Wrocław, 18 maja 2021

Prof. dr hab. Ryszard Magiera
Wydział Matematyki
Politechnika Wrocławska

RECENZJA

rozprawy doktorskiej

Estymacja kwantylowych wersji krzywej Lorenza

Estimation of quantile versions of the Lorenz curve

mgr Agnieszki Siedlaczek

Praca zawiera spis treści, spis rysunków, spis tabel, streszczenie (w jęz. polskim i angielskim), wstęp, pięć rozdziałów (zawierających rezultaty badań) oraz bibliografię składającą się z 90 pozycji. Cała praca zajmuje 94 strony.

W **Rozdziale 1**. podano podstawową definicję krzywej Lorenza oraz jej praktyczną interpretację. Podano też podstawowe własności krzywej Lorenza dołączając ich dowód (Tw. 1.1.1). W Tabeli 1.1 przedstawiono zestaw równań krzywych Lorenza dla wybranych rozkładów prawdopodobieństwa. W Podrozdziale 1.2 podano definicje (Prendergasta i Staudte'go) kwantylowych wersji krzywej Lorenza. Przedstawiono dowód niezmienniczości i monotoniczności omawianych trzech kwantylowych wersji (Tw. 1.2.1). Wykazano też relacje nierówności między tymi kwantylowymi wersjami krzywej Lorenza. W Podrozdziale 1.3 wprowadzono pojęcie kwantylowych wersji krzywej Bonferroniego związanych funkcyjnie z kwantylowymi wersjami krzywej Lorenza. Własności kwantylowych wersji krzywej Bonferroniego nie były dotychczas badane. W Podrozdziale 1.4 rozpatruje się miary koncentracji rozkładu w oparciu o krzywą Lorenza i jej wersje kwantylowe.

Rozdaiał 2 poświęcony jest nieparametrycznej estymacji kwantylowych wersji krzywej Lorenza. Jako estymator funkcji kwantylowej $Q(p) = F^{-1}(p)$ oparty na próbie $\mathbf{X}_n = (X_1, \ldots, X_n)$ rozważa się empiryczną funkcję kwantylową

$$\hat{Q}_n^E(p; \mathbf{X}_n) := \inf\{t : \hat{F}_n^E(t; \mathbf{X}_n) \geq p\},$$

gdzie

$$\hat{F}_n^E(t;\mathbf{X}_n) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{(-\infty,t]}(X_i)$$

jest jest dystrybuantą empiryczną z próby $\mathbf{X}_n$ i $\mathbb{I}_A(x) = 1$, gdy $x \in A$ i 0 poza tym.

Problem dotyczący estymacji funkcji kwantylowej znajduje wiele miejsca w światowej literaturze, również ze względu na duże znaczenie w praktycznych obszarach wiedzy. W rozprawie poszukuje się nowych konstrukcji estymatorów funkcji kwantylowej, które posiadałyby możliwie najwięcej z wymaganych własności w porównaniu ze znanymi w literaturze estymatorami. Sformułowano następujące pożądane własności estymatorów funkcji kwantylowych:

(W1) dla każdej realizacji $\mathbf{x}_n$ próby $\mathbf{X}_n$, $\hat{Q}_n(p)$ jest funkcją ciągłą p,

(W2) dla każdej realizacji $\mathbf{x}_n$ próby $\mathbf{X}_n$ i dla każdego $p \in (0,1)$,

$$\sum_{i=1}^{n} I_{(-\infty,\hat{Q}_n(p)]}(x_i) \geq np,$$

(W3) dla każdej realizacji $\mathbf{x}_n$ próby $\mathbf{X}_n$ i dla każdego $p \in (0,1)$,

$$\sum_{i=1}^{n} I_{(-\infty,\hat{Q}_n(p)]}(x_i) = \sum_{i=1}^{n} I_{[\hat{Q}_n(1-p),\infty)}(x_i),$$

(W4) dla każdej realizacji $\mathbf{x}_n$ próby $\mathbf{X}_n$ i dla każdego $k \in \{1,\ldots,n\}$,

$$\hat{Q}_n^{-1}(x_{k:n}) + \hat{Q}_n^{-1}(x_{(n-k+1):n}) = 1,$$

(W5) dla każdej realizacji $\mathbf{x}_n$ próby $\mathbf{X}_n$,

$$\hat{Q}_n^{-1}(x_{1:n}) > 0 \text{ i } \hat{Q}_n^{-1}(x_{n:n}) < 1,$$

(W6) $\hat{Q}_n(0.5)$ jest równy medianie próbkowej w postaci $(X_{l:n} + X_{(l+1):n})/2$, dla $n = 2l$, i $X_{(l+1):n}$, dla $n = 2l+1$, gdzie l jest liczbą naturalną.

Znane estymatory funkcji kwantylowej spełniają tylko niektóre z tych własności. Na przykład, estymator $\hat{Q}_n^E(p;\mathbf{X}_n)$ oparty na dystrybuancie empirycznej spełnia jedynie własności (W2), (W7) i dla n nieparzystego – własność (W6), tak jak w przypadku estymatora opartego na dystrybuancie empirycznej *level crossing*.

Zaproponowano nowy estymator będący modyfikacją estymatora funkcji kwantylowej opartego na jądrowym estymatorze dystrybuanty

$$\hat{F}_n^K(t) = \frac{1}{n}\sum_{j=1}^{n} \mathcal{K}\left(\frac{t - X_j}{h_n}\right),$$

gdzie $\mathcal{K}(v) = \int_{-\infty}^{v} K(z)dz$, K jest funkcją jądrową, przyjmując losową szerokość okna h_n. Estymator ten nie spełnia jedynie własności (W1) i własności (W6) dla parzystego n.

Zaproponowano również nowy estymator oparty na ciągłej i ściśle rosnącej wersji dystrybuanty empirycznej $\hat{F}_n^{JP}$ skonstruowanej w pracy Jokiel-Rokity i Pulita (2013). Udowodniono, że estymator funkcji kwantylowej $\hat{Q}_n^{JP}(p; \mathbf{X}_n)$ oparty na dystrybuancie $\hat{F}_n^{JP}$ spełnia własności (W1), (W2), (W5) i (W7).

Wprowadzając określoną w Podrozdziale 2.1.5 modyfikację $\hat{F}_n^M$ estymatora dystrybuanty $\hat{F}_n^{JP}$ udowodniono Twierdzenie 2.1.1 stanowiące o tym, że estymator $\hat{Q}_n^M(p; \mathbf{X}_n)$ funkcji kwantylowej, oparty na dystrybuancie $\hat{F}_n^M$ spełnia wszystkie własności (W1) – (W7). Wyznaczono również (Fakt 2.1.7) relację między estymatorem $\hat{Q}_n^M(p; \mathbf{X}_n)$ a estymatorem opartym na dystrybuancie empirycznej $\hat{Q}_n^E(p; \mathbf{X}_n)$.

W Podrozdziale 2.1.6 wykorzystując nierówność Hoefdinga wykazano (Tw. 2.1.3 i Wniosek 2.1.1), że rozpatrywane w rozprawie estymatory funkcji kwantylowej, dla każdego ustalonego $p \in (0, 1)$, są mocno zgodnymi estymatorami.

Udowodniono również, że rozpatrywane estymatory funkcji kwantylowej są asymptotycznie normalne o średniej równej kwantylowi x_p, rozkładu określonemu przez funkcję gęstości f, i wariancji $\frac{p(1-p)}{f^2(x_p)n}, p \in (0, 1)$ (Tw. 2.1.4 i wnioski 2.1.3 – 2.1.5).

W Podrozdziale 2.2 rozpatruje się estymatory typu *plug-in* kwantylowych wersji krzywej Lorenza, tzn. uzyskane poprzez zastąpienie nieznanych kwantyli odpowiednich rzędów nieparametrycznymi estymatorami tych kwantyli. Przy naturalnych założeniach wykazano mocną zgodność tych nieparametrycznych estymatorów w każdym punkcie, jak również zbieżność w normie supremum.

W Podrozdziale 2.3 wprowadzono estymatory kwantylowych wersji indeksu Giniego w oparciu o empiryczne estymatory kwantylowych wersji krzywej Lorenza i wykazano mocną zgodność tych estymatorów.

W **Rozdziale 3.** wyznaczono postać estymatorów typu plug-in kwantylowych wersji krzywej Lorenza w przypadku rozkładu Pareto typu II.

Celem przeprowadzonych symulacji w **Rozdziale 4.** jest numeryczne porównanie dokładności proponowanych w rozprawie estymatorów kwantylowych wersji krzywej Lorenza. Podano wnioski wskazując preferowane konstrukcje estymatorów. Ilustracją uzyskanych rezultatów są również przedstawione w **Rozdziale 5.** wyniki dotyczące wartości estymatorów kwantylowych wersji indeksu Giniego i estymatorów plug-in dla pewnych konkretnych rzeczywistych danych finansowych.

Uważam, że rezultaty prezentowane w rozprawie doktorskiej mgr Agnieszki Siedlaczek są nowe, interesujące i stanowią ważne uzupełnienie wiedzy w zakresie badań związanych z estymacją kwantylowych wersji krzywej Lorenza, zarówno z teoretycznego punktu widzenia jak i zastosowań praktycznych, np. w ekonomii, finansach, socjologii. W dowodach Autorka wykorzystuje umiejętnie zaawansowane metody, twierdzenia i fakty statystyki matematycznej.

Rezultaty zawarte w Podrozdziałach 2.1 i 2.2 rozprawy zostały opublikowane w 2 pracach (pozycje [36] i [37] w Bibliografii).

W rozprawie cytowana jest obszerna literatura.

Pod względem redakcyjnym rozprawa napisana jest bardzo dobrze używając poprawnie języka polskiego. Usterki pod tym względem są bardzo nieznaczne, np. "Udowodnione są", str. 8 (lepiej "Udowodniono"; na str. 54 lepiej byłoby podać odnośnik do strony 13-tej, na której sformułowane są założenia (A1) i (A2); lepiej byłoby użyć powszechnie stosowanego oznaczenia na zbieżność z prawdopodobieństwem 1 jakim jest p.w. (prawie wszędzie), zdecydowanie bardziej niż P1.

Na podstawie przeprowadzonej analizy dorobku naukowego zawartego w przedłożonej rozprawie stwierdzam, że rozprawa doktorska mgr Agnieszki Siedlaczek spełnia warunki określone w obowiązującej ustawie *o stopniach i tytule naukowym oraz o stopniach i tytule w zakresie sztuki* (art. 13 ust.1). Wnioskuję o dopuszczenie mgr Agnieszki Siedlaczek do dalszych etapów przewodu doktorskiego.

Prof. dr hab. Ryszard Magiera

Ryszard Magiera