

Kilka uwag na temat współczesnej statystyki i roli zastosowań.

Zgodnie z definicją, statystyka to nauka zajmująca się metodami zbierania, organizacji i analizy danych. W ostatnich latach statystyka bardzo prężnie się rozwija w związku z rozwojem technologii pomiarowych i informatycznych, umożliwiających zbieranie i przechowywanie wielkich zbiorów danych. Oprócz niezliczonej liczby rzeczywistych problemów praktycznych pojawia się też niezliczona liczba nowych i trudnych zagadnień pojęciowych i teoretycznych. Rozwój technologii informatycznych umożliwił zarówno powstanie i rozwój nowych, złożonych obliczeniowo, narzędzi statystycznych, jak również weryfikację tych narzędzi za pomocą symulacji komputerowych i technik walidacyjnych.

Od statystyka zajmującego się analizą takich wielowymiarowych danych wymaga się bardzo rozległej wiedzy matematycznej. Już dawno skończyły się czasy, kiedy statystykowi „stosowanemu” wystarczyła dobra znajomość modeli liniowych. W tej chwili statystyk zajmujący się analizą dużych zbiorów danych powinien biegłe poruszać praktycznie w każdym dziale matematyki. Do zupełnie podstawowych wymagań należy biegła znajomość zaawansowanej probabilistyki, algebry liniowej i analizy funkcjonalnej. Oprócz tego w tej chwili wymaga się także np. wiedzy z bardziej informatycznych dziedzin matematyki, jak teoria grafów, metody optymalizacji czy teoria złożoności algorytmów. Ostatnio w Stanford miałam przyjemność wysłuchania wykładu o wykorzystaniu złożonych pojęć z topologii do opisu kształtu i klasyfikacji komórek rakowych. Oprócz tych sporych wymagań od statystyka wymaga się też umiejętności komunikacji ze światem zewnętrznym. W szczególności dobry statystyk powinien umieć komunikować treści matematyczne wychodząc poza hermetyczny język matematyki. Tą cechą posiadają wszyscy czołowi statystycy, których miałam okazję poznać w Stanford, jak David Siegmund, Bradley Efron, Rob Tibshirani czy Emmanuel Candès, którzy publikują zarówno w teoretycznych czasopiśmie statystycznych jak i czasopiśmie genetycznych czy inżynierskich.

Pomimo bardzo rozległej wiedzy matematycznej wymaganej od statystyka wyraźnie widać różnicę między charakterem pracy statystyka a teoretycznego matematyka. Generalnie zadaniem statystyka jest wykorzystanie tej wiedzy do tworzenia narzędzi statystycznych i analizy ich własności. Obecnie czołowe teoretyczne pisma statystyczne jak np. *Annals of Statistics*, dowody matematyczne często publikują w dodatkach a dużo większy nacisk kładą na motywację i „praktyczną” interpretację teoretycznych wyników. Ponadto rozkwita rynek czasopism publikujących nowe metody analizy danych i oprogramowanie, które często są weryfikowane jedynie za pomocą symulacji komputerowych lub różnych technik walidacyjnych (np. na nowych rzeczywistych zbiorach danych lub za pomocą walidacji krzyżowej).

Pomimo tych różnic między matematyką a statystyką rozwój statystyki stosowanej służy rozwojowi matematyki. Zastosowania statystyki pokazują społeczeństwu, że pewne ważne rzeczywiste problemy można efektywnie rozwiązać jedynie posługując się zaawansowanym aparatem matematycznym. W tym miejscu gorąco zachęcam do wysłuchania krótkiego wywiadu z wybitnym statystykiem ze Stanford, Emmanuelem Candèsem, który między innymi opowiada o bardzo ważnym zagadnieniu praktycznym, którego rozwiązanie stało się początkiem rozwoju całej dziedziny tzw. „compressed sensing” (<http://vimeo.com/26792424>). Ponadto, czytelny przekaz o roli matematyki/statystyki w wielu innych dziedzinach nauki i życia codziennego i odpowiednie dostosowanie programów kształcenia daje szansę na rekrutację większej liczby dobrych i przedsiębiorczych studentów na studia matematyczne.

Oczywiście promując rozwój statystyki stosowanej należy bardzo starannie przemyśleć kryteria promocji w tej dziedzinie. W USA od statystyka na poziomie doktoratu wymaga się zaliczenia odpowiedniej liczby zaawansowanych kursów. Moim zdaniem jest to bardzo ważny warunek. Przy obecnym rozwoju statystyki studia magisterskie nie wystarczają nawet do tego żeby dobrze

przygotować „stosowanego” statystyka do pracy z wielowymiarowymi danymi. Poza tym od doktoratu wymaga się „innowacyjności”. Przy czym często ta innowacyjność sprowadza się do modyfikacji znanych metod i zastosowania ich w nowych problemach praktycznych. Tego typu wyniki publikuje się zwykle w specjalistycznych czasopismach z danej dziedziny. Choć charakter takich prac doktorskich zasadniczo odbiega od charakteru prac z klasycznej matematyki, w moim odczuciu powinny być one bronione/klasyfikowane w ramach nauk matematycznych. Rozwijane przez statystyków metody na ogół mogą być stosowane w wielu różnych dziedzinach nauki a badacze je tworzący są specjalistami ze statystyki/matematyki a nie z dziedziny której dotyczy konkretne zastosowanie. Zrozumienie problemu praktycznego, wybór odpowiedniego narzędzia, jego odpowiednia modyfikacja i weryfikacja (niekoniecznie za pomocą wyników teoretycznych) wymaga pokonania wielu trudności i dużej wiedzy matematycznej. Często jest to trudniejsze niż modyfikacja założeń i dowód teoretycznego twierdzenia za pomocą wyuczonej techniki, jak to się zdarza w przypadku doktoratów z teoretycznej matematyki/statystyki.

Promując statystykę stosowaną nie należy jednak zapominać o klasycznej statystyce matematycznej. Głębokie wyniki matematyczne uzyskiwane przez badaczy z tej dziedziny są coraz częściej wykorzystywane do analizy złożonych danych. Jest to problem ważny. Tak naprawdę jest dla mnie całkiem jasne, że statystyka stosowana wkrótce zostanie zauważona i będzie promowana. Takie są oczekiwania społeczne i decyzje rządzących. I wkrótce może się okazać, że w związku z nowymi kryteriami oceny/promocji (Impact Factor etc) klasyczna statystyka matematyczna nie będzie w stanie konkurować z zastosowaniami. Moim zdaniem należy wypracować system który zapewni równowagę między statystyką teoretyczną i stosowaną, bo obie dziedziny są bardzo ważne i są sobie nawzajem bardzo potrzebne.

Z wyrazami szacunku,

Małgorzata Bogdan