Lublin, 5.04.2023 r.

Streszczenie rozprawy doktorskiej mgr Luizy Pańczyk

Temat rozprawy: Wybór optymalnych L–statystyk w estymacji kwantyli

W rozprawie doktorskiej zajmujemy się klasycznym problemem estymacji kwantyli przez statystyki porządkowe z próby losowej oraz ich liniowe kombinacje, zwane L-statystykami. Dla próby ustalonego rozmiaru wyznaczone zostały najlepsze estymatory kwantyli w postaci pojedynczej statystyki porządkowej oraz w postaci kombinacji liniowej dwóch sąsiednich statystyk porządkowych. W znanej literaturze poświęconej temu problemowi wybór odpowiednich L-statystyk jako estymatorów kwantyli jest uzasadniany intuicyjnie. W pracy zaproponowane zostało całkowicie nowe podejście do problemu poprzez wprowadzenie nowego kryterium optymalności L-statystyki jako estymatora kwantyla. Kryterium to opiera się na zastosowaniu nieparametrycznych górnych i dolnych oszacowań obciążenia estymacji kwantyli. Wyznaczenie wartości tych oszacowań jest jednym z głównych wyników rozprawy. Zastosowanie nowego kryterium pozwala na wyznaczenie względnie prostych reguły wyboru jak najlepszych estymatorów oraz dowodzić analitycznie ich optymalności, a uzyskane estymatory mogą być łatwo zastosowane w praktyce. Podejście do problemu zaprezentowane w rozprawie jest w pełni nieparametryczne, a więc zostały wyznaczone estymatory, które optymalnie szacują kwantyl zadanego rzędu w sytuacji, gdy nie mamy żadnej uprzedniej wiedzy co do rozkładu badanej wielkości losowej. Dodatkowo, praca zawiera liczne przykłady obliczeń numerycznych, które ilustrują otrzymane wyniki i pozwalają na analizę własności wprowadzonych estymatorów.

W pierwszym rozdziale omówione zostały szczegółowo najważniejsze pojęcia rozważane w rozprawie, tj. kwantyle, funkcje kwantylowe, statystyki porządkowe oraz L–statystyki. Podano również liczne przykłady L–statystyk stosowanych w literaturze jako estymatory kwantyli.

W Rozdziale 2 wprowadzono nowe kryterium optymalności L–statystyk jako estymatorów kwantyla zadanego rzędu $p \in (0,1)$ przy ustalonym rozmiarze próby n. Kryterium to oparte jest na optymalnych oszacowaniach obciążenia estymacji kwantyla rzędu p przez zadaną L–statystykę i ma jasne uzasadnienie intuicyjne.

W kolejnym rozdziale wyznaczone zostały jawne wzory na górne i dolne oszacowania obciążenia estymacji funkcji kwantylowych przez pojedynczą statystykę porządkową. W szczególności wykazano, że dla ustalonych wartości $2 \leq j \leq n-1$ istnieje przedział (θ_j, ξ_j) zawierający liczbę $\frac{j}{n}$ taki, że dla dowolnego rozkładu F o skończonej wariancji σ_F^2 oraz dla dowolnego $p \in (\theta_j, \xi_j)$ zachodzą nierówności

$$-\frac{F_{j:n}(p)}{\sqrt{p(1-p)}} \leq \frac{\mathbb{E}_F X_{j:n} - F^{\rightarrow}(p)}{\sigma_F} \leq \frac{\mathbb{E}_F X_{j:n} - F^{\leftarrow}(p)}{\sigma_F} \leq \frac{1-F_{j:n}(p)}{\sqrt{p(1-p)}}.$$

Ponadto, oszacowania te są osiągnięte dla odpowiednio dobranych rozkładów dwupunktowych. Dla wartości p leżących poza tym przedziałem, albo lewa albo prawa strona w powyższej nierówności wyraża się bardziej skomplikowanym wzorem.

W Rozdziale 4 stosując kryterium wprowadzone w Rozdziale 2 oraz wartości oszacowań wyznaczone w Rozdziale 3, wyznaczono optymalny estymator kwantyla zadanego rzędu $p \in (0,1)$ w postaci pojedynczej statystyki porządkowej. Dokładnie mówiąc, udowodnimy, że istnieją jednoznacznie określone liczby $a_{j,n}$, $1 \leq j < n$, takie, że jeżeli $p \in (a_{j-1,n}, a_{j,n})$, to optymalnym wyborem jest $X_{j:n}$, a jeżeli $p = a_{j,n}$, to obydwie statystyki $X_{j:n}$ oraz $X_{j+1:n}$ są równie dobre. Ponadto liczby te spełniają warunki $a_{j,n} < \frac{j}{n} < a_{j+1,n}$ dla $1 \leq j < \frac{n}{2}$ oraz $a_{n-j,n} = 1 - a_{j,n}$. W szczególności, w odróżnieniu od klasycznego wyboru najlepszym estymatorem kwantyla $x_{k/n}$ jest $X_{k:n}$ jeżeli $1 \leq k < \frac{n}{2}$ lub $X_{k+1:n}$ jeżeli $\frac{n}{2} < k \leq n$. W przypadku gdy $k = \frac{n}{2}$, obie statystyki $X_{n/2:n}$ oraz $X_{n/2+1:n}$ są równie dobrymi estymatorami. W rozdziale tym rozważone zostało również inne kryterium optymalności pojedynczej statystyki porządkowej oparte o minimalizację maksymalnego obciążenia. Pokazano, że dla większości wartości parametru p obydwa rozważane kryteria są równoważne.

W kolejnym rozdziale ponownie zastosowano kryterium optymalności z Rozdziału 2, ale tym razem wyznaczony został najlepszy estymator kwantyla w postaci kombinacji liniowej dwóch sąsiednich statystyk porządkowych. Wyznaczone zostały wartości liczb $b_{k,n}$, $1 \leq k \leq n$, takie że jeśli $p \in (b_{k,n}, b_{k+1,n})$, to optymalnym estymatorem kwantyla x_p jest $(1-\alpha_{n,p})X_{k:n} + \alpha_{n,p}X_{k+1:n}$. Współczynnik $\alpha_{n,p}$ jest wyznaczony jednoznacznie tak, że $\alpha_{n,p} \in \left(0, \frac{1}{2}\right)$ wtedy i tylko wtedy, gdy $p \in (b_{k,n}, a_{k,n})$, gdzie liczby $a_{k,n}$ są opisane powyżej. Otrzymany wynik jest zgodny z klasyczną definicją mediany z próby. Należy podkreślić, że w przeciwieństwie do znakomitej większości estymatorów znanych w literaturze uzyskany estymator nie jest liniową interpolacją funkcji kwantylowej.

W ostatnim rozdziale przeanalizowany został dodatkowo błąd średnio–kwadratowy otrzymanych estymatorów, w celu porównania ich z estymatorami znanymi z literatury. Niestety, z uwagi na bardzo skomplikowane wyrażenia analityczne, większość wyników tego rozdziału to jedynie przykłady numeryczne.