Prof. Roman Zmyślony

Instytut Matematyki

Uniwersytet Zielonogórski

# Opinia o pracy doktorskiej Moniki Mokrzyckiej

# pt. „Aproksymacja macierzy kowariancji wybranymi strukturami w modelach podwójnie wielowymiarowych"

## 1. Ogólna charakterystyka pracy i uzyskane wyniki.

Praca składa się z abstraktu w języku angielskim, wstępu, zawiera 5 rozdziałów i bibliografię, na którą składa się 39 pozycji, w tym 3 prace współautorskie Moniki Mokrzyckiej. Jedna opublikowana jest w Linear Algebra, a dwie są w druku. Jest współautorką dwóch rozdziałów w książkowym wydaniu pt. Multivariate, Multilinear and Mixed Linear Models wydawnictwa Springer. Abstrakt jest krótkim streszczeniem zawartości pracy, natomiast wstęp nieco szerszy od angielskiego streszczenia, zawiera podsumowanie otrzymanych wyników w poszczególnych rozdziałach.

**Rozdział 1."Wprowadzenie".** W rozdziale tym opisany jest problem wyboru modelu podwójnie wielowymiarowego normalnego dla $n$ niezależnych macierzy $\boldsymbol{Y}_i$ o wymiarach $p \times q$, o kompletnie nieznanej średniej macierzowej $\boldsymbol{M}$ i nieznanej macierzy kowariancji $\boldsymbol{\Omega}$. Wybór modelu polega na dobraniu jednej z trzech struktur macierzy kowariancji, co jest uzasadnione, gdy obserwacje dotyczą pomiarów cech w czasie. Wówczas struktura macierzy kowariancji jest iloczynem Kroneckera $\boldsymbol{\Omega} = \boldsymbol{\Psi} \otimes \boldsymbol{\Sigma}$. Rozważane są trzy typy w zależności od struktury macierzy $\boldsymbol{\Psi}$ natomiast $\boldsymbol{\Sigma}$ jest dowolną macierzą dodatnio określoną. Pierwszy największy zbiór jest oznaczony w pracy jako $\boldsymbol{S}$, gdzie $\boldsymbol{\Psi}$ jest również dowolną macierzą dodatnio określoną. Drugi jest taki, że $\boldsymbol{\Psi}$ ma strukturę kompletnej symetrii, to znaczy $\boldsymbol{\Psi} = (1-\rho)\boldsymbol{I}_p + \rho\boldsymbol{1}_p\boldsymbol{1}_p'$, gdzie $\boldsymbol{I}_p$ oznacza macierz identycznościową natomiast $\boldsymbol{1}_p$ jest kolumnowym wektorem jedynek, oraz w trzecim zbiorze $\boldsymbol{\Psi}$ ma strukturę kowariancji dla procesu autoregresji pierwszego rzędu. Wyboru modelu dokonuje się przy trzech różnych funkcjach straty, nazwane w pracy funkcjami rozbieżności. Pierwsza oparta na normie Frobeniusa, czyli zdefiniowana jest jako ślad kwadratu różnicy $\boldsymbol{\Omega}-\boldsymbol{\Gamma}$. Funkcja ta była rozważana w pracy autorów K. Filipiak i D. Klein. W tej pracy służy do porównania wyników uzyskanych z dwoma innymi funkcjami rozbieżności. Funkcja entropijna zdefiniowana jest jako $\mathrm{tr}(\boldsymbol{\Omega}^{-1}\boldsymbol{\Gamma}) - \ln|\boldsymbol{\Omega}^{-1}\boldsymbol{\Gamma}| - pg$.

RS

Natomiast druga nazywana w pracy jako kwadratowa jest postaci $\mathrm{tr}[(\boldsymbol{\Omega}^{-1}\,\boldsymbol{\Gamma} - \boldsymbol{I}_{pq})^2]$. Dla tych trzech funkcji wybór odpowiedniej struktury definiuje się jako argument minimum odpowiedniej funkcji. Dalsza część tego rozdziału zawiera definicje i własności operatorów macierzowych zawartych w 13-tu lematach, a wykorzystanych w dalszej części rozprawy.

**Rozdział 2. Aproksymacja entropijną funkcją straty"** zawiera trzy twierdzenia dotyczące wyznaczenia minimów dla entropijnej funkcji rozdzielczości w odpowiednich trzech zbiorach struktury macierzy kowariancji. Twierdzenie 2.1 w równaniu (2.2) podaje dwa równania nieliniowe, których rozwiązanie daje dodatnio określoną macierz realizującą minimum funkcji entropijnej na zbiorze $\boldsymbol{S}$. Natomiast Twierdzenie 2.2 w równaniach (2.7) podaje sposób wyznaczenia minimum w zbiorze, w którym struktura macierzy $\boldsymbol{\Gamma} = (1-\rho)\boldsymbol{I}_p + \rho\boldsymbol{1}_p\boldsymbol{1}_p'$, czyli jest kompletnie symetryczna. Kolejne Twierdzenie 2.3 w równaniach (2.21) podaje układ równań, których rozwiązanie gwarantuje dodatnio określony operator kowariancji w zbiorze kowariancji o strukturze procesu autoregresji pierwszego rzędu.

**Rozdział 3. Aproksymacja kwadratową funkcją straty** zawiera analogiczne twierdzenia do poprzedniego rozdziału dla funkcji $\mathrm{tr}[(\boldsymbol{\Omega}^{-1}\,\boldsymbol{\Gamma} - \boldsymbol{I}_{pq})^2]$. Główne wyniki zawarte są w Twierdzeniu 3.1, przy czym minimum w zbiorze $\boldsymbol{S}$ jest rozwiązaniem układu dwóch nieliniowych równań macierzowych opisanych w (3.2) i (3.3). Dla struktury macierzy $\boldsymbol{\Psi} = (1-\rho)\boldsymbol{I}_p + \rho\boldsymbol{1}_p\boldsymbol{1}_p'$ (kompletnie symetrycznej) argumenty minimum podaje Twierdzenie 3.2, jako rozwiązania układu równań na $\rho$ i $\boldsymbol{\Sigma}$, a podane są w (3.5) i (3.6). Kolejne Twierdzenie 3.3, po rozwiązaniu równań nieliniowych podanych w pracy jako (3.8) i (3.9), podaje sposób wyznaczenia optymalnego doboru macierzy $\boldsymbol{\Psi}$, w zbiorze macierzy o strukturze procesu autoregresji pierwszego rzędu i $\boldsymbol{\Sigma}$ w zbiorze wszystkich dodatnio określonych macierzy.

**Rozdział 4. Badania symulacyjne oraz zastosowanie minimum funkcji rozbieżności do danych rzeczywistych** stanowi istotną część badań własności estymatorów macierzy kowariancji, w tym również zastosowanie do danych rzeczywistych. Badania symulacyjne prowadzono z jednej strony celem właściwego rozpoznania modelu na podstawie symulacji odpowiednich struktur macierzy kowariancji, a z drugiej strony badano moce testów oparte na estymatorach wyznaczonych przez minima tych funkcji. Istotą optymalnej symulacji (chodzi o skrócenie czasu wykonanych symulacji) jest wykorzystanie własności operatorów różniczkowania jak i niezmienniczości funkcji, oraz związku między statystykami testowymi MLE (największej wiarygodności), ELE (minimum entropijnej funkcji) oraz QLE (minimum kwadratowej funkcji).

Wyniki tych symulacji podane są w różnych tabelach. W Tabelach 4.1, 4.2, 4.3, 4.4 podane są uśrednione rozbieżności oraz procent poprawnie zidentyfikowanych struktur dla wybranych parametrów. Natomiast w Tabelach 4.5-4.11 podane są wartości estymatorów i ich błędy standardowe dla $\rho$, MLE, ELE i QLE dla modeli z macierzą kowariancji kompletnie symetryczną i strukturę kowariancji dla procesu autoregresji pierwszego rzędu dla wybranych generowanych rozkładów przy różnych parametrach $p, q$ , $\rho$. Z kolei w Tabelach 4.12 i 4.13 podano analogiczne wyniki symulacji dla normy Frobeniusa estymatora $\boldsymbol{\Sigma}$. Ostatnia część tego rozdziału dotyczy testowania hipotezo opartych na zmodyfikowanych rozbieżnościach polegającym na przekształceniu monotonicznym wartości funkcji rozbieżności na odcinek [0,1). Moce testu badano dla dwóch zmodyfikowanych losowo wybranych macierzy z hipotezy alternatywnej tak, aby miały rozbieżność 0.1 i 0.5, przy czym jedna z nich miała dwa elementy różne od wygenerowanej macierzy spełniającej hipotezę zerową. Wybrane moce dla testów LRT i RST w zależności od poziomu rozbieżności entropijnej przedstawiono odpowiednio w Tabelach 4.14 i 4.15, natomiast w Tabelach 4.16 i 4.17 moce te oparte są na normie kwadratowej, a w Tabelach 4.18-4.19 podano moce testów odpowiednio LRT i RST ze względu na skorygowane rozbieżności normy Frobeniusa. Dodatkowo, na wykresach (patrz Rysunki 4.1-4.2) porównano moce testów LRT i RST dla entropijnej i kwadratowej funkcji straty w celu wykrycia ich wpływu na moce testów. Krzywe funkcji mocy przecinają się, zatem nie można jednoznacznie wywnioskować, których z nich jest testem mocniejszym na całym analizowanym zbiorze. Przy okazji badań symulacyjnych wykryto błąd symulacyjnego obliczania mocy testu w pracy z 2005 roku (patrz [22]). W końcowym podrozdziale **4.5 Dane rzeczywiste** przedstawione są wyniki w Tabeli 4.22, w której to podano empiryczne p-wartości dla hipotez dla trzech cech jak i ich dwuelementowych podzbiorów.

**Rozdział 5. Podsumowanie** zawiera krótkie podsumowanie otrzymanych wyników i wnioski dotyczące stosowania ich dla danych rzeczywistych.

## 2. Ocena wyników rozprawy.

Praca wnosi istotny wkład do teorii estymacji i testowania hipotez w wielowymiarowych rozkładach normalnych. Autor uzyskał nowe wyniki teoretyczne z zakresu wyboru modelu spośród trzech struktur macierzy kowariancji. Słusznie dokonuje się wyboru modelu w oparciu o minimalne statystyki dostateczne. Statystyką ta jest jednocześnie estymatorem uzyskanym metodą największej wiarygodności w modelu z dowolną dodatnio określoną nieznaną macierzą kowariancji (brak tej uwagi w pracy). W uzyskaniu w/w wyników istotna była biegła znajomość i stosowanie formuł różniczkowania funkcji wektorowych i macierzowych w problemach statystycznych. Ponadto, autor pracy zaprezentował wykorzystanie wyników pracy do rzeczywistych danych. Z powyższych uwag i biorąc pod uwagę wysoką merytorycznie ocenę pracy wnoszę o jej wyróżnienie.

### 3. Konkluzja.

Z uwagi na wyżej przedstawioną opinię, stwierdzam, że rozprawa spełnia wymagania odnośnej ustawy o stopniach naukowych i tytule naukowym. W związku z powyższym wnoszę o dopuszczenie mgr Moniki Mokrzyckiej do dalszych etapów przewodu doktorskiego.

Roman Zmyślony

Roman Zmyślony