pojedynczej statystyki pozycyjnej jako estymatora kwantyla. Procedura ta opisana jest w rozdziale czwartym, w którym pani mgr Luiza Pańczyk pokazuje istnienie jednoznacznie wyznaczonych liczb  $a_{j,n}$ ,  $1 \leq j < n$ , spełniających pewne określone warunki, które w zależności od tego czy rząd kwantyla  $p$  należy do przedziału  $(a_{j-1,n}, a_{j,n})$ , czy też nie, umożliwiają wskazanie optymalnej jednej statystyki pozycyjnej  $X_{j:n}$  bądź dwóch, równie dobrych statystyk  $X_{j:n}$  i  $X_{j+1:n}$ . W szczególności, Autorka wskazuje najlepszą statystykę pozycyjną będącą estymatorem kwantyla rzędu  $\frac{k}{n}$ , której rząd zależy od tego, czy wartość  $k$  jest mniejsza od połowy liczebności próby (wówczas jest to  $X_{j:n}$ ), czy też większa (wówczas jest to  $X_{j+1:n}$ ). Wynik ten jest nieco inny niż zazwyczaj podawany w literaturze, jednak jego zaletą jest własność symetrii ( $X_{n:n}$  jest estymatorem kwantyla rzędu  $\frac{n-1}{n}$ , zaś  $X_1$  jest estymatorem kwantyla rzędu  $\frac{1}{n}$ ), która jest istotna w problemie estymacji kwantyli. Wyniki uzyskane w tym rozdziale zastosowane do estymacji mediany rozkładu są zgodne z wynikiem klasycznym w przypadku próby o nieparzystej liczebności  $n$ , kiedy to estymatorem mediany jest statystyka porządkowa rzędu  $\frac{n+1}{2}$ . W przypadku, gdy próba jest parzysta, otrzymano dwa równie dobre estymatory odpowiadające pojedynczym statystykom pozycyjnym rzędów  $\frac{n}{2}$  oraz  $\frac{n}{2} + 1$ , co oczywiście nie jest równoważne z klasycznym estymatorem, który jest ich średnią arytmetyczną. Znaczny nakład pracy został włożony do wyznaczenia liczb  $a_{j,n}$ , przy użyciu pomocniczych liczb  $p_j, q_j$ , których własności zbadano w podrozdziale 4.2. Wyniki tu uzyskane są raczej techniczne, wymagały dużej sprawności obliczeniowej Autorki rozprawy i w efekcie umożliwiły rozwiązanie problemu optymalnego wyboru pojedynczej statystyki pozycyjnej jako estymatora kwantyla zadanego rzędu. W dalszej części Autorka rozprawy rozważa równoważne kryterium, które polega na minimalizacji mak-

symalnego oszacowania obciążenia pojedynczej statystyki pozycyjnej. Jak się okazuje, dla większości wartości rzędu kwantyla  $p$  oba kryteria zwracają statystykę pozycyjną o tym samym indeksie, co nie jest dla Autorki zaskoczeniem, ponieważ oba kryteria optymalności równoważą górne i dolne oszacowania. Rozważania dotyczące optymalnego wyboru pojedynczej statystyki porządkowej jako estymatora kwantyla zadanego rzędu podsumowane zostało obliczeniami numerycznymi, które umożliwiają wyznaczenie przybliżonych wartości liczb pomocniczych, trudnych do wyznaczenia w sposób jawny.

Druga część pracy dotyczy optymalnego wyboru estymatora kwantyla spośród wypukłych kombinacji liniowych dwóch sąsiadujących statystyk pozycyjnych. Z wcześniejszych rozważań wiadomo już, że jest to jedyna postać  $L$ -statystyki angażująca więcej niż jedną statystykę porządkową, która daje możliwość zastosowania nowego kryterium optymalności. Analogicznie jak w przypadku pojedynczej  $X_{j:n}$ , zadanie polega na wyznaczeniu wartości liczb  $b_{j,n}$  (odpowiedników  $a_{j,n}$ ) i sprawdzeniu, czy rząd kwantyla należy do odpowiedniego przedziału wyznaczonego przez te liczby. Współczynniki kombinacji liniowej statystyk pozycyjnych są wyznaczone jednoznacznie. Wyniki, o których mowa, otrzymane dla mediany (również dla próby o parzystej liczebności) stanowią uogólnienie wyników klasycznych. Ważnym spostrzeżeniem jest to, że uzyskany w rozdziale piątym estymator nie jest interpolacją liniową funkcji kwantylowej, która jest rozważana w wielu dotychczasowych pracach. Analogicznie jak w pierwszej części, ilustracją uzyskanych wyników teoretycznych są obliczenia numeryczne.

Ostatni rozdział pracy dotyczy analizy błędu średniokwadratowego (MSE) otrzymanych estymatorów. Jako, że na MSE składa się zarówno obciążenie estymatora (a dokładniej kwadrat obciążenia), jak i wariancja, słusznym jest skupienie się również na minimalizacji wariancji estymatorów, jako, że dotychczas Autorka zajmowała się problemem minimalizacji obciążenia odpowiedniej  $L$ -statystyki. W tym celu posłużono się wynikami Kozyry [18], dotyczącymi oszacowań wariancji  $L$ -statystyk. Ta część pracy opiera się w znacznej mierze na obliczeniach numerycznych, niektórych przypuszczeń Autorki nie udało się uzasadnić teoretycznie z uwagi na skomplikowane postaci wzorów rozważanych funkcji i wyrażeń. Faktycznie, wartości, które są potrzebne do wyznaczenia estymatorów nie zawsze mogą być przedstawione w jawnej postaci i słusznym posunięciem jest wykonanie obliczeń numerycznych, które w obecnych czasach są raczej łatwe do realizacji. Dzięki temu uzyskano wiedzę o przeciętnym zachowaniu estymatorów, które potwierdza prawidłowość procesu wyznaczania estymatorów kwantyli. Generalnie, otrzymane przez Autorkę estymatory okazały się dobre, a w wielu przypadkach najlepsze spośród rozważanych, co utwierdza w przekonaniu, że rozważania, które są przedmiotem rozprawy są słuszne i zasadne, a osiągnięte wyniki prawidłowe. Ciekawą obserwacją jest fakt, że niektóre estymatory znane w literaturze, które mają uzasadnienie w intuicji, są bardzo dobre, ale często nieco gorsze niż estymatory zdefiniowane przez Autorkę rozprawy za pomocą wprowadzonego kryterium optymalności. Wyniki dla symulowanych danych i małych prób, porównujące obciążenia i MSE dla estymatorów znanych z literatury i estymatorów zaproponowanych przez Autorkę potwierdzają, że we wszystkich przypadkach zadanych rozkładów (standardowego normalnego, Gumbela, Cauchy'ego, Pareto) nowe estymatory zachowują się całkiem przyzwoicie (poza



pewnymi przykładami  $p$  i niektórych rozkładów), nawet w przypadku rozkładu Cauchy'ego.

Rozprawa doktorska od strony językowej została napisana dojrzałym językiem, bardzo starannie i przejrzysto. Numeracja wzorów oraz ich wykorzystanie są poprawne, bibliografia została dobrana starannie, a wszystkie jej pozycje zostały właściwie wykorzystane w treści pracy. W rozprawie znajduje się co prawda kilka drobnych usterek, ale nie mają one wpływu na jej wartość merytoryczną, m.in. niepoprawny zapis w języku polskim nazw cytowanych twierdzeń pochodzących od nazwisk ich autorów (twierdzenie Glivienki-Cantellego, Morigutiego, Meijera, Schwarza), „błąd średnio-kwadratowy” zamiast „błąd średniokwadratowy”, drobna pomyłka na str. 10, w wierszu 1 (powinno być: „ $0 \leq \alpha < 1$  oraz  $1 \leq j < n$ ”), „szukamy estymator” zamiast „szukamy estymatora”, niejednolity opis zbioru  $C_n^{(1)}$  (kwantyfikator na str.9, 35), zdublowane oznaczenie ostatniego wzoru na str. 28 itp.

Uważam, że zadanie, jakie postawił promotor rozprawy przed panią mgr Luizą Pańczyk było niełatwe, przede wszystkim z uwagi na to, że podjęty temat nie jest nowy, ma długą historię. Aby osiągnąć pożądaną cel, należało zaproponować oryginalne podejście, które do tej pory w literaturze nie było rozważane. W recenzowanej rozprawie jest to nowe kryterium optymalności, które umożliwia wyznaczenie optymalnych  $L$ -statystyk jako estymatorów kwantyla zadanego rzędu. Muszę przyznać, że wraz z zapoznawaniem się z kolejnymi częściami rozprawy, rodziło się we mnie wiele pytań, które w zasadzie od razu były w dalszej części rozstrzygane przez Autorkę. Czuję również pewien niedosyt, że nie udało się uzyskać estymatorów będących kombinacjami liniowymi więcej niż dwóch statystyk pozycyjnych. W mojej opinii praca, jaką wykonała pani mgr Luiza Pańczyk stanowi porządny i matematycznie poprawny wkład w teorię estymacji kwantyli.

### Konkluzja

Podsumowując, w mojej ocenie przedstawiona mi do recenzji rozprawa doktorska pani mgr Luizy Pańczyk pt. „*Wybór optymalnych  $L$ -statystyk w estymacji kwantyli*” spełnia wymagania stawiane pracom doktorskim przez ustawę Prawo o szkolnictwie wyższym z dnia 20 lipca 2018 r. (Dz. U. 2018 poz. 1668). Recenzowana rozprawa stanowi oryginalne rozwiązanie problemu naukowego, jakim jest problem estymacji kwantyli rozkładu, dowodzi ogólnej wiedzy teoretycznej Autorki pracy w zakresie matematyki oraz umiejętności prowadzenia samodzielnie pracy naukowej. Wnoszę więc o dopuszczenie pani mgr Luizy Pańczyk do dalszych etapów postępowania o nadanie stopnia doktora w dyscyplinie matematyka

*Gorany Agnieszka*