Recenzja pracy doktorskiej mgr Agnieszki Szumery

Probabilistyczne struktury regresji

Rozprawa doktorska "Probabilistyczne struktury regresji" mgr Agnieszki Szumery zawiera rozważania na temat analizy regresji rozumianej jako badanie zależności pomiędzy wielkościami podlegającymi obserwacji (które są tradycyjnie nazywane zmiennymi objaśniającymi) oraz wielkościami ocenianymi na podstawie wartości zmiennych obserwowanych. Głównym zadaniem analizy regresji jest znalezienie możliwie najlepszego modelu teoretycznego, czyli zależności funkcyjnej która dobrze odzwierciedla związki między różnymi rodzajami zmiennych. Praca została zainspirowana wynikami badań zawartymi w artykule *Generalized approach to the problem of regression*, którego autorami są D. Partyka i J. Zając, poz. [25] Bibliografii. W odróżnieniu od badań opisanych w wymienionym artykule, poświęconych głównie asynchronicznym strukturom regresji, czyli związkom między funkcjami zależnymi od różnych zmiennych, rozważania przedstawione w omawianej pracy doktorskiej dotyczą tzw. synchronicznych (probabilistycznych) struktur regresji $\mathfrak{P}$, w których badany jest związek między funkcjami zależnymi od tej samej zmiennej przebiegającej przestrzeń probabilistyczną.

Metody badań problemu regresji zastosowane w rozprawie doktorskiej mgr A. Szumery różnią się od klasycznego ujęcia tego zagadnienia z kilku względów:

1^0 w klasycznym ujęciu problemu regresji badana jest najczęściej regresja liniowa (tzn. poszukiwane jest najlepsze przybliżenie nieznanych wartości zmiennej y poprzez wartości dowolnej funkcji liniowej $ax + b$ zmiennej objaśniającej x, gdzie a, b są stałymi), lub badana jest tzw. regresja pierwszego rodzaju (tzn. poszukiwane jest możliwie najlepsze przybliżenie wartości zmiennej y przy pomocy wartości dowolnej funkcji $f(x)$, zwykle borelowskiej) na podstawie otrzymanych doświadczalnie wartości zmiennej x. W omawianej pracy doktorskiej opisane zostały różne własności rodziny funkcji regresji (tj. funkcji najlepszego przybliżenia) minimalizujących średniokwadratowy błąd oceny w obszernych klasach funkcji $\mathcal{F}$ spełniających dość ogólne założenia (zamkniętych ze względu na kombinacje liniowe funkcji), a więc rozważania mogą np. dotyczyć regresji w klasie wielomianów stopnia $\leq p < \infty$, regresji w klasie funkcji potęgowych lub wykładniczych zmiennej x z różnymi parametrami, regresji w klasie funkcji trygonometrycznych, a także regresji w klasie funkcji mierzalnych względem ustalonego σ-ciała, itp.

2^0 zamiast przybliżania nieznanych surowych wartości zmiennej y, w pracy badana jest regresja dla wartości $g(y)$ w probabilistycznych synchronicznych strukturach regresji $\mathfrak{P} = \mathfrak{P}_g$, przy czym zmienna objaśniająca x oraz zmienna y są określone na tej samej przestrzeni probabilistycznej $(\Omega, \mathcal{A}, P)$, a przekształcenie g jest traktowane jako funkcja skalująca surowe wartości nieznanej zmiennej y,

3^0 zamiast poszukiwania pojedynczych elementów będących najlepszym przybliżeniem, które w przypadku minimalizowania średniokwadratowego błędu przybliżonej oceny faktycznie są klasami równoważności funkcji równych prawie wszędzie, w pracy opisywane są własności rodziny funkcji regresji $\mathrm{Reg}(\mathcal{F}, \mathfrak{P}_g) \subset \mathcal{F}$ w zależności od klas funkcji $\mathcal{F}$ w których poszukiwane jest najlepsze przybliżenie wartości $g(y)$ minimalizujące błąd średniokwadratowy

$$\delta(f \circ x, g \circ y) = \int_{\Omega} |f \circ x - g \circ y|^2 \, dP, \quad f \in \mathcal{F}.$$

Rozszerzając informacje podane w p. 2^0 warto podkreślić, że wprowadzenie funkcji g skalującej surowe wyniki zmiennej zależnej y jest dość interesującym zabiegiem, gdyż pozwala wyrazić regresję w sposób uwikłany. Ponadto wprowadzenie funkcji g umożliwia udowodnienie głównego rezultatu pracy, jakim jest Twierdzenie 3.8, na którym opiera się większość dalszych rozważań. W szczególności, na podstawie Twierdzenia 3.8 w pracy znaleziony został m.in. ogólny wzór dla rodziny funkcji regresji $\mathrm{Reg}(\mathcal{F}, \mathfrak{P}_g)$ (Tw. 3.15), a następnie wzór dla rodziny funkcji regresji w ramach klasy $\mathcal{F}$ funkcji rozpiętych na skończonym zbiorze elementów (Tw. 4.11). Warto jednak uświadomić sobie, że funkcja g nie może być całkowicie dowolna, w zastosowaniach praktycznych powinna to być raczej funkcja różnowartościowa (chociaż takie założenie nie jest potrzebne w dowodach twierdzeń), bowiem funkcja skalująca która nie spełnia tego wymogu może istotnie zawężać informacje zawarte w zbiorze wartości zmiennej y, w skrajnym przypadku możemy przybliżać jakąś stałą $g(y) = const.$ W niektórych punktach pracy, np. na str. 9, jako funkcje skalujące wymienia się wielomiany stopnia pierwszego, wydaje się jednak, że takie zawężenie możliwości wyboru funkcji skalujących g jest zbyt ograniczające, gdyż dla (różnowartościowej) funkcji liniowej można łatwo znaleźć funkcję odwrotną g^{-1}, a następnie rozważać problem regresji w klasie funkcji złożonych $g^{-1} \circ f$, $f \in \mathcal{F}$.

Natomiast metoda badań problemu regresji naszkicowana powyżej w p. 3^0 ma dalsze istotne konsekwencje. Mianowicie, zamiast poszukiwania minimum funkcjonału wyrażającego błąd przybliżenia w obrębie podzbiorów przestrzeni Hilberta badania przedstawione w pracy są prowadzone w przestrzeniach pseudo-hilbertowskich, czyli przestrzeniach pseudo-metrycznych zupełnych z pseudo-metryką generowaną przez iloczyn skalarny, w kórych istnieją różne elementy (funkcje różniące się od siebie), ale ich odległości są równe zero. Z tego względu rozważania zaprezentowane w pracy są raczej nietypowe i w konsekwencji trudniejsze od badań w ramach dobrze znanych przestrzeni Hilberta, ponieważ niektóre powszechnie znane twierdzenia dla przestrzeni Hilberta nie są słuszne w przestrzeniach pseudo-hilbertowskich. Ponadto w przypadku, gdy błąd oceny przybliżenia jest wyrażony w metryce średniokwadratowej, pomijając sytuację, gdy mamy do czynienia z dyskretną miarą probabilistyczną o dodatnich masach przyporządkowanych wszystkim punktom przeliczalnej przestrzeni Ω, właściwie nie można znaleźć dokładnych funkcji minimalizujących błąd średniokwadratowy, lecz tylko klasy równoważności takich funkcji (ewentualnie podklasy tych klas po odpowiednim zawężeniu teoretycznego modelu funkcyjnego $\mathcal{F}$), co dodatkowo komplikuje rozważania. Z drugiej strony opisane podejście do problemu regresji może być zaletą, jeśli poszukujemy regresji w klasie zwykłych funkcji, a nie w jakiejś przestrzeni składającej się z klas równoważności funkcji.

Po tych ogólnych uwagach wstępnych możemy przejść do dokładniejszego opisu wyników zawartych w pracy.

W rozdziale I omówione zostały ogólne problemy dotyczące zagadnienia regresji, m.in. przedstawiony został zarys historycznego rozwoju badań naukowych związanych z problematyką regresji. Ponadto wprowadzone zostały podstawowe pojęcia dotyczące zagadnienia regresji, w szczególności sformułowane zostały definicje asynchronicznej struktury regresji, oraz synchronicznej struktury regresji, będącej przedmiotem badań opisanych w pracy.

Rozdział II zawiera pomocnicze rezultaty z zakresu teorii miary i całki wykorzystywane w dalszych częściach pracy. Jednym z głównych wyników podanych w tym rozdziale jest Lemat 2.4, który razem z bezpośrednim dowodem zajmuje prawie 4 strony pracy. Po dokładniejszej analizie okazuje się jednak, że Lemat 2.4 jest właściwie specjalnym przypadkiem Lematu 2.3, zawierającego wzór na całkowanie przez podstawienie (dobrze znany w przypadku całki funkcji rzeczywistych) rozszerzony w rozprawie na całki funkcji zespolonych względem miary, przy czym dowód Lematu 2.3 jest znacznie krótszy i prostszy od dowodu Lematu 2.4. Fakt, iż Lemat 2.4 jest specjalnym przypadkiem Lematu 2.3 można uzasadnić w następujący sposób (przy oznaczeniach analogicznych jak w pracy).

Niech $(\Omega, \mathcal{B})$, $(\Omega', \mathcal{B}')$ i $(\Omega'', \mathcal{B}'')$ będą przestrzeniami mierzalnymi i niech $\Omega_0 = \Omega' \times \Omega''$ będzie przestrzenią z σ-ciałem produktowym $\mathcal{B}' \otimes \mathcal{B}'' = \sigma\left(A' \times A'' : A' \in \mathcal{B}', A'' \in \mathcal{B}''\right)$, tzn. najmniejszym σ-ciałem zawierającym wszystkie iloczyny kartezjańskie zbiorów mierzalnych $A' \in \mathcal{B}'$, $A'' \in \mathcal{B}''$. Załóżmy, że $x : \Omega \to \Omega'$ jest funkcją $(\mathcal{B}, \mathcal{B}')$-mierzalną i $y : \Omega \to \Omega''$ jest funkcją $(\mathcal{B}, \mathcal{B}'')$-mierzalną. Wtedy odwzorowanie $(x, y) : \Omega \to \Omega_0$ określone wzorem $(x, y)(\omega) = (x(\omega), y(\omega))$, $\omega \in \Omega$ jest funkcją $(\mathcal{B}, \mathcal{B}' \otimes \mathcal{B}'')$-mierzalną. W konsekwencji, dla dowolnej miary $\mu : \mathcal{B} \to \overline{R}$, funkcja zbioru $\pi_{x,y} : \mathcal{B}' \otimes \mathcal{B}'' \to \overline{R}$ dana wzorem $\pi_{x,y}(C) = \mu\left((x, y)^{-1}(C)\right)$, $C \in \mathcal{B}' \otimes \mathcal{B}''$, jest miarą na σ-ciele $\mathcal{B}' \otimes \mathcal{B}''$. Niech teraz $G : \Omega_0 \to \mathbb{C}$ będzie dowolną funkcją $(\mathcal{B}' \otimes \mathcal{B}'', \mathcal{B}(\mathbb{C}))$-mierzalną, gdzie $\mathcal{B}(\mathbb{C})$ jest σ-ciałem zbiorów borelowskich w przestrzeni $\mathbb{C}$ z metryką $|z_1 - z_2|$, $z_1, z_2 \in \mathbb{C}$. Wówczas zgodnie z Lematem 2.3 mamy

$$G \in L^1\left(\Omega_0, \mathcal{B}' \otimes \mathcal{B}'', \pi_{x,y}\right) \quad \Leftrightarrow \quad G \circ (x, y) \in L^1(\Omega, \mathcal{B}, \mu),$$

oraz

$$\bigwedge_{G \in L^1(\Omega_0, \mathcal{B}' \otimes \mathcal{B}'', \pi_{x,y})} \quad \int_{\Omega} G \circ (x, y)\, d\mu = \int_{\Omega_0} G d\pi_{x,y}.$$

Załóżmy następnie, że $f : \Omega' \to \mathbb{C}$ jest funkcją $(\mathcal{B}', \mathcal{B}(\mathbb{C}))$-mierzalną, oraz $g : \Omega'' \to \mathbb{C}$ funkcją $(\mathcal{B}'', \mathcal{B}(\mathbb{C}))$-mierzalną. Zauważmy, że przekształcenie $G = f \cdot g : \Omega_0 = \Omega' \times \Omega'' \to \mathbb{C}$ określone wzorem $G(t_1, t_2) = f(t_1) \cdot g(t_2)$, $t_1 \in \Omega'$, $t_2 \in \Omega''$, jest funkcją $(\mathcal{B}' \otimes \mathcal{B}'', \mathcal{B}(\mathbb{C}))$-mierzalną, gdyż jest złożeniem $(\mathcal{B}' \otimes \mathcal{B}'', \mathcal{B}(\mathbb{C}) \otimes \mathcal{B}(\mathbb{C}))$-mierzalnej funkcji (f, g),

$$\Omega_0 = \Omega' \times \Omega'' \ni (t_1, t_2) \mapsto (f(t_1), g(t_2)) \in \mathbb{C} \times \mathbb{C}$$

oraz funkcji ciągłej $\mathbb{C} \times \mathbb{C} \ni (z_1, z_2) \mapsto z_1 \cdot z_2 \in \mathbb{C}$. Przyjmując $G = f \cdot g$ otrzymujemy tezę Lematu 2.4, tj.

$$G = f \cdot g \in L^1\left(\Omega_0, \mathcal{B}' \otimes \mathcal{B}'', \pi_{x,y}\right) \Leftrightarrow G \circ (x, y) = (f \circ x) \cdot (g \circ y) \in L^1(\Omega, \mathcal{B}, \mu),$$

oraz

$$\bigwedge_{f \cdot g \in L^1(\Omega_0, \mathcal{B}' \otimes \mathcal{B}'', \pi_{x,y})} \int_{\Omega} (f \circ x) \cdot (g \circ y)\, d\mu = \int_{\Omega_0} f \cdot g d\pi_{x,y} = \int_{\Omega' \times \Omega''} f(t_1) \cdot g(t_2)\, d\pi_{x,y}(t_1, t_2).$$

Ponadto w rozdziale II na str. 24 w 12 linii od góry pod znakiem ostatniej całki powinna być funkcja $|f|^p$ zamiast $|f|$, można to jednak uznać za błąd typograficzny, który nie utrudnia zrozumienia sensu tego wzoru.

Rozdział III opisuje najważniejsze własności rodzin funkcji regresji dla teoretycznych modeli $\mathcal{F}$ będących zbiorami zamkniętymi ze względu na skończone kombinacje liniowe elementów. W trakcie dokładniejszej analizy tej części pracy stwierdzamy, że formułując Lemat 3.14 oraz Twierdzenie 3.15 warto dodać założenie $\mathrm{Reg}\,(\mathcal{F}, \mathfrak{P}_g) \neq \emptyset$, bowiem dowodzenie własności elementów zbioru pustego, które faktycznie nie istnieją, raczej nie ma sensu. Wypowiedź Wniosku 3.16, który wynika z Lematu 3.14 i Twierdzenia 3.15, zawiera już założenie $\mathrm{Reg}\,(\mathcal{F}, \mathfrak{P}_g) \neq \emptyset$, zatem w pierwszym zdaniu dowodu tego wniosku nie ma potrzeby pisać "założmy, że istnieje $f_0 \in \mathrm{Reg}\,(\mathcal{F}, \mathfrak{P}_g)$", bowiem przy założeniu $\mathrm{Reg}\,(\mathcal{F}, \mathfrak{P}_g) \neq \emptyset$ faktycznie istnieje jakiś element tego zbioru, a więc wystarczy napisać "załóżmy, że $f_0 \in \mathrm{Reg}\,(\mathcal{F}, \mathfrak{P}_g)$". Ponadto dwa ostatnie zdania Uwagi 3.12 w rozdziale III na str. 40 nie są słuszne w całej rozciągłości, ponieważ warunkowe wartości oczekiwane względem σ-ciał określa się dla zmiennych losowych całkowalnych (mających skończony pierwszy moment), natomiast minimalizując błąd średniokwadratowy w klasie wszystkich funkcji mierzalnych względem ustalonego σ-ciała, żeby osiągnąć jakiś konstruktywny wynik należy założyć, że rozważane funkcje są całkowalne z kwadratem (a takie założenie dla zmiennych losowych jest mocniejsze od istnienia momentów rzędu pierwszego).

W rozdziale IV podane zostały dokładne wzory precyzujące postać funkcji regresji w przypadku, gdy klasa $\mathcal{F}$ jest zbiorem rozpiętym na skończenie wielu elementach bazowych. W tym rozdziale warto byłoby wprowadzić drobne poprawki w dowodzie Lematu 4.6. Zamiast drugiego zdania w pierwszej, drugiej i trzeciej linii od góry na str. 55, które brzmi "Ustalmy dowolnie $f \in L^2(A, \mathcal{A}_x, P_x)$ oraz ciąg $\mathbb{N} \ni n \mapsto f_n \in \Theta + \mathcal{F}$ taki, że $\|f_n - f\| \to 0$, gdy $n \to +\infty$" należy raczej napisać "Ustalmy dowolnie ciąg $\mathbb{N} \ni n \mapsto f_n \in \Theta + \mathcal{F}$ zbieżny do pewnej granicy $f \in L^2(A, \mathcal{A}_x, P_x)$, tzn. $\|f_n - f\| \to 0$, gdy $n \to +\infty$. Musimy wykazać, że $f \in \Theta + \mathcal{F}$." Wynika to z faktu, że dla dowolnie ustalonego elementu $f \in L^2(A, \mathcal{A}_x, P_x)$, ciąg $\mathbb{N} \ni n \mapsto f_n \in \Theta + \mathcal{F}$ spełniający warunek $\|f_n - f\| \to 0$, gdy $n \to +\infty$, może nie istnieć. Następnie w liniach 11-13 od góry na str. 56 po stwierdzeniu, że zbiór $\Theta + \mathcal{F}$ jest domknięty w przestrzeni $H(\mathfrak{P})$, pojawia się konkluzja "na mocy Twierdzenia 3.10 istnieje $f \in \mathrm{Reg}\,(\Theta + \mathcal{F}, \mathfrak{P}_g)$." Otóż z Twierdzenia 3.10 rzeczywiście wynika, że $\mathrm{Reg}\,(\Theta + \mathcal{F}, \mathfrak{P}_g) \neq \emptyset$, a więc istnieje jakiś element $f \in \mathrm{Reg}\,(\Theta + \mathcal{F}, \mathfrak{P}_g)$.

Jednakże element $f \in \operatorname{Reg}(\Theta + \mathcal{F}, \mathfrak{P}_g)$ nie musi być tym samym elementem $f \in \Theta+\mathcal{F}$, który jest granicą ciągu $\mathbb{N} \ni n \mapsto f_n \in \Theta+\mathcal{F}$ rozważanym w pierwszej części dowodu, zatem w tym miejscu należy dopisać zdanie "Wtedy istnieje ciąg $\mathbb{N} \ni n \mapsto f_n \in \Theta+\mathcal{F}$ zbieżny do elementu $f \in \operatorname{Reg}(\Theta + \mathcal{F}, \mathfrak{P}_g)$", w przeciwnym razie nie widać związku między elementem $f \in \operatorname{Reg}(\Theta + \mathcal{F}, \mathfrak{P}_g)$ i pierwszą częścią dowodu. Na str. 65 w pierwszej linii wzoru (4.48) występuje zmienna t_1, której nie ma w drugiej linii tego wzoru, natomiast w drugiej linii pojawia się zmienna y_k, która nie ma związku ze zmienną t_1 (w pierwszej linii tego wzoru zamiast dwukrotnie występującej zmiennej $t_1 \in \mathbb{Z}_{1,n}$ powinna być wartość $k \in \mathbb{Z}_{1,n}$). Ponadto w ostatnim wzorze na str. 65 oraz wzorach w drugiej i czwartej linii na str. 66, a także w drugiej linii na str. 68, po zastosowaniu ogólnej równości (4.33) pojawiają się całki po zbiorach $A \times \mathbb{C}$ i $\mathbb{C}$ względem miar $P_{x,y}$ i $P_{x,y}(x_k, \cdot)$, podczas gdy formalnie biorąc miary $P_{x,y}$ i $P_{x,y}(x_k, \cdot)$ nie są określone w takich przestrzeniach. Mianowicie, miara $P_{x,y}$ została określona wzorem (3.6) na str. 29 w przestrzeni $\mathbb{Z}_{1,n} \times \mathbb{R}$, a nie $\mathbb{Z}_{1,n} \times \mathbb{C}$, zatem $P_{x,y}(x_k, \cdot)$ jest miarą określoną na przestrzeni $\mathbb{R}$, a nie na $\mathbb{C}$. Taką nieścisłość można jednak łatwo usunąć, jeśli rozszerzymy $P_{x,y}$ do miary $\overline{P}_{x,y}$ na $\mathbb{Z}_{1,n} \times \mathbb{C}$ przyjmując $\overline{P}_{x,y}(\{k\} \times B) = P_{x,y}(\{k\} \times (B \cap \mathbb{R}))$ dla zbiorów borelowskich B w przestrzeni $\mathbb{C}$, gdyż wówczas $\overline{P}_{x,y}(\{k\} \times (\mathbb{C} \setminus \mathbb{R})) = 0$. Warto również wspomnieć, że w czwartej linii od dołu na str. 51 i podobnie na str. 58 występuje równość $\operatorname{lin}(\emptyset) = \{\theta\}$, gdzie θ jest elementem zerowym przestrzeni $H(\mathfrak{P})$, tymczasem zgodnie z określeniem podanym w 14 linii od góry na str. 28 symbol $\operatorname{lin}(S)$ oznacza otoczkę liniową niepustego zbioru $S \subset (A \to B)$ (zbiór pusty jest podzbiorem każdej przestrzeni, a więc bez dodatkowego objaśnienia można przyjąć, że $\operatorname{lin}(\emptyset)$ jest zbiorem zawierającym element zerowy dowolnej przestrzeni; czytelnik tylko na podstawie kontekstu może domyślać się o jaki element zerowy tu chodzi). Warto również wspomnieć, że definicja (4.30) wielkości $E(u|v)$ w 4 linii na str. 60 może być nieco myląca, ponieważ w teorii prawdopodobieństwa symbolem $E(u|v)$ oznacza się warunkową wartość oczekiwaną zmiennej losowej u względem zmiennej v, natomiast wzór (4.30) definiuje pewien operator (można tu byłoby zastosować np. oznaczenie $E(u/v)$). Wreszcie w zdaniu "Ustalmy dowolnie $n \in \mathbb{Z}_{1,p-1}$ ..." w drugiej linii od dołu na str. 74 indeks n powinien być zastąpiony przez l.

Nie mam żadnych uwag ani zastrzeżeń dotyczących rozdziału piątego oraz załącznika na końcu rozprawy. W tej części pracy znajdują się interesujące przykłady zastosowań teorii rozwiniętej w poprzednich rozdziałach.

Warto natomiast wspomnieć o kilku drobnych błędach literowych jakie wkradły się do pracy. Np. kilka razy błędnie zostało napisane nazwisko Lebesgue, m.in. na str. 17 w drugiej linii tekstu od góry i na str. 24 w drugiej linii od góry mamy słowo Lebesque (napisane przez "q"), na str. 16 w trzynastej linii od dołu powinno być "wyznaczeniu wszystkich funkcji $f_0 \in \mathcal{F}$ minimalizujących funkcjonał F", a nie "minimalizujący funkcjonał F", na str. 42 w jedenastej linii od góry pojawia się dwukrotnie to samo słowo "jest jest", itp. Kilka błędów można też dostrzec w Bibliografii, np. poprawny tytuł książki M. Fisza brzmi *Probability Theory and Mathematical Statistics,* a nie *Statistic* (poz. [11]), prawidłowy tytuł książki podanej jako poz. [28] Bibliografii powinien brzmieć *Regressions: Why Are Economists Obsessed with Them?* a nie *Obessessed,* w poz. [30] zamiast *Real and complex analysis,* McGRAF-HILL, tytuł książki należy napisać dużymi literami (konsekwentnie jak w innych pozycjach Bibliografii), a nazwę wydawnictwa też nieco inaczej, tzn. *Real and Complex Analysis,* McGraw-Hill.

Kilka uwag należy także poświęcić terminologii stosowanej w rozprawie. Np. na str. 9 jest mowa o funkcji próbkującej y, co sugeruje, że wartości y są znane, a więc powstaje pytanie: dlaczego mamy oceniać wartości y przy pomocy funkcji (regresji) zmiennej x? Również nazwa rozmaitość liniowa użyta w dziesiątej linii na str. 28 i później stosowana w dalszym ciągu pracy na oznaczenie zbioru elementów $\mathcal{F}$ zamkniętego ze względu na skończone kombinacje liniowe wydaje się nieodpowiednia. Ponadto nazwa przestrzeń pseudo-Hilberta stosowana w całej pracy jest raczej niefortunnie wybrana, gdyż może nasuwać przypuszczenie, że chodzi o przestrzeń skonstruowaną przez kogoś kto podszywa się pod Hilberta (nieco lepsza wydaje się nazwa przestrzeń pseudo-hilbertowska). Warto także wspomnieć o braku konsekwencji dotyczącej oznaczeń, np. w rozdziale II σ-ciała podzbiorów przestrzeni Ω (z dodatkowymi wskaźnikami) są oznaczane literą $\mathcal{B}$, a w innych rozdziałach pracy literą $\mathcal{A}$, miara $\pi_{x,y}$ powstaje z miary μ, a więc powinna być oznaczona symbolem $\mu_{x,y}$, podobnie jak miara $P_{x,y}$ powstała z miary P, itd.

Przechodząc do merytorycznej oceny pracy, w mojej opinii najciekawszymi rezultatami otrzymanymi w wyniku badań probabilistycznych struktur regresji są Twierdzenia 3.8, 3.10 i 3.11, oraz Twierdzenia 4.10, 4.11 i 4.12. Wspomniane rezultaty, jak również kilka lematów pomocniczych, są wynikami które mogą odegrać ważną rolę w teorii regresji, przy czym wydaje się, że większość z nich można uogólnić (np. na kombinacje wypukłe, czy funkcje o wartościach w przestrzeniach nieskończenie wymiarowych).

Pomimo sporej liczby wymienionych powyżej drobnych usterek należy podkreślić, że wszystkie rozważania zawarte w omawianej pracy doktorskiej zostały przedstawione w sposób jasny i przejrzysty, przytoczone dowody większości rezultatów są raczej krótkie i zwięzłe, a sposoby argumentacji i metody rozumowania – precyzyjne i przekonujące. Jak zwykle w tego rodzaju obszernych publikacjach Autorka nie ustrzegła się niestety drobnych błędów i niedociągnięć, o których wspomniałem powyżej. Większość tych usterek można jednak uznać za błędy typograficzne, a wszystkie wspomniane wcześniej usterki można łatwo usunąć (przygotowując np. wyniki pracy do publikacji w formie artykułów). Ponadto dostrzeżone niedociągnięcia nie utrudniają zbytnio zrozumienia rezultatów przedstawionych w pracy.

Podsumowując powyższe uwagi można stwierdzić, że opisywana rozprawa doktorska stanowi niewątpliwie oryginalne rozwiązanie niektórych problemów teorii regresji, a w szczególności charakteryzacji rodzin funkcji minimalizujących średniokwadratowy błąd przybliżenia nieznanych wielkości przy pomocy wielkości obserwowanych. Warto również wspomnieć, że w trakcie prowadzonych w rozprawie rozważań Autorka wykazała się dużą wiedzą teoretyczną, a także umiejętnością samodzielnego myślenia oraz rozwiązywania problemów matematycznych. Biorąc pod uwagę powyższe spostrzeżenia ogólna ocena rozprawy doktorskiej mgr Agnieszki Szumery powinna być pozytywna.

Uważam, że rozprawa doktorska "Probabilistyczne struktury regresji" mgr Agnieszki Szumery spełnia wszelkie wymagania ustawy z dn. 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki wraz z późniejszymi zmianami (załącznik do obwieszczenia Marszałka Sejmu Rzeczypospolitej Polskiej z dn. 15 września 2017 r. Dz. U. poz. 1789). W związku z tym składam wniosek o dopuszczenie jej Autorki, mgr Agnieszki Szumery, do dalszych etapów przewodu doktorskiego.

August Zapała

dr hab. August Zapała