

Warszawa, 8 maja 2026

prof. dr hab. Wojciech Niemirow
Instytut Matematyki Stosowanej i Mechaniki
Wydział Matematyki, Informatyki i Mechaniki
Uniwersytet Warszawski

Recenzja rozprawy doktorskiej Wojciecha Wójcika

**pt. „Multi-Domain Optimum Sample Allocation
with Controlled-Precision under Upper-Bound
Constraints”**

Zawartość rozprawy

Rozprawa liczy 148 stron, składa się z 7 rozdziałów, 7 dodatków i bibliografii. Napisana jest w języku angielskim.

Rozdział 1 jest wprowadzeniem do tematyki i przedstawia problem rozpatrywany w pracy. W Rozdziałach 2 i 3 problem z dziedziny statystyki jest reformułowany jako abstrakcyjny problem optymalizacyjny. Rozdział 4 zawiera opis skonstruowanego przez Autora algorytmu RDCA, który rozwiązuje postawiony problem. Dalsze dwa rozdziały zawierają szczegółowy dowód poprawności algorytmu. Zamieszczone na końcu Dodatki ułatwiają czytanie rozprawy.

Problem

Rozpatrywany w rozprawie problem (Problem 1.1.1) dotyczy optymalnej alokacji próbek w zadaniu estymacji sum w obszarach (t_d) i sumy całkowitej (t). Populacja jest podzielona na obszary (domains; d), a te z kolei na warstwy (strata; d, h). Z każdej warstwy pobierana jest próbka zgodnie ze schematem SRSWOR (prosta próbka bez zwracania w warstwach) i stosowany jest estymator Horwitza-Thompsona. Zadanie polega na alokacji próbek o całkowitej liczności n do poszczególnych warstw tak, aby zminimalizować względną wariancję

$$V = \text{Var}(\hat{t})/t^2$$

przy ograniczeniach $V = \sum_d \kappa_d V_d$, gdzie wagi κ_d są dane a $V_d = \text{Var}(\hat{t}_d)/t_d^2$. Kluczowe jest dodatkowe ograniczenie $n_{d,h} \leq N_{d,h}$ (liczność próbek nie przekracza liczności warstwy). Bez tego ograniczenia rozwiązanie jest łatwiejsze i wcześniej znane (algorytm DCA, Wieczorkowski i Wesołowski 2017). Również w sytuacji gdy rozpatrujemy populację niepodzielną na obszary, zadanie było wcześniej rozwiązane przy współudziale Autora (algorytm RNA, Wieczorkowski, Wesołowski i Wójcik 2022). Obecna rozprawa doktorska przedstawia algorytm RDCA, który znajduje rozwiązanie problemu w ogólnej postaci.

Ocena wyników

Konstrukcja algorytmu RDCA wraz z dowodem poprawności jest osiągnięciem, które zasługuje na wysoką ocenę. Uzasadnienie tej opinii i szczegółowe uwagi zamieszczam poniżej.

Ocena matematycznej i informatycznej zawartości rozprawy

Chociaż rozprawa dotyczy statystyki matematycznej i w szczególności metody reprezentacyjnej, to probabilistyka odgrywa tu niewielką rolę. Ważna jest teoria optymalizacji wypukłej i dość wyrafinowane pomysły algorytmiczne. Autor rozpatruje zadanie sformułowane elementarnie i używa klasycznych narzędzi matematycznych, przede wszystkim sławnych warunków Kuhna-Tuckera. Punktem wyjścia rozprawy Wójciaka są dwa kluczowe pomysły. Pierwszy to spostrzeżenie (Wieczorkowski i Wesołowski 2017), że zadanie alokacji przy założonej równej dokładności w domenach sprowadza się do obliczenia wektora własnego pewnej macierzy. Drugi to algorytm RNA (Särndal et al. 1992), formalizujący dość intuicyjny sposób uwzględnienia naturalnych ograniczeń w Neymanowskiej alokacji. Wójciak oba te pomysły rozwija i łączy. Na pierwszy rzut oka nie wydaje się to nowatorskie, ale trudność leży w szczegółach, które wymagały od Autora dużej pomysłowości i zostały opracowane niezwykle dokładnie.

Idea algorytmu RDCA polega na wielokrotnym obliczaniu rozwiązania bez ograniczeń $n_{dh} \leq N_{dh}$, „wypełnianiu” tych warstw ($n_{dh} = N_{dh}$), które naruszają ograniczenie i rozwiązywaniu zadania dla pozostałych warstw. Ale naiwna realizacja tej idei nie prowadzi do znalezienia prawidłowego rozwiązania. Powstają problemy związane z blokowaniem obszarów (jeśli $n_d = N_d$ dla jakiegoś obszaru to jedynym rozwiązaniem musi być $n = N$). Autor rozprawy opracował algorytm wykorzystujący w pomysłowy sposób rekurencję do przezwyciężenia trudności (Algorytm 1 w podrozdziale 4.1). Nie jestem kompetentny do oceny informatycznej (algorytmicznej) strony rozwiązania, ale wydaje mi się, że jest ona bardzo istotną wartością. (Nie wiem, czy równoważny lub analogiczny do RDCA algorytm dałoby się sformułować bez użycia rekursji.)

Z punktu widzenia matematycznego, podstawowym osiągnięciem pracy jest szczegółowy i dokładny dowód poprawności algorytmu RDCA (dowód Twierdzenia 6.4.3). Dowód zajmuje 50 stron (Rozdziały 5 i 6). Co prawda, formalnie dowód mieści się na połowie strony 116, ale poprzedzające 50 stron zajmują wyniki pomocnicze i konieczne przygotowania! Dowody zamieszczone w rozprawie są bardzo techniczne. Trudność rozumowań wynika ze skomplikowanej logicznej struktury algorytmu RDCA, co podkreśla przerażający diagram na stronie 88.

Warto zauważyć, że poziom ścisłości rozumowań matematycznych w rozprawie Wojciecha Wójciaka nie jest całkiem powszechny w literaturze dotyczącej „statystyki reprezentacyjnej” (*survey statistics*). Rozprawa jest zatem kolejnym krokiem w kierunku dostosowania teorii *survey statistics* do standardów obowiązujących w głównej nurcie statystyki matematycznej.

Dobre wrażenie robi naszkicowanie w końcowej części rozprawy dalszych kierunków badań. Uwagi Autora są dojrzałe i kompetentne.

Ocena praktycznej wartości rozprawy

Algorytm, zaproponowany przez Autora, rozwiązuje zadanie wyrastające z potrzeb *survey statistics* i będzie miał zastosowanie w praktyce. Wybór optymalnej alokacji poprawia dokładność estymatorów. Co prawda, optymalność otrzymanego rozwiązania jest udowodniona przy nierealistycznym założeniu znajomości szeregu wielkości (w tym $\sum t_d$, które są obiektem estymacji!), ale jest to nieunikniona idealizacja. W praktyce zastępuje się nieznane wielkości przez ich wstępne oszacowania, co prowadzi do alokacji optymalnej tylko w przybliżeniu.

Autor poświęca nieco uwagi usprawiedliwieniu innego przybliżenia w sformułowaniu zadania: mianowicie potraktowania rozmiarów próbek jako liczb rzeczywistych, niekoniecznie całkowitych. Wydaje się, że ten problem jest drugorzędny, a zastosowanie programowania całkowitoliczbowego doprowadziłoby do niepraktycznych algorytmów.

Implementacja w środowisku R

W poprzednich częściach recenzji zwracałem uwagę wartość informatyczną zawartość rozprawy. Tym bardziej godne uznania jest opracowanie pakietu R-owskiego „stratallo” realizującego między innymi algorytm RDCA. Autor rozprawy jest współautorem tego pakietu i domyślam się, że jego wkład jest niebagatelny.

Redakcja rozprawy

Rozprawa jest napisana z ogromną dbałością o szczegóły, o kompletność rozumowań i spójność oznaczeń. W połączeniu ze skomplikowaną strukturą logiczną dowodów dało to w rezultacie tekst obszerny, trudny do czytania i naszpikowany setkami ponumerowanych wzorów. Oznaczenia są ścisłe i konsekwentne, ale nie specjalnie przejrzyste. Autor jednak pomaga czytelnikowi, uzupełniając wzory objaśnieniami i komentarzami w języku naturalnym. Pseudo-kody podstawowych algorytmów RDCA i DCA w Rozdziale 4 są zwięzłe i eleganckie, ale niełatwe do zrozumienia. Następujące po nich objaśnienia pozwalają znaleźć główną ideę, a jeszcze ważniejsza jest przytoczona dalej analiza kilku przykładów (Podrozdział 4.2). Autor zamieścił świetnie obmyślane diagramy umożliwiające śledzenie działania RDCA i zrozumienie rekurencyjnej struktury.

Ciekawym pomysłem jest włączenie do rozprawy wielu dodatków (apendiksów), które w założeniu czynią tekst samowystarczającym i dostępnym dla praktycznie zorientowanych czytelników bez szerokiej wiedzy matematycznej. Napisałem „w założeniu”, bo nie bardzo wierzę aby taki praktyk przeczytał 50 stron dowodu poprawności.

Jedno tylko redakcyjne potknięcie wydało mi się wyjątkowo niefortunne. W Streszczeniu czytamy: *The objective is the simultaneous minimization of the variance of the global total estimator and the variances of domain total estimators under a fixed total sample size, while ensuring that allocations do not exceed the population sizes within strata.* Dalej w Rozdziale 1 ten sam „skrót myślowy” się powtarza uparcie: *The objective is to simultaneously minimize all*

domain-level relative variances V_d , $d \in \{1, \dots, D\}$, as well as the overall relative variance V , subject to a fixed total sample size n and [...]. Gdyby to ktoś czytał (powierzchownie), toby miał wrażenie, że minimalizacja V daje się pogodzić z minimalizacją V_d dla każdego d , co nie jest prawdą (dopóki nie wprowadzimy żądania typu $V_d = \kappa_d V$). To „redakcyjne potknięcie” nie jest formalnie błędem, ale jest nadmiernie uproszczonym „tłumaczeniem języka matematyki na język naturalny”.

Podsumowanie i konkluzja

Wyniki zamieszczone w rozprawie oceniam wysoko i wnoszę o dopuszczenie Pana Wojciecha Wójciaka do dalszych etapów przewodu doktorskiego.

Wojciech Niemirowicz