Zbigniew Ciesielski i Ryszard Zieliński
Instytut Matematyczny PAN

Wygładzanie dystrybuanty empirycznej

Rozpatrujemy nieparametryczny model statystyczny z rodziną $\mathcal{F}$ wszystkich ciągłych dystrybuant. Na podstawie próby $X_1, \ldots, X_n$ z rozkładu o pewnej nieznanej dystrybuancie $F \in \mathcal{F}$ chcemy oszacować F. Standardowym estymatorem jest dystrybuanta empiryczna

$$F_n(x) = n^{-1} \sum_{j=1}^{n} 1_{(-\infty,x]}(X_j).$$

Dokładność estymatora jest opisana nierównością Dvoretzkiego-Kiefera-Wolfowitza (DKW)

$$P_F\{||F_n - F||_\infty > \varepsilon\} < 2\exp\{-2n\varepsilon^2\},$$

która pozwala na wyznaczenie takiego n, żeby błąd estymacji z zadanym prawdopodobieństwem równym $2\exp\{-2n\varepsilon^2\}$ nie przekraczał zadanej liczby ε. Nierówność DKW pozwala również obliczać przedziały ufności dla F.

Ale F_n jest funkcją schodkową, więc wydaje się, że jej wygładzenie przynajmniej do funkcji ciągłej będzie w bardziej naturalny sposób odpowiadało estymacji ciągłej funkcji F. Okazuje się jednak, że po wygładzeniu za pomocą jąder, wielomianów lub splajnów uzyskanie nierówności typu DKW napotyka na trudności.

W naszej pracy konstruujemy kawałkami wielomianowy estymator $F_{m,n}$ klasy $C^m(R)$, dla którego nierówność DKW działa w całej przestrzeni $\mathcal{F}$ wszystkich ciągłych dystrybuant; tutaj m jest parametrem wybieranym przez statystyka.

Ponadto dla pewnych innych interesujących estymatorów wielomianowych i splajnowych podajemy podrodziny rodziny $\mathcal{F}$ dystrybuant, w których spełniona jest nierówność typu nierówności DKW.