

Adaptacyjny test dla semiparametrycznego modelu liniowej regresji

Przedstawimy nową konstrukcję testu dla orzekania zgodności rozkładu obserwacji z następującym modelem regresji

$$\mathcal{M}_0 : Y = \beta[v(X)]^T + \epsilon, \quad (1)$$

gdzie X i ϵ są niezależnymi zmiennymi losowymi o nieznanymi gęstościach g i f , odpowiednio. Zakładamy, że X przyjmuje wartości w $[0, 1]$, ϵ w R , $E_f \epsilon = 0$, $E_f \epsilon^2 < \infty$. Ponadto, $\beta \in R^q$ jest wektorem nieznanymi parametrów, a $v(x) = (v_1(x), \dots, v_q(x))$ jest wektorem znanych, liniowo niezależnych, funkcji.

Konstrukcja testu polega na połączeniu metod klasycznych z nowoczesnymi metodami semiparametrycznej estymacji i doboru modelu.

Klasyczna, głównie analityczna, część składa się z dwóch kroków.

- W pierwszym kroku zanurzamy hipotetyczny model (1) w rozszerzony model

$$\mathcal{M}_k(\theta) : Y = \theta[u(X)]^T + \beta[u(X)]^T + \epsilon, \quad (2)$$

gdzie, dla ustalonego k , $k \geq 1$, $\theta \in R^k$ jest wektorem nieznanymi parametrów, a $u(x) = (u_1(x), \dots, u_k(x))$ jest wektorem znanych, liniowo niezależnych, funkcji. Gęstość $Z = (X, Y)$ w rozszerzonym modelu (2) ma postać

$$p(z; \kappa) = g(x)f(y - (u(x), v(x))(\theta, \beta)^T), \quad \kappa = (\theta, \beta, f, g). \quad (3)$$

Testowanie zgodności z \mathcal{M}_0 , w klasie modeli (2), sprowadza się do testowania hipotezy $\theta = 0$, $0 \in R^k$, dla gęstości (3), przy zakłócających parametrach (β, f, g) .

- W drugim kroku wyznacza się efektywny test wynikowy dla tego problemu. Ta konstrukcja składa się z dwóch etapów.

Po pierwsze, liczy się odpowiednie pochodne $p(z; \kappa)$, po κ , dla κ z odpowiednio dobranej przestrzeni Banacha. Te pochodne zdefiniowane są przez wektor, zwany wektorem funkcji wynikowych.

Po drugie, liczy się rzuty składowych wektora funkcji wynikowych, związanych z θ , na podprzestrzeń rozpiętą przez pozostałe składowe, związane z parametrami zakłócającymi β, f, g .

Ta analityczna część pracy zawarta jest w raporcie [1]. Powyższe dwa kroki konstrukcji prowadzą do prototypu statystyki testowej. Oznaczając będziemy go przez $W_k(\beta, f, g)$. Dla uzyskania właściwego rozwiązania opracowuje się następne elementy konstrukcji, wykorzystując głównie narzędzia statystyki i rachunku prawdopodobieństwa. Te etapy konstrukcji można streścić następująco.

- Używając metod wypracowanych w teorii semiparametrycznej estymacji, proponuje się estymatory $\hat{\beta}$, \hat{f} i \hat{g} nieznanymi parametrów β, f, g oraz innych wielkości z nimi związanych. Istota tego kroku polega na sprytnym doborze estymatorów, eliminującym wpływ parametrów zakłócających na rozkład statystyki testowej. Dokładniej,

trzeba zagwarantować, że asymptotyczny rozkład $W_k(\hat{\beta}, \hat{f}, \hat{g})$ jest, przy hipotezie zerowej, taki sam jak asymptotyczny rozkład prototypu $W_k(\beta, f, g)$.

- Stosując metody doboru modeli decyduje się, który z modeli $\mathcal{M}_k(\theta)$ powinien być użyty do konstrukcji prototypu $W_k(\beta, f, g)$.

Niech $D = D(\hat{\beta}, \hat{f}, \hat{g})$ oznacza wymiar k modelu $\mathcal{M}_k(\theta)$ wybrany, na podstawie obserwacji, poprzez regułę doboru. Adaptacyjna wersja statystyki testowej jest zdefiniowana przez

$$W_D(\hat{\beta}, \hat{f}, \hat{g})$$

W pracy [2] zaproponowano dwie reguły doboru modelu i odpowiadające im statystyki testowe. Wyniki teoretyczne pokazują, że konstrukcja rzeczywiście eliminuje wpływ parametrów zakłócających na asymptotyczny rozkład statystyk testowych, przy prawdziwości hipotezy zerowej. Wynik ten pozwala na kontrolę poziomu istotności nowych testów. Wyniki symulacyjne wskazują na dużą czułość nowych testów w wykrywaniu rozmaitych odstępstw od hipotezy zerowej.

Literatura

- [1] Inglot, T., Ledwina, T. (2003). *Semiparametric regression: Hadamard differentiability and efficient score functions for some testing problems*. Preprint 012. Instytut Matematyki Politechniki Wrocławskiej.
- [2] Inglot, T., Ledwina, T. (2005). *Data driven score tests of fit for semiparametric homoscedastic linear regression model*. Przesłana do druku.