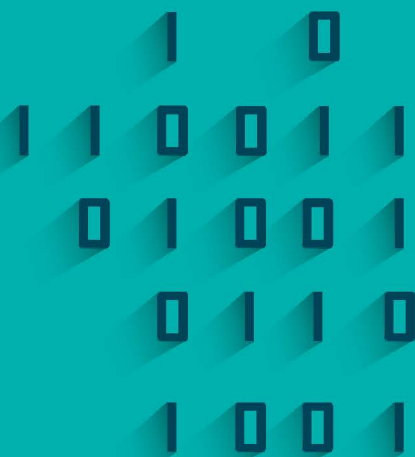


Dorota Pekasiewicz

Statystyki pozycyjne w procedurach estymacji i ich zastosowania w badaniach ekonomicznych



WYDAWNICTWO
UNIwersYTETU
ŁÓDZKIEGO

Statystyki pozycyjne
w procedurach
estymacji i ich
zastosowania
w badaniach
ekonomicznych



WYDAWNICTWO
UNIwersYTETU
ŁÓDZKIEGO

Dorota Pekasiewicz

Statystyki pozycyjne w procedurach estymacji i ich zastosowania w badaniach ekonomicznych

 WYDAWNICTWO
UNIwersytetu
ŁÓDZKIEGO

ŁÓDŹ 2015

Dorota Pekasiewicz – Uniwersytet Łódzki, Wydział Ekonomiczno-Socjologiczny
Katedra Metod Statystycznych, 90-214 Łódź, ul. Rewolucji 1905 r. nr 41/43

RECENZENT
Wojciech Zieliński

REDAKTOR WYDAWNICTWA UŁ
Iwona Gos

SKŁAD KOMPUTEROWY
Barbara Lebioda

PROJEKT OKŁADKI
Stämpfli Polska Sp. z o.o.

Zdjęcie na okładce: © shutterstock.com

© Copyright by Uniwersytet Łódzki, Łódź 2015

Wydane przez Wydawnictwo Uniwersytetu Łódzkiego
Wydanie I. W.06340.13.0.H

ISBN 978-83-7969-519-5 (wersja papierowa)
978-83-7969-520-1 (wersja elektroniczna)

Wydawnictwo Uniwersytetu Łódzkiego
90-131 Łódź, ul. Lindleya 8
www.wydawnictwo.uni.lodz.pl
e-mail: ksiegarnia@uni.lodz.pl
tel. (42) 665 58 63, faks (42) 665 58 62

SPIS TREŚCI

Wprowadzenie	7
1. Statystyki pozycyjne i ich własności	13
1.1. Uwagi wstępne	13
1.2. Podstawowe statystyki pozycyjne	13
1.3. Charakterystyki liczbowe i funkcyjne statystyk pozycyjnych	19
1.4. Graniczne rozkłady statystyk pozycyjnych	33
1.5. Uwagi końcowe	55
2. Metody estymacji oparte na statystykach pozycyjnych	57
2.1. Uwagi wstępne	57
2.2. Metoda kwantyli	58
2.3. Kwantylowa metoda najmniejszych kwadratów	68
2.4. Modyfikacje kwantylowej metody najmniejszych kwadratów	69
2.4.1. Kwantylowa metoda najmniejszych kwadratów z uciętą liczbą kwantyli	70
2.4.2. Medianowo-kwantylowa metoda najmniejszych kwadratów	74
2.5. Metoda momentów ważonych prawdopodobieństwami	75
2.6. Zmodyfikowana metoda momentów ważonych prawdopodobieństwami	84
2.7. Bayesowskie metody estymacji	89
2.8. Bootstrapowe metody estymacji	93
2.9. Uwagi końcowe	98
3. Analiza własności opartych na statystykach pozycyjnych estymatorów parametrów wybranych rozkładów	99
3.1. Uwagi wstępne	99
3.2. Badania własności estymatorów otrzymanych metodą kwantyli	100
3.3. Symulacyjne badania własności estymatorów otrzymanych kwantylową metodą najmnie- szych kwadratów z uciętą liczbą kwantyli	120
3.4. Symulacyjne badania własności estymatorów uzyskanych medianowo-kwantylową metodą najmniejszych kwadratów	132
3.5. Symulacyjne badania własności estymatorów otrzymanych metodami momentów ważo- nych prawdopodobieństwami	133
3.6. Analiza porównawcza własności wybranych estymatorów	138
3.7. Zastosowanie procedur estymacji opartych na statystykach pozycyjnych w badaniach eko- nomicznych	143
3.8. Uwagi końcowe	144

4. Procedury estymacji parametrów pozycyjnych zmiennej losowej i ich zastosowania	147
4.1. Uwagi wstępne.....	147
4.2. Estymatory kwantyli.....	148
4.3. Klasyczne metody wyznaczania przedziałów ufności dla kwantyli.....	156
4.4. Bayesowska estymacja kwantyli.....	163
4.5. Bootstrapowe procedury estymacji kwantyli.....	168
4.6. Estymacja dominanty.....	173
4.7. Przykłady zastosowań estymatorów parametrów pozycyjnych.....	178
4.7.1. Szacowanie miar ubóstwa i bogactwa w analizach dochodów ludności.....	178
4.7.2. Estymacja miar ryzyka rynkowego.....	184
4.7.3. Konstrukcja kart kontrolnych z wykorzystaniem estymatorów mediany.....	192
4.8. Uwagi końcowe.....	195
5. Statystyki pozycyjne w analizach zdarzeń ekstremalnych	197
5.1. Uwagi wstępne.....	197
5.2. Estymacja parametrów uogólnionych rozkładów statystyk ekstremalnych.....	198
5.3. Semiparametryczne metody szacowania indeksu ekstremalnego.....	201
5.4. Estymacja ogona rozkładu zmiennej losowej i jej zastosowanie.....	207
5.5. Bootstrapowa estymacja kwantyli wykorzystująca oszacowanie ogona rozkładu zmiennej losowej.....	216
5.6. Zastosowanie statystyk ekstremalnych w wybranych procedurach estymacji.....	219
5.6.1. Szacowanie ryzyka ekstremalnego.....	219
5.6.2. Konstrukcja kart kontrolnych w oparciu o statystyki ekstremalne.....	223
5.7. Uwagi końcowe.....	226
6. Wybrane empiryczne zastosowania statystyk pozycyjnych w badaniach ekonomicznych ...	227
6.1. Uwagi wstępne.....	227
6.2. Zastosowanie statystyk pozycyjnych w analizach dochodów i wydatków ludności.....	228
6.3. Zastosowanie kwantyli z próby do estymacji miar ryzyka na rynku finansowym.....	234
6.4. Zastosowanie metod estymacji opartych na statystykach pozycyjnych na rynku ubezpieczeniowym.....	242
6.5. Wykorzystanie statystyk pozycyjnych w ocenie działalności przedsiębiorstw.....	248
6.6. Uwagi końcowe.....	251
Zakończenie	253
Order statistics in estimation procedures and their applications in economic research (Summary)	259
Aneks. Charakterystyki funkcyjne i liczbowe wybranych rozkładów	263
Wybrane oznaczenia	275
Literatura	279
Od Redakcji	287

WPROWADZENIE

We współczesnych badaniach ekonomicznych, będących podstawą podejmowania decyzji na różnych poziomach – przedsiębiorstwa, regionu czy też kraju – zauważa się wzrost zapotrzebowania na metody statystyczne. Odgrywają one rolę w procesach zbierania informacji, ich analizowania i interpretowania, a także udostępniania otrzymanych wyników. Ze względu na złożoność i różnorodność gromadzonych obserwacji metody statystyczne oparte na klasycznych parametrach i ich estymatorach, wykorzystywane do analizy zjawisk ekonomicznych, nie zawsze pozwalają na przeprowadzenie pogłębionych analiz i sformułowanie prawidłowych wniosków. Brak momentów zwykłych i centralnych odpowiednich rzędów analizowanych zmiennych losowych, z którymi utożsamiane są badane cechy statystyczne, jak również występowanie obserwacji nietypowych utrudnia wnioskowanie statystyczne klasycznymi metodami. W takich przypadkach mogą być przydatne procedury oparte na statystykach pozycyjnych.

Statystyki pozycyjne stanowią grupę statystyk wyznaczanych na podstawie uporządkowanych prób losowych. Znajdują one zastosowanie w konstrukcji estymatorów parametrów zmiennych losowych wykorzystywanych w procedurach parametrycznej i nieparametrycznej estymacji oraz przy weryfikacji hipotez statystycznych.

Do podstawowych statystyk pozycyjnych zalicza się kwantyle z próby, w tym medianę, statystyki ekstremalne, tj. maksimum i minimum, oraz dominantę z próby. Medianę z próby stosuje się do szacowania wartości średniej, gdy rozkład populacji jest asymetryczny bądź charakteryzuje się tzw. grubymi ogonami. Jest ona znacznie stabilniejsza niż średnia arytmetyczna, która jest bardzo wrażliwa na wartości ekstremalne. Kwantyle rozkładu empirycznego używa się do pomiarów ryzyka rynkowego, finansowego i operacyjnego. Miary oparte na statystykach pozycyjnych stosowane są także w analizach dochodów oraz analizach zjawisk bardzo rzadko występujących, których pojawienie się powoduje duże straty finansowe. Oszacowanie wielkości tych strat możliwe jest przy użyciu statystyk ekstremalnych, ich rozkładów dokładnych lub granicznych. Statystyki pozycyjne i ich funkcje wykorzystywane są również w statystycznej kontroli jakości do tworzenia kart kontrolnych stosowanych w monitorowaniu

i regulacji procesu produkcyjnego oraz w wielu innych analizach dotyczących różnorodnych problemów ekonomicznych.

Głównym celem rozprawy jest przedstawienie metod estymacji parametrów rozkładu populacji wykorzystujących statystyki pozycyjne oraz propozycji ich modyfikacji wraz z zaprezentowaniem wyników przeprowadzonych analiz własności estymatorów stanowiących wskazówki w praktycznych zastosowaniach. W rozważaniach uwzględnione jest klasyczne ujęcie procedur estymacji oraz podejście nieklasyczne – bayesowskie i bootstrapowe, zarówno parametryczne, jak i nieparametryczne.

Aby zrealizować tak sformułowany cel główny, określono cele szczegółowe, do których należą:

- analiza własności statystyk pozycyjnych, w szczególności ich rozkładów dla wybranych klas rozkładów zmiennych losowych;
 - prezentacja metod opartych na statystykach pozycyjnych wykorzystywanych do szacowania parametrów rozkładów zmiennych losowych oraz analiza ich własności;
 - propozycje modyfikacji procedur szacowania parametrów rozkładu zmiennej losowej, prowadzące do otrzymania estymatorów o mniejszych obciążeniach i mniejszych błędach średniokwadratowych;
 - porównanie rozważanych metod dla wybranych klas rozkładów zmiennych losowych oraz sformułowanie wniosków dotyczących ich efektywności;
 - prezentacja parametrycznych i nieparametrycznych metod estymacji kwantyli, w tym mediany;
 - analiza wybranych metod estymacji stosowanych w badaniach zjawisk ekstremalnych, w szczególności metod wykorzystujących oszacowania ogonów rozkładów rozważanych zmiennych;
 - wskazanie obszarów zastosowań rozważanych procedur statystycznych opartych na kwantylach w badaniach ekonomicznych.
- Weryfikacji poddano następujące hipotezy badawcze:
- zastosowanie metody kwantyli z odpowiednio dobranymi rangami stosowanych statystyk pozycyjnych umożliwia uzyskanie estymatorów nieobciążonych lub asymptotycznie nieobciążonych o małych błędach średniokwadratowych;
 - modyfikacje kwantylowej metody najmniejszych kwadratów prowadzą do otrzymania estymatorów parametrów rozkładów populacji o mniejszych obciążeniach i błędach średniokwadratowych niż estymatory uzyskane kwantylową metodą najmniejszych kwadratów oraz metodą kwantyli;
 - modyfikacja metody momentów ważonych prawdopodobieństwami, polegająca na zastosowaniu dystrybuanty empirycznej typu *level crossing*, pozwala otrzymać estymatory o lepszych własnościach w stosunku do estymatorów

uzyskanych metodą momentów ważonych prawdopodobieństwami z klasyczną dystrybuantą empiryczną;

– procedury nieparametrycznej estymacji bootstrapowej umożliwiają uzyskanie przedziałów ufności pokrywających wartość szacowanego parametru z prawdopodobieństwem w przybliżeniu równym ustalonemu współczynnikowi ufności o dokładności większej niż nieparametryczne metody klasyczne.

Praca składa się z sześciu rozdziałów, w których omówiono zagadnienia metodologiczne związane z procedurami estymacji opartymi na kwantylach z próby oraz podano przykłady ich zastosowań.

W rozdziale pierwszym przedstawiono statystyki pozycyjne i ich matematyczne funkcje. Zaprezentowano, znane z literatury przedmiotu, podstawowe twierdzenia dotyczące ich charakterystyk liczbowych, funkcyjnych, w tym rozkładów granicznych, uzupełniając je twierdzeniami dotyczącymi własności statystyk pozycyjnych wyznaczanych w oparciu o ciągi zmiennych losowych o wybranych rozkładach. Są one niezbędne do konstrukcji estymatorów przedstawionych w dalszej części pracy.

W rozdziale drugim omówiono metody estymacji punktowej parametrów rozkładu zmiennej losowej, wykorzystujące statystyki pozycyjne. Prezentowane w literaturze metody: kwantyli (por. J. Bartoszewicz [1996]), kwantylowa metoda najmniejszych kwadratów (por. E. Castillo i in. [2004]), metoda momentów ważonych prawdopodobieństwami (por. J. A. Greenwood i in. [1979]), bootstrapowa (por. B. Efron, R. J. M. Tibshirani [1993]), uzupełnione są autorskimi propozycjami ich modyfikacji pozwalającymi uzyskać estymatory o mniejszym obciążeniu i mniejszej wariancji. Dwie proponowane metody stanowią modyfikacje kwantylowej metody najmniejszych kwadratów, a trzecia – metody momentów ważonych prawdopodobieństwami. Pierwsza z nich polega na pominięciu w estymacji kwantylową metodą najmniejszych kwadratów ustalonej liczby k skrajnych kwantyli z próby, natomiast druga na wyznaczeniu estymatorów kwantylową metodą najmniejszych kwadratów z pominięciem różnej liczby skrajnych kwantyli, a następnie wyznaczeniu mediany z otrzymanych oszacowań. Inna propozycja modyfikacji dotyczy wykorzystania dystrybuanty empirycznej *level crossing* w metodzie momentów ważonych prawdopodobieństwami. Ponadto w rozdziale tym prezentowane są metody estymacji bayesowskiej konstruowane przy ustalonym rozkładzie *a priori* szacowanego parametru i ustalonej funkcji straty. Liniowa funkcja straty sprawia, że estymatorami szacowanych parametrów są kwantyle rozkładu *a posteriori*, czyli pewne funkcje statystyk pozycyjnych. W metodach bootstrapowych, omówionych w jednym z podrozdziałów, istotne znaczenie mają kwantyle rozkładów bootstrapowych stosowane do konstrukcji przedziałów ufności.

W rozdziale trzecim przedstawiono wyniki badań własnych dotyczących własności metod estymacji opartych na statystykach pozycyjnych, ze szczególnym

uwzględnieniem autorskich propozycji. W przypadku rozważanych metod nie zawsze możliwe jest analityczne zbadanie obciążeń i błędów średniokwadratowych otrzymanych estymatorów, dlatego stosowano metody Monte Carlo. Dzięki dostępnemu oprogramowaniu komputerowemu, szybkim procesorom istnieje możliwość wykonania tak dużej liczby powtórzeń analizowanych procedur, że wyniki badań symulacyjnych są praktycznie identyczne z wynikami obliczeń analitycznych. Przeprowadzone badania pozwalają ocenić własności rozpatrywanych metod dla wybranych klas rozkładów populacji, porównać je oraz sformułować wnioski dotyczące ich efektywności i praktycznego zastosowania.

W kolejnym rozdziale pracy zaprezentowano wykorzystanie statystyk pozycyjnych w estymacji parametrów pozycyjnych rozkładu zmiennej losowej, czyli kwantyli i dominanty. Problematyce estymacji punktowej i przedziałowej, parametrycznej oraz nieparametrycznej kwantyli rozkładu badanej zmiennej, w szczególności parametru położenia – mediany, poświęconych jest wiele prac R. Zielińskiego (m.in. [2001], [2003], [2005a]) oraz W. Zielińskiego (np. [2008], [2009]). Oprócz klasycznych metod estymacji, w rozdziale tym przeanalizowano również wybrane bayesowskie i bootstrapowe metody szacowania parametrów pozycyjnych. Rozważano także metody szacowania dominanty, wykorzystujące statystyki pozycyjne (por. np. D. R. Bickel [2002], A. Sokołowski [2013], J. Wywiół [2000b]). W ostatnich podrozdziałach przedstawiono zastosowanie rozważanych estymatorów kwantyli, w tym własnych propozycji do konstrukcji estymatorów miar stosowanych w badaniach ekonomicznych.

W rozdziale piątym omówiono metody estymacji wykorzystywane w analizach zjawisk ekstremalnych, rzadko występujących, których źródłem są załamania na rynkach finansowych, katastrofy czy też nietypowe warunki pogodowe. Podobnie jak w przypadku estymacji kwantyli, do estymacji parametrów rozkładu statystyk ekstremalnych mogą być stosowane parametryczne i nieparametryczne metody prezentowane w literaturze (por. m.in. R. A. Davis, S. T. Resnick [1984], A. L. M. Dekkers i in. [1989], B. M. Hill [1975], J. R. M. Hosting i in. [1985], J. Pickands [1975]) oraz proponowane w rozdziale drugim zmodyfikowane metody estymacji. Istotnym zagadnieniem jest szacowanie indeksu ekstremalnego – parametru określającego kształt rozkładu statystyk ekstremalnych. Jego wartość związana jest z klasą rozkładu populacji. Gdy rozkład populacji charakteryzuje się grubymi (ciężkimi) ogonami, to jego wartość jest dodatnia, gdy cienkimi (lekkimi) ogonami – indeks wynosi zero, natomiast dla rozkładów o krótkich ogonach (ograniczonym przedziale wartości) przyjmuje on wartość ujemną. Ma to znaczenie przy wykrywaniu wartości nietypowych, rzadko występujących, przy obliczaniu prawdopodobieństw zajścia zdarzeń ekstremalnych oraz szacowaniu wielkości pojawiających się katastrof, przy ustalonym prawdopodobieństwie ich wystąpienia. Ponadto podano przykłady wykorzystania statystyk

ekstremalnych i ich funkcji do określania miar stosowanych w analizach ekonomicznych, w tym finansowych.

W rozdziale szóstym zaprezentowano empiryczne przykłady zastosowań metod estymacji opartych na statystykach pozycyjnych rozważanych w pracy. Ograniczono się do wspomnianych już wcześniej trzech obszarów badań ekonomicznych: analizy dochodów, bogactwa i ubóstwa, statystycznej kontroli jakości oraz zarządzania ryzykiem, tzw. zwykłym i ekstremalnym, a także wskazano możliwość ich wykorzystania w ubezpieczeniach majątkowych. Na podstawie rzeczywistych danych statystycznych pochodzących z Głównego Urzędu Statystycznego, jednostki kontrolującej jakość w przedsiębiorstwie produkującym urządzenia gospodarstwa domowego, publikowanych indeksów polskiej i amerykańskiej giełdy papierów wartościowych oraz danych dotyczących ubezpieczeń komunikacyjnych pochodzących z pewnego zakładu ubezpieczeń zaprezentowano zastosowanie wybranych metod.

W zamieszczonym aneksie przedstawiono podstawowe charakterystyki funkcyjne i liczbowe rozkładów zmiennych losowych rozważanych w pracy.

W niniejszej monografii zaprezentowano zarówno znane z literatury procedury estymacji, jak i własne propozycje. W poszczególnych rozdziałach monografii przedstawiano rezultaty analitycznych rozważań oraz badań symulacyjnych przeprowadzonych w oparciu o samodzielnie przygotowane programy napisane w środowisku Gauss i Mathematica.

Pragnę serdecznie podziękować Recenzentowi – Panu Profesorowi zw. dr. hab. Wojciechowi Zielińskiemu – za cenne uwagi i sugestie, które wpłynęły na poprawę jakości publikacji.

1. STATYSTYKI POZYCYJNE I ICH WŁASNOŚCI

1.1. Uwagi wstępne

Statystyki pozycyjne, zwane również porządkowymi, definiuje się na podstawie prób losowych uporządkowanych w sposób niemalejący lub nierosnący.

W rozdziale przedstawiono pojęcia i własności podstawowych statystyk pozycyjnych, do których należą kwantyle z próby, w szczególności mediana, kwartyle, decyle i percentyle z próby, statystyki ekstremalne oraz dominanta z próby. Ponadto rozważano statystyki będące funkcjami statystyk porządkowych, wykorzystywane w estymacji parametrów położenia i zróżnicowania.

Dla wybranych klas rozkładów sformułowano twierdzenia określające funkcje gęstości, dystrybuanty oraz charakterystyki liczbowe statystyk pozycyjnych. Analizowano również rozkłady graniczne statystyk ekstremalnych, wykorzystywanych w badaniach zjawisk nietypowych. Wyboru rozpatrywanych rozkładów dokonano na podstawie analizy rozkładów mających praktyczne zastosowanie w badaniach społeczno-ekonomicznych. W szczególności rozważano rozkłady zmiennych losowych, które nie mają momentów centralnych pierwszego i drugiego rzędu. Wykorzystanie zatem we wnioskowaniu statystycznym takich estymatorów, jak średnia arytmetyczna czy wariancja jest niemożliwe.

Przedstawione statystyki pozycyjne oraz ich funkcje stosowane są w estymacji parametrów rozkładów zmiennych losowych występujących w badaniach ekonomicznych oraz do szacowania różnego rodzaju miar definiowanych w oparciu o kwantyle rozkładów.

1.2. Podstawowe statystyki pozycyjne

Niech X_1, X_2, \dots, X_n będzie ciągiem niezależnych zmiennych losowych o rozkładzie określonym za pomocą dystrybuanty F , x_1, x_2, \dots, x_n – ciągiem ich wartości, natomiast $x_{(1)}^{(n)}, x_{(2)}^{(n)}, \dots, x_{(n)}^{(n)}$ – uporządkowanym niemalejąco ciągiem tych wartości.

Statystyka pozycyjna jest funkcją wektora losowego (X_1, X_2, \dots, X_n) zdefiniowaną w następujący sposób (por. np. M. Fisz [1967, s. 389–390], C. Domański i in. [1998, s. 176]).

Definicja 1.2.1. Statystyką pozycyjną $X_{(k)}^{(n)}$, gdzie $k = 1, 2, \dots, n$, nazywamy zmienną losową, której wartościami są k -te co do wielkości wartości realizacji, uporządkowanego w sposób niemalejący, wektora losowego (X_1, X_2, \dots, X_n) , stanowiącego próbę losową, czyli wartości $x_{(k)}^{(n)}$.

Liczbę k nazywamy rangą statystyki pozycyjnej $X_{(k)}^{(n)}$, natomiast wielkość $\frac{k}{n}$ określamy jako rangę względną tej statystyki.

Statystyki pozycyjne zwane są również statystykami porządkowymi (por. J. Bartoszewicz [1996, s. 68]).

We wnioskowaniu statystycznym wykorzystuje się statystyki wyznaczane w oparciu o n -elementową próbę prostą, którą stanowi ciąg niezależnych zmiennych losowych X_1, X_2, \dots, X_n , czyli wektor losowy (X_1, X_2, \dots, X_n) . Za pomocą statystyk pozycyjnych definiuje się kwantyle z próby, w szczególności medianę, kwartyle, decyle i percentyle z próby.

Definicja 1.2.2. Kwantylem rzędu p , gdzie $p \in (0, 1)$, z n -elementowej próby prostej X_1, X_2, \dots, X_n nazywamy statystykę postaci:

$$X_{p;n} = \begin{cases} X_{(np)}^{(n)}, & \text{gdy } np \in N, \\ X_{(\lfloor np \rfloor + 1)}^{(n)}, & \text{gdy } np \notin N, \end{cases} \quad (1.2.1)$$

gdzie $\lfloor np \rfloor$ oznacza część całkowitą liczby np , natomiast N jest zbiorem liczb naturalnych.

Kwantyl rzędu $p = 0,5$ z próby losowej X_1, X_2, \dots, X_n nazywany jest medianą. Ze względu na symetrię często definiuje się medianę w poniższy sposób (por. R. Zieliński [2011, s. 33]).

Definicja 1.2.3. Medianą Me z n -elementowej próby prostej X_1, X_2, \dots, X_n nazywamy statystykę określoną wzorem: