mgr inż. Małgorzata Żak-Szatkowska
Politechnika Wrocławska
Instytut Matematyki i Informatyki

Rangowe kryterium wyboru modelu w lokalizacji genów

Zagadnienie lokalizacji genów wpływających na cechę ilościową, tzw. QTLi, można sprowadzić do problemu wyboru modelu wielokrotnej regresji liniowej. Budując model regresji decydujemy, ile i które zmienne do niego włączymy. Możemy w tym celu wykorzystać kryteria wyboru modelu, jak np. AIC lub BIC.

W przypadku lokalizacji genów zbiór danych zawiera na ogół kilkaset lub kilka tysięcy zmiennych, spośród których chcemy wybrać tylko kilka. W takim przypadku klasyczne kryteria przeszacowują liczbę QTLi. Zaproponowana w [1] modyfikacja, mBIC, dzięki dodatkowej karze za wymiar modelu jest przystosowana do takiej sytuacji.

Kryterium mBIC zostało skonstruowane przy standardowym założeniu, że szum ma rozkład normalny. W przypadku, gdy rozkład ten ma ciężkie ogony albo pewien procent danych to obserwacje odstające, wyniki uzyskiwane przez BIC oraz mBIC znacznie się pogarszają.

W tej prezentacji przedstawimy nową, rangową wersję kryterium mBIC. Kryterium rBIC, zaproponowane w [2], pozwala na kontrolę błędu FWER (the family wise error rate), a także osiąga dobrą moc w lokalizacji QTLi, niezależnie od rozkładu, z jakiego pochodzi szum. Kryterium sprawdza się również w przypadku dużych zbiorów danych niegenetycznych.

Bibliografia

[1] M. Bogdan, J. K. Ghosh, R. W. Doerge, *Modifying the Schwarz Bayesian Information Criterion to locate multiple interacting quantitative trait loci*, Genetics 167 (2004), 989–999.

[2] M. Żak, A. Baierl, M. Bogdan, A. Futschik, *Locating multiple interacting quantitative trait loci using rank-based model selection*, Genetics 176 (2007), 1845–1854.

[3] M. Bogdan, J. K. Ghosh, M. Żak-Szatkowska, *Selecting eplanatory variables with the modified version of the Bayesian Information Criterion*, QREI 24 (2008), 627–641.