

Prof. UW dr hab. Anna Wójtowicz
Zakład Logiki
Wydział Filozofii i Socjologii UW

RECENZJA ROZPRAWY DOKTORSKIEJ

mgra Adama Pawła Kubiaka

Jerzy Neyman's Conception of Scientific Inference and Its Relation to Bayesianism

(Koncepcja wnioskowań naukowych Jerzego Neymana i jej związku z bayesianizmem)

Przygotowanej pod kierunkiem Prof. KUL dr hab. Pawła Kawalca (promotor główny) i dr Adama Kiersztyna (promotor pomocniczy)

Rozprawa mgr Adama Kubiaka poświęcona jest przedstawieniu koncepcji wnioskowań naukowych Jerzego Neymana, a także obronie tej koncepcji przed zarzutami, formułowanymi z pozycji bayesianizmu. Niezależnie od oceny merytorycznej tych wysiłków chcę podkreślić, że w świetle powszechnej krytyki statystyki częstościowej (wyrażającej się w szczególności deprecjonowaniem osiąganego jej metodami wyników) taką postawę uważam za odważną i godną uznania. Niewątpliwie bowiem łatwiej i bezpieczniej byłoby Neymana krytykować niż go bronić.

Ze względu na stopień oryginalności formułowanych tez rozprawę można podzielić na dwie zasadnicze części:

W części pierwszej, obejmującej wstęp i rozdziały 1-5, wprowadzona jest odpowiednia aparatura pojęciowa, a następnie – za jej pomocą – przedstawiona zostaje koncepcja Neymana. Autor referuje tu głównie poglądy obecne w bogatej literaturze przedmiotu. Ocenie podlega więc przede wszystkim to, jak wyczerpujące i rzetelne jest to przedstawienie.

I tak **w rozdziale 1** podana jest aksjomatyczna *definicja* pojęcia prawdopodobieństwa (1.1), a następnie przedstawiona jest klasyczna *definicja* prawdopodobieństwa (1.2), geometryczna *definicja* prawdopodobieństwa (1.3), logiczna *interpretacja* prawdopodobieństwa (1.4), subiektywistyczna *interpretacja* prawdopodobieństwa (1.5), częstościowa *interpretacja* prawdopodobieństwa (1.6) i częstościowa *interpretacja* prawdopodobieństwa (1.7). O ile omówienie poszczególnych punktów jest dobre i ciekawie przedstawione, to nieco wątpliwości budzi rozróżnienie *definicji* prawdopodobieństwa i *interpretacji* prawdopodobieństwa (nie do końca zresztą konsekwentne, bo w podsumowaniu (1.8) Autor mówiąc o „klasycznym prawdopodobieństwie” pisze, że „ta interpretacja wydaje się być...” (s. 30, w. 13 od dołu)). Do pewnego stopnia rozumiem stojące za tym intencje – w przypadku (1.2) i (1.3) mamy zadane pewne dodatkowe warunki na funkcję P, które nie występują w definicji aksjomatycznej. Można jednak twierdzić, że wynika to raczej z tego,

jaką przyjmuje się w nich *interpretację* Ω , tzn. co zakłada się o elementach tego zbioru. W moim przekonaniu lepiej i spójniej trzymać się tu ugruntowanej (np. przez Hájek, A. "Interpretations of Probability", *The Stanford Encyclopedia of Philosophy*) terminologii – i nazywać wszystko od (1.2) do (1.7) interpretacjami.¹ Jeśli się tego nie robi – trzeba taką decyzję uzasadnić.

W **rozdziale 2** omówione są różne metody wnioskowania, które angażują pojęcie prawdopodobieństwa: paradygmat bayesowski, paradygmat częstościowy i paradygmat wiarygodnościowy. Autor słusznie zauważa, że wnioskowania prowadzone w ramach tych paradygmatów stawiają sobie z gruntu różne cele.

Wnioskowania oparte na twierdzeniu Bayesa służą do ustalania jak silne powinniśmy mieć przekonanie co do prawdziwości danej tezy w oparciu uzyskane informacje, która z dwóch konkurencyjnych hipotez jest lepiej uzasadniona przez uzyskaną informację, jak należy modyfikować zbiór naszych przekonań po uzyskaniu nowych informacji.

Trochę zabrakło mi w tym miejscu zwrócenia uwagi tzw. zasadę regularności przyjmowaną w tym paradygmacie. Zgodnie z nią, jeśli dane zdarzenie A jest możliwe (logicznie), to $P(A) \neq 0$. Przyjęcie tej zasady stanowi warunek konieczny twierdzenia o konwergencji (o którym Autor wspomina pośrednio w przypisie 107) – por. np. [Weisberg 2011], cytowany w rozprawie. Twierdzenie to osłabia subiektywistyczny charakter wnioskowań opartych na tw. Bayesa.

Wnioskowania w paradygmacie częstościowym to przede wszystkim pewnego rodzaju metody testowania hipotez (omówione na klasycznym przykładzie damy rozpoznającej, co najpierw wlano do filiżanki i ustalaniu rzetelności monety w oparciu o obserwowane wyniki rzutów), jak również rozszerzone o pojęcia probabilistyczne schematy modus tollens. Schematy te stanowią podstawę rozumowania określanego skrótem NHST, w którym główna przesłanka opiera się na stwierdzeniu, jak niskie jest prawdopodobieństwo uzyskania danego wyniku przy założeniu prawdziwości testowanej hipotezy. Autor wskazuje na kilka istotnych własności tego paradygmatu:

Wszelkie stwierdzenia dotyczące ewidencji i hipotez są stwierdzeniami częstościowymi; Centralnym pojęciem jest pojęcie błędu będącego wynikiem wnioskowania i ograniczenia częstości jego występowania (a stąd kluczowe znaczenie pojęcia poziomu istotności); Wnioskowania takie nie są zgodne z zasadą wiarygodności (*likelihood principle*).

Moim zdaniem rozdział ten jest dobrym wprowadzeniem, pokazującym na czym polegają zasadnicze różnice między podejściem statystycznym i podejściem bayesowskim.

W **rozdziale 3** zaprezentowane są główne osiągnięcia Neymana w dziedzinie metodologii prowadzenia badań – konstruowania eksperymentów, strategii doboru reprezentatywnych próbek, interpretacji uzyskanych wyników. Są one bardzo istotne z punktu widzenia dalszych

¹ Na stronie 29 (3 w. od dołu i przypis 24) mamy też literówkę we wzorze: zamiast $P(B|A) > P(A)$ powinno być $P(B|A) > P(B)$.

rozważań, bo pokazują, jakie znaczenie przy konstruowaniu eksperymentów i doborze próbek mają aprioryczne informacje (w szczególności dotyczące aspektów ekonomicznych, społecznych czy hierarchii przyjmowanych przez naukowców wartości).

W **rozdziale czwartym** omówione są własności wnioskowań statystycznych, czyli wnioskowań, które mają ustalać (przybliżyć) pewne interesujące nas cechy danej populacji. Przedstawione są dwa główne typy takich wnioskowań: estymacja parametryczna i metoda testowania hipotez (Neymana-Pearsona). Są to rozważania kluczowe z punktu widzenia jednej z tez rozprawy – ponieważ pojawiające się tu pojęcia posłużą do oceny jakości procedur stosowanych na gruncie koncepcji Neymana. Ważne miejsce zajmuje również paragraf 4.2.3. w którym analizowane są powody, dla których dwa typy błędów (błąd typu I – odrzucenie hipotezy prawdziwej i błąd typu II – przyjęcie hipotezy fałszywej) mogą mieć przypisane inne wagi. Jest to kolejny moment, w którym przekonania badacza wykonującego test mogą mieć istotny wpływ na jego kształt i interpretacje jego wyników.

Rozdział piąty jest poświęcony przedstawieniu filozoficznych poglądów Neymana na temat prawdopodobieństwa, statystyki, i ich zastosowań przy wnioskowaniach i opisie danych.² Wykorzystane są założenia, które jawnie zostały sformułowane w rozdziale 2. Wynika z nich w szczególności, że rzetelność testu powinna być mierzona częstością błędów, które generuje oparte na jego wynikach, wielokrotnie powtarzane rozumowanie. Powtarzany jest jednak nie jeden i ten sam test, zastosowany do jednej hipotezy, ale ten sam rodzaj testu do różnych hipotez. Tym co pozwala zachować spójność całej koncepcji jest odwołanie się do Centralnego Twierdzenia Granicznego, zgodnie z którym średni błąd z testowania wielu hipotez daną metodą będzie dążył do średniej z błędów z testowania poszczególnych hipotez.

Zaprezentowane są tu również poglądy Neymana na temat paradygmatu opartego na twierdzeniu Bayesa. Według Neymana rozumowanie bayesowskie – chociaż oczywiście poprawne matematycznie – ma zastosowanie tylko do bardzo szczególnych i rzadkich sytuacji, w których znane są prawdopodobieństwa a priori testowanych hipotez. Co więcej, Neyman twierdzi, że hipotezy są (najczęściej) stałymi (parametrami pewnego modelu), nie mają statusu zmiennych losowych i nie można im w związku z tym przypisywać prawdopodobieństwa a priori. Trzeba jednak zwrócić uwagę, że Neyman ma tu na myśli prawdopodobieństwo w sensie częstościowym a nie subiektywnym, czyli takim, jaki przyjmuje się w bayesianizmie. Argument ten nie jest więc przekonujący. Równie dobrze Neyman mógłby krytykować zwolenników Bayesa mówiąc, że niesłusznie przypisują oni prawdopodobieństwo zdarzeniom jednostkowym. Podobnie nie jest dla mnie jasne, w jakim sensie zastosowanie twierdzenia Bayesa jest sprzeczne z prawem wielkich liczb. Przypis 107 tego nie wyjaśnia – bo nie bardzo wiadomo, co to znaczy, że ten sam eksperyment został wykonany jednocześnie w wielu miejscach, dając takie same wyniki E.

Zasadnicze znaczenie dla dalszych rozważań ma teza Neymana dotycząca wyników wnioskowań statystycznych. Nie sprowadzają się one do uznania, że pewna hipoteza jest prawdopodobna lub prawdziwa, czy do nabycia określonych przekonań. Akceptacja hipotezy

² W przykładzie podanym na początku paragrafu 5.2 jest błąd – w pierwszej urnie powinna być jedna kula biała i jedna kula czarna – a nie, jak pisze autor – tylko jedna kula czarna.

skutkuje podjęciem pewnej decyzji (pewnego działania). Dzięki temu Neyman mógł twierdzić, że stosowane przez niego procedury statystyczne nie są wnioskowaniami indukcyjnymi i nie podlegają w związku tym wyłącznie epistemicznej ocenie.

Mam świadomość, że spór między Neymanem a Fisherem, dotyczącym tej kwestii, ma długą tradycję i bogatą literaturę, muszę jednak przyznać, że nie zgadzam się z powyższą tezą Neymana i przekonuje mnie prosty argument zawarty np. w pracy Szaniawskiego („Wnioskowanie czy *behaviour*?”³). Jeśli podejmujemy jakąś decyzję w oparciu o wyniki testu dotyczącego hipotezy H i działamy racjonalnie, to wybieramy decyzję stosując kryterium maksymalizowania wartości oczekiwanej (lub minimalizowania oczekiwanej straty). To z kolei opiera się (pośrednio lub bezpośrednio) na założeniu, że H jest prawdziwa z pewnym prawdopodobieństwem – bo tylko w taki sposób wartość oczekiwaną naszej decyzji możemy obliczyć. Jeśli więc nawet uznamy, że w efekcie wykonania procedury statystycznej podejmiemy pewną decyzję, to u jej podłoża tkwi przekonanie na temat hipotezy H. Jest to niezgodne z tym, co twierdzi Neyman.

Oczekiwałam, że Autor rozprawy spróbuje Neymana przed Szaniawskim obronić.

Podsumowując – w moim przekonaniu rozdziały 1-5 rozprawy bardzo dobrze realizują postawione przed nimi cele. Pozwalają w dostatecznie szerokim kontekście przedstawić stanowisko metodologiczno-filozoficzne Neymana, co umożliwi – w rozdziale następnym – dokonanie jego oceny. Dodatkowo należy podkreślić, że otrzymujemy tu bardzo pełną i nie spotykaną dotąd w literaturze rekonstrukcję całej koncepcji wnioskowań naukowych Neymana. Już samo to zasługuje na duże uznanie.

Rozdział 6 zawiera szereg oryginalnych tez. Do najważniejszych z nich moim zdaniem należą następujące:

- 1) Stosowane przez Neymana procedury statystyczne są w dostatecznym stopniu rzetelne.
- 2) Przy stosowaniu procedur statystycznych, w koncepcji Neymana, wykorzystywane są różne rodzaje wiedzy apriorycznej (teoretycznej, empirycznej, praktycznej, dotyczącej hierarchii wartości).
- 3) To, że na szczegółowe charakterystyki procedur statystycznych (a więc również na ich wyniki) mają wpływ nie tylko ściśle poznawcze wartości ale również inne (ekonomiczne, społeczne, praktyczne) nie jest ich wadą ale wręcz zaletą. Dzięki temu realizowane są różne pragmatyczne cele, które powinny stać przed badaniami naukowymi.

Tezy te stanowią bezpośrednie odpowiedzi na zarzuty, jakie stawiają tej koncepcji jej przeciwnicy (szczególnie z obozu Bayesa). Wspierające je argumenty uważam za przekonujące w odniesieniu do (2) i (3), i w mniejszym stopniu w odniesieniu do (1).

³ K. Szaniawski *O nauce, rozumowaniu i wartościach*, Warszawa, PWN 1994, 139-147

Niezamierzonym – jak sędzę – efektem zaprezentowanej przez Autora argumentacji jest to, że tak opisane metody Neymana prowadzenia badań naukowych bardzo zbliżają je do **heurystyk**. Od razu chcę podkreślić, że terminu tego nie używam w pejoratywnym sensie.

Heurystyki – w rozumieniu wprowadzonym do współczesnej literatury przez Gigerenzera – to sposoby wnioskowania, które

- Są nastawione na osiągnięcie praktycznych celów, tzn. wyciąganie wniosków dostatecznie bliskich prawdzie, aby zapewnić wnioskującemu osiągnięcie założonego progu poprawności;
- Pozwalają ograniczyć koszty procedury wnioskowania (koszty poszukiwania informacji, czasu jej przetwarzania itp.)
- Są dobierane ze względu na bardzo konkretne, specyficzne własności danego problemu i bazują na często nawet nie uświadamianej wiedzy wnioskującego na temat środowiska, którego wnioskowanie dotyczy.

Stosowanie heurystyk jest uważane za racjonalne korespondencyjnie (tzn. działają one skutecznie) ale nie racjonalne koherencyjnie (tzn. nie działają one w oparciu o normatywne reguły, nakazujące np. w każdych okolicznościach wykorzystywanie wszystkich dostępnych informacji, maksymalizowanie wartości oczekiwanej itp.).

Wydaje się, że wskazane cechy charakterystyczne heurystyk są zbliżone do tych, które w tezach (1)-(3) autor rozprawy przypisuje procedurom statystycznym Neymana. Dlatego sędzę, że warto zastanowić się nad tym, czy również procedur tych nie zakwalifikować raczej jako działań racjonalnych korespondencyjnie a nie koherencyjnie. Zupełnie inaczej można będzie wtedy podchodzić do problemu ich uniwersalnej stosowalności i niezawodności.

Podsumowując uwagi krytyczne i polemiczne, chciałabym wskazać kilka zagadnień, które uważam za wymagające wyjaśnienia:

1. Czym się różnią – co do istoty – definicje (o których mówi się w punktach (1.2) i (1.3)) od interpretacji prawdopodobieństwa?
2. Czy argumentacja Neymana przeciwko koncepcji wnioskowań opartych na tw. Bayesa jest zasadna tylko z punktu widzenia specyficznej, częstościowej interpretacji prawdopodobieństwa? Czy rzeczywiście tw. Bayesa jest sprzeczne z prawem wielkich liczb?
3. Jak można bronić teorio-decyzyjnej interpretacji wyników wnioskowań statystycznych przed prostym argumentem zawartym np. w Szaniawski „Wnioskowanie czy behaviour?”⁴
4. Czy w świetle tez (1) – (3) można powiedzieć, że koncepcja wnioskowań statystycznych jest heurystyką w rozumieniu Gigerenzera (por. np. G. Gigerenzer, T. Sturm “How (far) can rationality be naturalized?”, *Synthese* (2012) 187:243–268)?

Niezależnie od tych drobnych uwag krytycznych rozprawę uważam za bardzo dobrze i ciekawie napisaną. Autor wypełnił istniejącą w historii rozwoju statystyki lukę,

⁴ K. Szaniawski *O nauce, rozumowaniu i wartościach*, Warszawa, PWN 1994, 139-147

przedstawiając w sposób pełny poglądy Jerzego Neymana. Na pochwałę zasługuje również wnikliwa argumentacja przedstawiona w rozdziale 6. W wielu wypadkach poprawia ona i rozszerza to, co można znaleźć w istniejącej w literaturze przedmiotu.

WNIOSEK

Rozprawę doktorską mgra Adama Pawła Kubiaka „Koncepcja wnioskowań naukowych Jerzego Neymana i jej związki z bayesianizmem” oceniam bardzo wysoko i nie mam najmniejszych wątpliwości, że spełnia ona wymogi stawiane rozprawom doktorskim. W związku z tym wnoszę o dopuszczenie mgra Adama Pawła Kubiaka do dalszych etapów przewodu doktorskiego

Warszawa, 6.12.2018

