METODY ILOŚCIOWE W BADANIACH EKONOMICZNYCH
Tom XV/3, 2014, str. 111 – 121

SZACOWANIE MEDIANY PRZY UŻYCIU DOKŁADNEJ METODY BOOTSTRAPOWEJ

Joanna Kisielińska
Wydział Nauk Ekonomicznych
Szkoła Główna Gospodarstwa Wiejskiego w Warszawie
e-mail: joanna_kisielinska@sggw.pl

Streszczenie: W artykule przedstawiono dokładną metodą bootstrapową, którą wykorzystano do szacowania mediany. Obliczenia potwierdziły jej skuteczność, ponieważ dla próby o nieparzystej liczbie elementów, oszacowany rozkład standardowego estymatora mediany dokładnie pokrywał się z rozkładem teoretycznym. Dla próby o parzystej liczbie elementów pokazano, że standardowy estymator mediany może być obciążony, co wskazuje na konieczność poszukiwania innej jego postaci.

Słowa kluczowe: estymacja mediany, dokładna metoda bootstrapowa

WPROWADZENIE

Znajomość wartości jakie przyjmują rzeczywiste cechy bądź zjawiska w wielu zagadnieniach, również dotyczących problemów ekonomicznych, jest sprawą kluczową. Problem nie występuje, jeśli badacz dysponuje wszystkimi ich realizacjami. Możliwe jest wówczas wyznaczenie rozkładu oraz interesujących jego parametrów. W praktyce jednak sytuacja tak komfortowa rzadko ma miejsce. Wynikać to może ze zbyt dużych kosztów uzyskania wszystkich realizacji (rozważana populacja jest zbyt duża), bądź faktu, że realizacji jest nieskończenie wiele (populacja nie jest skończona). Zwykle dysponujemy jedynie pewnym podzbiorem możliwych realizacji, zwanych próbą. Na podstawie próby formułujemy pewne stwierdzenia dotyczące całej populacji.

Dalsze rozważania ograniczymy do przypadku poszukiwania wartości parametrów rozkładu. Jeśli nie dysponujemy informacją o całej populacji, parametru obliczyć się nie da, można go jedynie oszacować na podstawie próby. W tym celu dobieramy odpowiednią statystykę zwaną estymatorem.

Niech cechę bądź zjawisko reprezentuje pewną zmienna losową X, której dystrybuantę oznaczymy jako F, zaś poszukiwany parametr jako θ. Ze skończonej populacji można pobrać wiele n-elementowych prób, zaś z nieskończonej – nieskończenie wiele. Możliwe do wylosowania próby są w ciągiem zmiennych losowych o jednakowych rozkładach, które oznaczymy jako $\mathbf{X} = (X_1, X_2, \ldots, X_n)$. Dokonując losowania próby wybieramy w istocie pewna jej realizację $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. Estymatorem (oszacowaniem) parametru θ jest funkcja $\hat{\theta} = t(\mathbf{X})$. Aby formułować pewne stwierdzenia dotyczące parametru θ należy znać rozkład estymatora $\hat{\theta}$. Podkreślić należy, że jedynie w pewnych przypadkach zakładając rozkład F zmiennej X, można wyznaczyć rozkład estymatora.

Jeśli rozkład F nie jest znany, lub znajomość jego nie pozwala na wyznaczenie rozkładu estymatora można zastosować zaproponowaną przez Efrona (1979) metodę bootstrapową, polegającą na wtórnym próbkowaniu próby $\mathbf{x}$. Z próby pierwotnej $\mathbf{x}$ losowanych jest ze zwracaniem N prób wtórnych. Próby wtórne są tak, jak próba pierwotna n-elementowe. Jeśli wszystkie elementy próby pierwotnej są różne, prawdopodobieństwa wylosowania jednej z wartości x_i są jednakowe i równe $1/n$.

Próby wtórne oznaczane w literaturze zwykle jako $\mathbf{X}^* = \left(X_1^*, X_2^*, \cdots, X_n^*\right)$ są ciągami zmiennych losowych o jednakowych rozkładach. Każda ze zmiennych losowych X_l^* może przyjmować wartości ze zbioru $\mathbf{x}$. Zmienne te mają rozkład zwany rozkładem bootstrapowym oznaczanym jako $\hat{F}$, który jest równoważny rozkładowi empirycznemu. Pojedynczą próbę wtórną można zapisać jako $\mathbf{x}^* = \left(x_1^*, x_2^*, \cdots, x_n^*\right)$, zaś estymator $\hat{\theta}$ dla próby bootstrapowej jako $\hat{\theta}^* = t(\mathbf{X}^*)$.

Jak wspomniano wcześniej do formułowania stwierdzeń dotyczących parametru θ potrzebna jest znajomość rozkładu jej estymatora $\hat{\theta}$. W metodzie bootstrapowej rozkład ten aproksymowany jest rozkładem statystyki $\hat{\theta}^*$. Zamiast poszukiwać rozkładu $\hat{\theta}$ poszukiwany jest rozkład $\hat{\theta}^*$.

W klasycznej metodzie bootstrapowej rozkład $\hat{\theta}^*$ wyznaczany jest na podstawie N prób wtórnych. Dokładne jego wyznaczenie wymagałoby znajomości wszystkich prób wtórnych. Dlatego metoda bootstrapowa, wykorzystująca do szacowania wszystkie próby wtórne nazywa się metodą dokładną. Metoda ta przedstawiona zostanie w następnym punkcie.

Metoda dokładnego bootstrapu wykorzystana zostanie do szacowania mediany. Rozkład estymatora mediany dla dowolnej próby o nieparzystej liczbie elementów jest znany, co pozwoli na weryfikację skuteczności metody dokładnego bootstrapu. Dla prób o nieparzystej liczbie elementów natomiast, metoda ta pozwoli na wygenerowanie rozkładu mediany.

ALGORYTM DOKŁADNEGO BOOTSTRAPU

Do dokładnego wyznaczenia rozkładu $\hat{\theta}^*$ potrzebna jest znajomość wszystkich prób wtórnych, których liczbę oznaczmy jako B. Wartość ta określa liczbę sposobów, na które można wylosować n liczb ze zbioru n-elementowego. Z kombinatoryki wiadomo, że jest to liczba wariacji z powtórzeniami, czyli $B = n^n$. Klasyczna metoda bootstrapowa (Efron (1979)) polega więc na losowaniu ze zwracaniem N elementów ze zbioru B-elementowego.

Na możliwość wygenerowania całego zbioru prób wtórnych zwracali uwagę Fisher i Hall (1991), podkreślając jednak, że jest to możliwe jedynie dla małych prób. Ograniczenia wynikają z czasu trwania obliczeń. Zauważmy bowiem, że dla n=5 liczba wszystkich prób wtórnych jest równa B=3125, dla n=10 już B=10 mld, a dla n=20 otrzymujemy $B \approx 10{,}5 \cdot 10^{25}$. Ze względu na postęp w technice komputerowej możliwa jest realizacja metody dokładnego bootstrapu dla coraz większych prób, co pokazała Kisielińska (2011),(2013).

Liczba wszystkich prób wtórnych jest bardzo duża. Zwróćmy jednak uwagę, że wiele z nich różni się jedynie kolejnością wylosowanych elementów. Liczba „różnych" prób wtórnych B_r jest znacznie mniejsza i wynosi (Feller (1950), Fisher i Hall (1991)):

$$B_r = \binom{2n-1}{n}. \tag{1}$$

Aby zastosować dokładną metodę bootstrapową należy wobec tego wygenerować wszystkie „różne" próby wtórne, a następnie obliczyć ile razy każda z nich występuje w zbiorze wszystkich prób wtórnych.

Załóżmy, że w próbie pierwotnej $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ wszystkie elementy są różne (przypadek powtarzających się elementów zostanie omówiony dalej). W próbach wtórnych natomiast elementy mogą się powtarzać. Oznaczmy jako m liczbę różnych elementów próby wtórnej, zaś liczbę wystąpień elementu l jako n_l, przy czym $1 \leq n_l \leq n$. Wartości te spełniać muszą warunek:

$$n_1 + n_2 + \ldots + n_m = n. \tag{2}$$

Z n-elementowej próby pierwotnej m różnych elementów można wylosować na $\binom{n}{m}$ sposobów (liczba m-elementowych kombinacji ze zbioru n-elementowego). Dla każdej kombinacji należy określić ile razy w próbie każdy jej element wystąpi. Gdyby wszystkie wartości n_l były różne, byłoby to $m!$, czyli liczba permutacji m-elementowego zbioru $\{n_1, n_2, \ldots, n_m\}$. Oznaczmy jako s liczbę różnych wartości w zbiorze liczb $\{n_1, n_2, \ldots, n_m\}$, zaś jako m_j liczbę wystąpień n_j w tym zbiorze. Wartości te spełniać muszą warunek:

$$m_1 + m_2 + \ldots + m_s = m. \tag{3}$$

Elementów każdej kombinacji nie należy permutować na pozycjach odpowiadających jednakowym wartościom n_l, a wobec tego liczba wystąpień każdej kombinacji jest równa $\frac{m!}{m_1!\cdot m_2!\cdot\ldots\cdot m_s!}$.

Zauważmy, że w próbach wtórnych, każdy element kombinacji może wystąpić, na różnych pozycjach. Podobnie jak wyżej, gdyby wszystkie elementy próby wtórnej były różne, liczba permutacji równa byłaby $n!$. Ze względu na powtórzenia, liczba różnych ustawień m elementów powtarzających się po n_l razy jest równa $\frac{n!}{n_1!\cdot n_2!\cdot\ldots\cdot n_m!}$.

Dla każdego ciągu wartości $\{n_1,n_2,...,n_m\}$ liczba możliwych prób wtórnych jest ostatecznie równa:

$$\binom{n}{m}\cdot\frac{m!}{m_1!\cdot m_2!\cdot\ldots\cdot m_s!}\cdot\frac{n!}{n_1!\cdot n_2!\cdot\ldots\cdot n_m!}. \tag{4}$$

Pozostaje pytanie jak wygenerować wszystkie ciągi $\{n_1,n_2,...,n_m\}$. Problem ten jest dość skomplikowanym, zwanym w kombinatoryce zadaniem podziału liczb. Algorytm opracowany został już przez Eulera i przedstawiony jest przez Ryttera (2013). Feller (1950) natomiast zadanie to nazywa problemem rozmieszczenia (str. 38). Opis zostanie tu pominięty.

Schemat generowania wszystkich różnych prób wtórnych jest więc następujący:

1. Wygeneruj dla danego n wszystkie ciągi $\{n_1,n_2,...,n_m\}$.
2. Dla każdego ciągu $\{n_1,n_2,...,n_m\}$ wylosuj wszystkie m-elementowe kombinacje losowane ze zbioru n-elementowego. Dla każdej z kombinacji określ wszystkie permutacje ciągu $\{n_1,n_2,...,n_m\}$ określające liczbę powtórzeń elementów próby.

Mając wygenerowane wszystkie różne próby wtórne, można wyznaczyć dokładne bootstrapowe oszacowanie parametru θ oraz błąd tego oszacowania. Każda próba wtórna pozwala wyznaczyć pojedynczą realizację statystyki $\hat{\theta}*$. Oszacowanie boostrapowe parametru θ oblicza się jako średnią arytmetyczną wszystkich realizacji. Wyznaczenie wszystkich realizacji statystyki $\hat{\theta}*$ jedynie z różnych prób bootstrapowych (pozwalające na znaczną oszczędność czasu) wymaga wprowadzenia wag równych liczbie możliwych permutacji tych prób. Jeżeli jako x^{*b} oznaczymy b-tą różną próbę wtórną oszacowaniem bootstrapowym parametru θ będzie:

$$\hat{\theta}^*(\bullet)=\frac{1}{B_r}\sum_{b=1}^{B_r}\hat{\theta}^*\left(x^{*b}\right)\cdot\frac{n!}{n_1!\cdot n_2!\cdot\ldots\cdot n_m!}. \tag{5}$$

Błąd tego oszacowania będzie zaś równy:

$$\hat{s}^* = \sqrt{\frac{1}{B_r}\sum_{b=1}^{B_r}\left(\hat{\theta}^*\left(x^{*b}\right)-\hat{\theta}^*(\bullet)\right)^2 \cdot \frac{n!}{n_1!\cdot n_2!\cdot\ldots\cdot n_m!}}\,. \tag{6}$$

Jeżeli w próbie pierwotnej występują powtórzenia, konieczna jest korekta wzorów (5) i (6), wynikająca z faktu, że poszczególne próby wtórne nie są już jednakowo prawdopodobne.

Jeśli w próbie pierwotnej nie występują powtórzenia, prawdopodobieństwo wylosowania każdego z jej elementów do próby wtórnej jest równe $1/n$. Jeśli zaś powtórzenia występują prawdopodobieństwo p_l wylosowania elementu l-tego będzie równe liczbie powtórzeń podzielonej przez n. Prawdopodobieństwo wylosowania b-tej próby wtórnej zawierającej elementy $\{x_{b1}, x_{b2}, \ldots , x_{bm}\}$ jest wobec tego równe $p_b = p_{b1}\cdot p_{b2}\cdot\ldots\cdot p_{bm}$. Oszacowanie bootstrapowe parametru θ będzie równe wówczas:

$$\hat{\theta}^*(\bullet) = \sum_{b=1}^{B_r}\hat{\theta}^*\left(x^{*b}\right)\cdot\frac{n!}{n_1!\cdot n_2!\cdot\ldots\cdot n_m!}\cdot p_b\,, \tag{7}$$

zaś błąd tego oszacowania:

$$\hat{s}^* = \sqrt{\sum_{b=1}^{B_r}\left(\hat{\theta}^*\left(x^{*b}\right)-\hat{\theta}^*(\bullet)\right)^2 \cdot \frac{n!}{n_1!\cdot n_2!\cdot\ldots\cdot n_m!}\cdot p_b}\,. \tag{8}$$

ROZKŁAD MEDIANY

Medianą w uporządkowanym zbiorze jest wartość, powyżej i poniżej której znajduje się jednakowa liczba jego elementów. Najczęściej definiuje się medianę jako wartość od której przynajmniej połowa elementów zbioru przyjmuje wartości mniejsze lub równe medianie i przynajmniej połowa - wartości większe lub jej równe. O ile dla zbioru o nieparzystej liczbie elementów obydwie definicje jednoznacznie pozwalają wskazać medianę, o tyle dla parzystej liczby potencjalnych wartości mediany jest wiele.

Jeśli rozpatrywanym zbiorem jest uporządkowany zbiór liczb $\mathbf{x} = \left(x_{(1)}, x_{(2)}, \ldots, x_{(n)}\right)$, mediana obliczana jest jako:

$$me = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & \text{jeśli } n \text{ jest nieparzyste} \\ \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\right) & \text{jeśli } n \text{ jest parzyste} \end{cases} \tag{9}$$

Zauważmy, że dla próby parzystej medianą może być każda liczba z przedziału $\left(x_{\left(\frac{n}{2}\right)}, x_{\left(\frac{n}{2}+1\right)}\right)$, zwyczajowo jednak przyjmuje się wartość leżącą w połowie przedziału między środkowymi wartościami.

Oznaczmy wszystkie możliwe do wylosowania uporządkowane próby jako $\mathbf{X} = \left(X_{(1)}, X_{(2)}, ..., X_{(n)}\right)$. Wówczas standardowy estymator mediany (Zieliński (2010)) określony jest wzorem:

$$Me = \begin{cases} X_{\left(\frac{n+1}{2}\right)} & \text{jeśli } n \text{ jest nieparzyste} \\ \frac{1}{2}\left(X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}\right) & \text{jeśli } n \text{ jest parzyste} \end{cases} \tag{10}$$

Dla próby o nieparzystej liczbie elementów znany jest rozkład teoretyczny mediany (Efron (1979)). Oznaczmy pozycję mediany jako m, zaś możliwe do wylosowania uporządkowane próby bootstrapowe jako $\mathbf{X}^* = \left(X^*_{(1)}, X^*_{(2)}, ..., X^*_{(n)}\right)$. W próbach bootstrapowych medianą $X^*_{(pm)}$ może być każdy element próby $x_{(l)}$. Aby wyznaczyć rozkład mediany, należy dla każdego $x_{(l)}$ obliczyć prawdopodobieństwo tego, że jest on medianą. Prawdopodobieństwo to można zapisać jako:

$$P\left(X^*_{(pm)} = x_{(l)}\right) = P\left(X^*_{(pm)} > x_{(l-1)}\right) - P\left(X^*_{(pm)} > x_{(l)}\right) \tag{11}$$

Aby mediana była większa od $x_{(l)}$, należy w próbie wtórnej wylosować co najwyżej pm-1 elementów mniejszych lub równych $x_{(l)}$, zaś pozostałe wylosowane elementy muszą być od $x_{(l)}$ większe. Prawdopodobieństwo wylosowania pojedynczego elementu mniejszego lub równego $x_{(l)}$ jest równe $\frac{l}{n}$. Prawdopodobieństwo wylosowania co najwyżej pm-1 elementów spełniających ten warunek można obliczyć stosując schemat Bernoulliego, czyli:

$$P\left(X^*_{(pm)} > x_{(l)}\right) = \sum_{j=0}^{pm-1} \binom{n}{j} \left(\frac{l}{n}\right)^j \left(1 - \frac{l}{n}\right)^{n-j}. \tag{12}$$

Analogicznie obliczamy drugie prawdopodobieństwo we wzorze (11):

$$P\left(X^*_{(pm)} > x_{(l-1)}\right) = \sum_{j=0}^{pm-1} \binom{n}{j} \left(\frac{l-1}{n}\right)^j \left(1 - \frac{l-1}{n}\right)^{n-j}. \tag{13}$$

Ostatecznie rozkład estymatora mediany określają następujące prawdopodobieństwa:

$$P\left(X^*_{(pm)} = x_{(l)}\right) = \sum_{j=0}^{pm-1} \binom{n}{j} \left(\frac{l-1}{n}\right)^j \left(1 - \frac{l-1}{n}\right)^{n-j} - \sum_{j=0}^{pm-1} \binom{n}{j} \left(\frac{l}{n}\right)^j \left(1 - \frac{l}{n}\right)^{n-j}. \qquad (14)$$

Zauważmy, że tak określone prawdopodobieństwa nie zależą od próby pierwotnej, a jedynie od liczby jej elementów *n*. Własności tej nie będzie miał rozkład mediany określony regułą (10) dla próby o parzystej liczbie elementów. Medianą dla wtórnych prób bootstrapowych mogą być bowiem wszystkie elementy próby oraz średnie arytmetyczne wszystkich możliwych do wylosowania par jej elementów. Nie można z góry określić, między którymi elementami próby, średnie się znajdują. Ponieważ niektóre średnie arytmetyczne będą się powtarzać, prawdopodobieństwa w ich przypadku należy dodać. Wynika z tego, że prawdopodobieństwa dla poszczególnych realizacji standardowego estymatora mediany zależeć będą od próby.

WYNIKI OBLICZEŃ

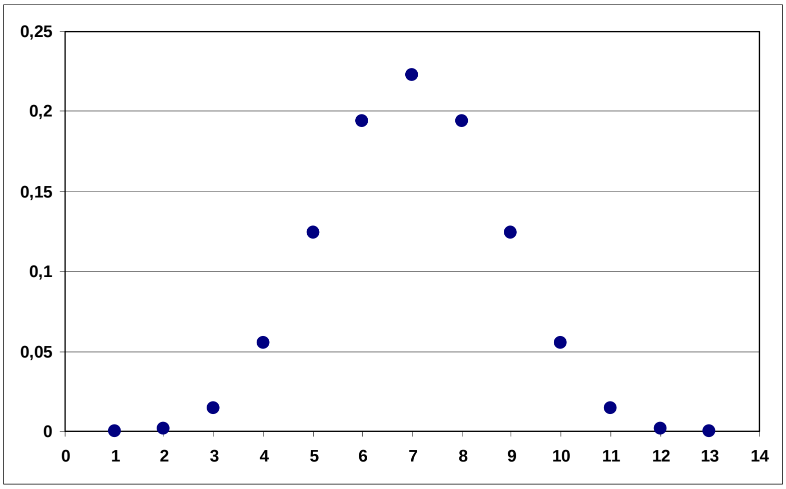
Estymację mediany przeprowadzono dla 13-elementowej próby. Przykład pochodzi z pracy Erona (1979). Ponieważ próba ma nieparzystą liczbę elementów rozkład mediany nie zależy od elementów próby, a jedynie od jej liczebności. W tabeli 1 oraz na rysunku 1 przedstawiony został teoretyczny rozkład mediany, oraz rozkład uzyskany metodą dokładnego bootstrapu. Obydwie metody dały takie same wyniki, co potwierdza poprawność wykorzystanego algorytmu.

Tabela 1. Teoretyczny rozkład mediany i rozkład uzyskany metodą dokładnego bootstrapu dla próby nieparzystej

x_l	1 i 13	2 i 12	3 i 11	4 i 10	5 i 9	6 i 8	7
Rozkład teoretyczny	1,8E-05	0.00146	0.01423	0.05495	0.12427	0,19361	0,22294
Rozkład bootsrapowy	1,8E-05	0.00146	0.01423	0.05495	0.12427	0,19361	0,22294

Źródło: obliczenia własne

Rysunek 1. Rozkład mediany - teoretyczny i uzyskany metodą dokładnego bootsrapu dla próby nieparzystej



Źródło: obliczenia własne

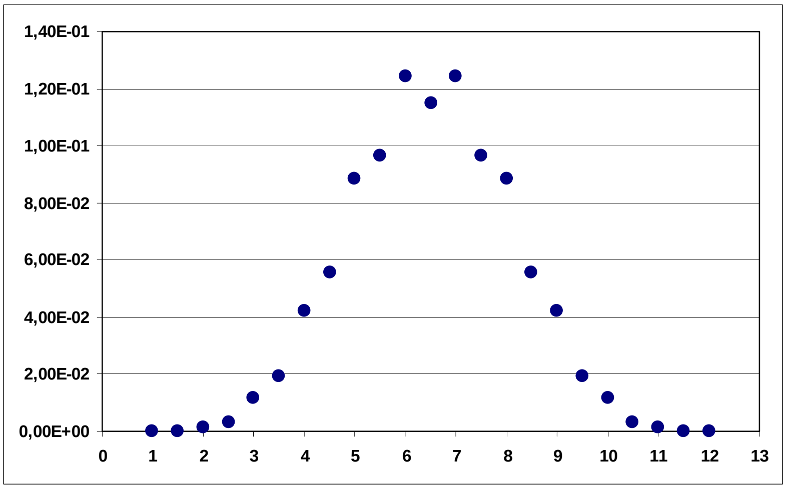
Dla 12-elementowej próby rozkład mediany oszacowany dokładną metodą bootstrapową przedstawiono w tabeli 2 i na rysunku 2. W przykładzie jako elementy próby przyjęto kolejne liczby naturalne począwszy od jedynki. Uzyskany rozkład jest również symetryczny. Zauważmy, że najbardziej prawdopodobnymi wartościami mediany są 6 i 7, a nie wartość faktyczna, jaką jest 6,5. Ponieważ w wykorzystanej próbie różnice między jej kolejnymi elementami są jednakowe (równe 1), liczba możliwych wartości mediany nie jest duża. Są to jedynie elementy próby oraz wartości leżące pomiędzy nimi. Wartość oczekiwana estymatora mediany jest równy jej faktycznej wartości dla próby, czyli 6,5.

Tabela 2. Teoretyczny rozkład mediany i rozkład uzyskany metodą dokładnego bootstrapu dla 12-elementowej próby parzystej

x_l	1 i 12	1,5 i 11,5	2 i 11	2,5 i 10,5	3 i 10	3,5 i 9,5
Rozkład bootsrapowy	1,5E-05	0.00008	0,00125	0,00309	0,01163	0,01951
x_l	4 i 9	4,5 i 8,5	5 i 8	5,5 i 7,5	6 i 7	6,5
Rozkład bootsrapowy	0,04227	0,05552	0,08847	0,09647	0,12432	0,11479

Źródło: obliczenia własne

Rysunek 2. Rozkład mediany uzyskany metodą dokładnego bootstrapu dla 12-elementowej próby parzystej



Źródło: obliczenia własne

Dla 12-elementowej próby (przedstawionej w Tabeli 3), w której różnice między kolejnymi elementami nie są takie same, rozkład mediany oszacowany dokładną metodą bootstrapową przedstawiono na rysunku 3. Pominięto prezentacje tabelaryczną, ponieważ w rozkładzie mediany jest w tym przypadku 75 realizacji. Wielu spośród nich odpowiada bardzo małe prawdopodobieństwo, ponieważ w nielicznych próbach wtórnych są medianą. Najmniej prawdopodobną medianą w próbach wtórnych jest 9,25 (średnia arytmetyczna 1,1 i 17,4). W próbach tych 1,1 musi wystąpić dokładnie 6 razy. Pozostałe elementy to 17,4 przynajmniej raz oraz 20. Prób wtórnych, w których medianą jest 9,25 będzie więc: $\frac{6!}{6!1!5!}+\frac{6!}{6!2!4!}+\frac{6!}{6!3!3!}+\frac{6!}{6!4!2!}+\frac{6!}{6!5!1!}+\frac{6!}{6!6!}=58212$. Prawdopodobieństwo uzyskania mediany 9,25 jest więc równe $P(me=9{,}25)=\frac{58212}{12^{12}}\approx 6{,}52886\cdot 10^{-9}$, co zgadza się z prawdopodobieństwem wyznaczonym metodą dokładnego bootstrapu.

Dokładny rozkład mediany na wykorzystanej w obliczeniach próby nie jest symetryczny. Posiada ponadto wyraźnie niepożądane własności, polegające na tym, że wartości bardziej oddalone od wartości oczekiwanej mają wyższe prawdopodobieństwa realizacji, niż niektóre wartości znajdujące się bliżej niej.

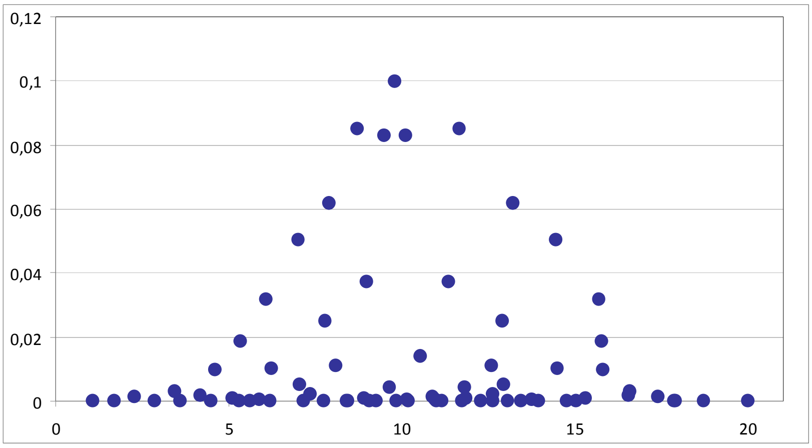
Mediana próby jest równa 9,8, natomiast bootstrapowe jej oszacowanie dokonane metodą dokładnego bootstrapu wynosi w przybliżeniu 10,26. Oznacza to, że estymator użyty do jej szacowania jest obciążony. Potwierdza to wyniki teoretycznych rozważań Zielińskiego (2010 str. 14), że dla próby z parzystą liczbą obserwacji „mediana rozkładu estymatora może być dowolnie daleko od estymowanej mediany rozkładu". Wskazuje to na konieczność poszukiwania innego niż standardowy estymatora.

Tabela 3. Próba 12-elementowa, w której różnice między kolejnymi elementami nie są takie same

l	1	2	3	4	5	6	7	8	9	10	11	12
x_l	1,1	2,3	4,6	6,1	7,9	9,5	10,1	13,2	15,7	15,8	17,4	20

Źródło: opracowanie własne

Rysunek 3. Rozkład mediany uzyskany metodą dokładnego bootstrapu dla 12-elementowej próby parzystej, w której różnice między kolejnymi elementami nie są takie same



Źródło: obliczenia własne

PODSUMOWANIE

Z przedstawionych w artykule badań wynikają następujące wnioski:

1. Dokładna metoda bootstrapowa pozwala wygenerować dokładny rozkład estymatora mediany. Potwierdzają to obliczenia wykonane dla próby o nieparzystej liczbie elementów, dla której znany jest rozkład teoretyczny. Prawdopodobieństwa dla wszystkich jego realizacji zależą jedynie od liczebności próby, a nie jej elementów. Rozkład teoretyczny jest taki sam, jak uzyskany dokładną metodą bootstrapową.
2. Dla próby o nieparzystej liczbie elementów rozkład, prawdopodobieństwa poszczególnych realizacji mediany zależą od próby, a nie tylko od jej liczebności. Przeprowadzone badania pokazały (potwierdzając rozważania teoretyczne), że standardowy estymator mediany może mieć rozkład o wartości oczekiwanej odbiegającej od wartości mediany dla próby, co wskazuje, że jest to estymator obciążony. Dodatkowo, w wielu przypadkach, prawdopodobieństwa poszcze-

gólnych realizacji bardziej oddalone od wartości oczekiwanej rozkładu mają wartości wyższe, niż znajdujące się bliżej niej.
3. Badania wskazują na konieczność poszukiwania innego estymatora mediany dla próby o nieparzystej liczbie elementów.

BIBLIOGRAFIA

Efron B. (1979) Bootstrap methods: another look at the jackknife. The Annals of Statistics. Vol. 7, No. 1, 1-26.

Feller W. (1950): An introduction to probability theory and its application. John Wiley & Sons. New York, London, Sydney.

Fisher N.I., Hall P. (1991) Bootstrap algorithms for small samples. Journal of Statistical Planning and Inference. Vol. 27, 157-169.

Kisielińska J. (2011): Dokładna metoda bootstrapowa na przykładzie estymacji średniej. Metody Ilościowe w Badaniach Ekonomicznych. Tom 12/ nr 2. str. 191-211

Kisielińska J. (2013) The exact bootstrap method shown on the example of the mean and variance estimation. Computational Statistics (Volume 28, Issue 3, June 2013) str. 1061-1077

Rytter W. (2013) Zliczanie podziałów liczb: algorytm Eulera. Publikacja elektroniczna: http://www.deltami.edu.pl/temat/informatyka/algorytmy/2013/01/30/algorytm_eulera.pdf (1.09.2014)

Zieliński R. (2010) O średniej arytmetycznej i medianie. Matematyka Stosowana, Tom 11/52.

MEDIAN ESTIMATING USING THE EXACT BOOTSTRAP METHOD

Abstract: The article presents the exact bootstrap method, which was used to estimate the median. Calculations have confirmed its effectiveness, because attempts with an odd number of elements, the estimated distribution of the standard median estimator exactly coincide with the theoretical distribution. For the sample with an even number of items showing that the standard estimator of the median may be biased, which indicates the need to seek its other form.

Keywords: median estimating, exact bootstrap method