

Kryterium Akaike: prostota w języku statystyki

Łukasz Kukier*

*Katedra Fizyki Teoretycznej, Katolicki Uniwersytet Lubelski im. Jana Pawła II,
Al. Racławickie 14, 20-950 Lublin*

Marek Szydłowski†

*Katedra Fizyki Teoretycznej, Katolicki Uniwersytet Lubelski im. Jana Pawła II,
Al. Racławickie 14, 20-950 Lublin
i Centrum Układów Złożonych, Uniwersytet Jagielloński,
Reymonta 4, 30-059 Kraków*

Paweł Tambor‡

*Katedra Fizyki Teoretycznej, Katolicki Uniwersytet Lubelski im. Jana Pawła II,
Al. Racławickie 14, 20-950 Lublin*

Innymi słowy, nie szukamy po prostu prawdy,
szukamy bowiem prawdy bardziej interesującej i wyjaśniającej,
szukamy teorii, które przynoszą rozwiązania interesujących problemów.
Szukamy, jeśli to możliwe, teorii głębokich.
K. Popper, Wiedza obiektywna

Abstract

Celem pracy jest pokazanie możliwości sformułowania pojęcia prostoty w języku statystyki klasycznej (niebayesowskiej). Prezentujemy teoretyczne podstawy kryterium *AIC* w ramach teorii informacji oraz wskazujemy na zalety i wady ujmowania pojęcia prostoty modelu w terminach tego kryterium. Wskazujemy na ważność *AIC* w kontekście Forstera-Sobera koncepcji filozofii nauki. Pokazujemy, że intuicje K. R. Poppera, odnośnie pojęcia prostoty odnajdują swoją formalizację w kryterium Akaike. Przedstawiamy przykład, który *explicite* wskazuje, że *AIC* automatycznie nie odrzuca modelu bardziej skomplikowanego na rzecz prostszego.

1 Wstęp

Niewątpliwie nową cechą kosmologii współczesnej jest to, że staje się ona w coraz większym stopniu nauką empiryczną. Dzieje się to wraz z rozwojem coraz bardziej wyrafinowanych technik obserwacyjnych. Mimo kontrowersji niektórych uczonych skupionych wokół rozważań nad możliwością przeprowadzania w kosmologii eksperymentów w sensie ścisłym [1], uzyskiwana baza empiryczna dostarcza coraz lepszych przesłanek do wnioskowań falsyfikujących w sensie Poppera. Na przykładzie oglądu kosmologii współczesnej z dużą wyrazistością możemy także obserwować efekt przesunięcia zainteresowań naukowych od

*e-mail: lkukier@op.pl

†e-mail: uoszydlo@cyf-kr.edu.pl

‡e-mail: xpt76@poczta.fm

teorii naukowych w kierunku modeli teoretycznych¹. Na tym gruncie jest w pewnym sensie realizowana koncepcja skrzynki narzędziowej zaproponowana przez Nancy Cartwright [2], zgodnie z którą modele teoretyczne reprezentują badane układy empiryczne, podczas gdy same teorie naukowe służą raczej do ich konstrukcji jako pewne konceptualne narzędzia. We współczesnej kosmologii obserwacyjnej, co warto podkreślić, modele są w dużej mierze autonomicznymi narzędziami praktyki badawczej².

Zasygnalizowany powyżej optymizm, dotyczący pojawiających się coraz skuteczniejszych narzędzi testowania teorii, zbiega się, między innymi, z próbami wyjaśnienia problemu akceleracji Wszechświata; a w związku z nim kwestii istnienia i natury tzw. ciemnej energii [7, 8, 9]. Generalnie istnieje przynajmniej kilkadziesiąt różnych propozycji ciemnej energii³. Nie ma powodu tutaj omawiać ich bardziej szczegółowo. Wszystkie te mechanizmy są czysto fenomenologicznymi opisami jednymi z wielu, podczas gdy rozwiązanie zdaje się dostarczyć fizyka cząstek elementarnych.

W kontekście analiz, które zamierzamy podjąć w tej pracy, warto już na początku zasygnalizować pewien wyróżniony punkt widzenia w ocenie danego modelu, mianowicie jakość i liczbę parametrów użytkowanych przez model. Istnieje powszechna zgoda, że prostota wyraża się w terminach liczby wolnych (istotnych) parametrów konkurujących hipotez (Forster, Sober). Każda zmienna w modelu wnosi pewną informację, czyli po części odpowiada za zmienność zjawiska. Oczywiście dążymy do tego, żeby opisać w terminach modelu to zjawisko jak najdokładniej. W pewnych sytuacjach modelujących – przykładowo układów złożonych ze swej natury – uzyskanie takiej 100% zgodności z obserwacją jest niemożliwe. Pomijamy tutaj sytuację, w której proste układy opisują złożone zachowania, jak w przypadku chaosu deterministycznego. W takim przypadku argument prostoty mówi nam, że powinniśmy uwzględnić minimalną liczbę czynników, które dadzą nam zakładany poziom dokładności. Innymi słowy, jeśli uporządkować czynniki według ich siły wpływu, to bierzemy wszystkie powyżej kreski. Takie jest klasyczne statystyczne rozumienie, np. w duchu współczynnika determinacji, a także dla *AIC*. Mając współczynnik determinacji *explicite*, widzimy granice dokładności modelu (stopnia wyjaśniania modelu, twierdząc że np. $R^2 = 0.90$). Natomiast *AIC* mówi, nie tyle o granicy dokładności, co o tym, że od pewnego momentu uwzględnienie każdego następnego czynnika daje coraz mniejszy wkład do wyjaśnienia zjawiska. W przypadku *AIC* nie kontrolujemy tej granicy tak, jak dla współczynnika determinacji, który ma sens dla układów liniowych. Te obserwacje stanowią podstawowe intuicje przy konstruowaniu wskaźników prostoty w pojęciach statystycznych. Wyobraźmy sobie następującą sytuację: dysponujemy opisem jakiegoś zjawiska przy pomocy dwóch różnych modeli, w których występują różne pojęcia mające swoje matematyczne formalizacje. Jeżeli teraz te modele w jednakowym stopniu wyjaśniają dane zjawisko (albo oba opisy mieszczą się w akceptowalnej granicy dokładności), należy wybrać model, który jest konceptualnie prostszy (np. z mniejszą liczbą pojęć matematycznych). Gdyby

¹Por. Zeidler P., „Modele teoretyczne jako narzędzia badawcze nauk empirycznych”, wykład na Ogólnopolskim Interdyscyplinarnym Seminarium Naukowym z Filozofii Przyrody i Nauk Przyrodniczych, 15 luty 2007 KUL Lublin.

²Dzień 12 lutego 2003r., kiedy Wilkinson Microwave Anisotropy Probe [WMAP] ogłasza wyniki bazujące na jednorocznych obserwacjach [3, 4, 5], jest uważany za początek *golden age* w kosmologii. Obserwacje te umożliwiły wyznaczenie aż 22 parametrów. Rodzi się wówczas koncepcja minimalnego modelu kosmologicznego opartego na 6 parametrach (τ , Ω_Λ , ω_d , ω_b , A_s , n_s) [6]. Parametry te określają tzw. *vanilla model* Wszechświata. Model ten można zdefiniować przy pomocy kilku parametrów, których pomiar jest możliwy przez WMAPa, jeśli przyjmiemy pewne określone założenia teoretyczne.

³Dynamiczna stała kosmologiczna, samooddziaływające pola skalarnie z potencjałem – *kwintesencje*, zmienne w czasie kosmologicznym równania stanu, zmodyfikowane równania pola w uogólnionym nieliniowym lagrangianem $L(R)$ (R – skalar Ricciego), wszechświat branowy z dodatkowymi wymiarami, ciecz Chaplygina, oddziałująca ciemna materia i ciemna energia i wiele innych.

się odwołać do kosmologii cierpiącej na problem ciemnej energii, to oznaczałoby wybór modelu Λ CDM zamiast innych alternatywnych modeli z ciemną energią. Istotnie jest on prostszy i lepszy w stosunku do konkurencyjnych, podczas gdy wszystkie posiadają, z grubsza biorąc, podobną moc wyjaśniającą. Oczywiście model Λ CDM jest już modelem bardziej złożonym w stosunku do CDM, lecz jego moc wyjaśniająca jest zdecydowanie lepsza i dlatego go preferujemy.

Zaistniałą sytuację nazwijmy problemem istnienia tzw. *multiple explanation*. W kontekście naszej pracy chcemy postawić nie tylko, a może nie tyle, kwestię falsyfikacji danych propozycji teorio-modelowych, ale zaproponować narzędzia oceny wartości modelu przy pomocy kryterium Akaike. Zdając sobie sprawę z wad i ograniczeń związanych z tym podejściem, sformułujemy określone kontekstualne warunki stosowania tego narzędzia i zwrócimy uwagę na jego skuteczność w osiąganiu zamierzonych specyficznych celów w „środowisku” badawczym kosmologii współczesnej.

Aby osiągnąć ten cel, nakreślimy na początku pewną perspektywę dotychczasowych rozważań dotyczących kwestii *prostoty* - czym jest, a czym nie jest w kontekście nauki. Tym samym część pierwsza pracy – ogólna, o charakterze metanaukowym – poprzedzi ściśle zaprezentowanie formalizmu kryterium Akaike (sekcja 2). W części trzeciej pracy powrócimy do rozważań o charakterze filozoficznym, ale już będących oceną prezentowanego i badanego kryterium. Wreszcie zaprezentujemy konkretne *studium przypadku* stosowania AIC w odniesieniu do orbit keplerowskich w układzie planetarnym.

2 Prostota uwikłana w statystykę

Najbardziej podstawowym kontekstem metanaukowych rozważań dotyczących kryterium prostoty jest kwestia oceny wyjaśniania jako takiego. Celem naszej pracy jest nie tylko wyodrębnienie tych kryteriów, które oznacza i reprezentuje sobą, w odniesieniu do nauki, prostota, ale poszukanie kryteriów dla niej samej, czyli warunków jej *detekcji* jako takiej. Jest to tym bardziej konieczne, że historia nauki pokazuje, jak różnie nie tylko definiowano (jeśli definiowano), ale i rozumiano zakres nazwy „prostota”. Wystarczy porównać ze sobą różne sformułowania tzw. *brzytwy Ockhama*, która zwykle towarzyszy lub zastępuje takie określenia stosowane w odniesieniu do hipotez, jak ekonomiczność, oszczędność czy prostota właśnie. Wymieńmy na początek kilka klasycznych sformułowań:

- Kiedy teoria nie może być potwierdzona przez dowód, z dwóch przeciwstawnych sobie poglądów akceptować należy ten, co do którego mamy najmniej wątpliwości⁴
- Prostszy jest coś, co może być równie dobrze wyjaśnione przy pomocy mniejszej liczby przesłanek.
- J. L. Rodríguez-Fernández [10, s. 121] cytuje dwa ciekawe sformułowania zasady prostoty podane przez fizyków:

We are to admit no more causes of natural things than such as are both true and sufficient to explain their appearances. To this purpose the philosophers say that Nature does nothing in vain and more is in vain when less will serve; for Nature is pleased with simplicity and affects not the pomp of superfluous causes [I. Newton].

⁴Obviously then it would be better to assume a finite number of principles. Then should, in fact, be as few as possible, consistently with proving what has to be proved. (por. Arystoteles, w; The Works of Aristotle into English, Vol. 2. Clarendon).

Everything should be made as simple as possible but not simpler [A. Einstein].

Zarówno w tym „*but not simpler*”, jak i w klasycznym ujęciu sensu *brzytwy* – by nie mnożyć bytów bez potrzeby – kryje się samo sedno interpretacji: „Gdzie są granice tego *bez potrzeby*?”. Tu kryją się odpowiedzi na pytania typu: dlaczego proste teorie są często niedokładne, a teorie złożone są zgodne z doświadczeniem. W pracy pokażemy, że doprecyzowania w tym kontekście domaga się samo rozumienie funkcji *zgodne z doświadczeniem*, które pociąga za sobą dwa, okazuje się nie zawsze zintegrowane, skutki: dokładność fitowania danych empirycznych i zdolność do predykcji nowych.

Jedną z pierwszych rzeczy, którą trzeba ustalić w rozważaniach na temat prostoty, jest podanie podstawowych konotacji, w których to kryterium występuje. Wymienia się prostotę przyrody, prostotę praw, prostotę teorii naukowych, prostotę, metodologiczną [11]⁵. Prostota stosowana do charakterystyki teorii naukowych także objawia różne oblicza („versions of simplicity”): określa się ją w kategoriach liczby parametrów, prawdopodobieństwa, stabilności, zawartości treściowej, informatywności (Sober) czy falsyfikowalności (Popper). Tym dwóm ostatnim kategoriom poświęcimy w pracy nieco więcej miejsca.

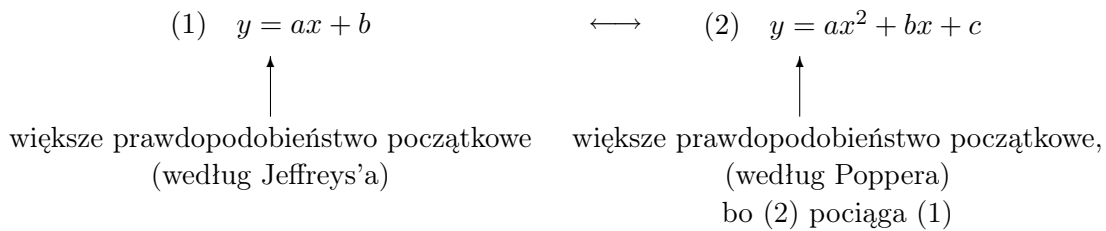
Prostota rozumiana była najczęściej w bardzo intuicyjny sposób. Newton [10, s. 121] łączy prostotę z liczbą przyczyn zaangażowanych w proces wyjaśniania. Wydaje się, że prostotę można analizować niezależnie od statystyki, ale już same intuicje Newtona dotyczące podstawowych cech badanego świata – „(...)Nature is pleased with simplicity (...)” – sugerują przeświadczenie o pewnej ekonomii, która przynajmniej zakłada większe prawdopodobieństwo.

Wśród podstawowych sposobów szacowania miary prostoty należy wymienić:

- Minimum Description Length [12].
- Kryterium informacyjne Akaike – *AIC* (Sober, Forster). Prezentowane szczegółowo w dalszej części pracy.
- Kryteria oparte na statystyce bayesowskiej. Tym samym można powiedzieć, że *brzytwa Ockhama* może być usprawiedliwiona matematycznie jako konsekwencja stosowania zasady logiki bayesowskiej: hipoteza wyrażona zdaniem zawierającym mniejszą liczbę parametrów uzyskuje większe prawdopodobieństwo *a posteriori*, z tej racji, że predykcje, na które pozwala, są bardziej ostre.

Gdy wybieramy krzywą, która ma przechodzić przez dwa dane punkty, rodzi się pytanie dlaczego wybierzemy postać $y = ax + b$, a nie $y = ax^2 + bx + c$. Ten prosty przykład można wykorzystać, by pokazać trudność w wykorzystaniu narzędzi probabilistycznych w analizie kryterium prostoty. Przedstawmy istotę znanego sporu Jeffreys’a z Popperem, który koncentrował się wokół prawa falsyfikacjonizmu wyrażonego w zdaniu: z jakimi cechami teorii związana jest jej mniejsza lub większa falsyfikowalność. Postawmy intuicyjnie zgodny następujący punkt wyjścia: prostsze hipotezy — większe prawdopodobieństwo; złożone — mniejsze prawdopodobieństwo. Jaskrawość interpretacji widoczna jest w ocenie *prior probability* dwóch powyższych równań: rodziny prostych i rodziny parabol.

⁵Jak się okazuje, tę deskrypcję można znacznie rozszerzać, mówiąc o prostocie strukturalnej, pojęciowej, dynamicznej, ontologicznej, epistemologicznej, formalnej, pragmatycznej, semantycznej, indukcyjnej, logicznej, itd.



W rzeczy samej, jedną z najlepszych prób reprezentacji problemu prostoty jest procedura wyboru najprostszej krzywej fitującej dane empiryczne. Sam Popper opisuje wyniki swojego sporu z H. Jeffreyssem [13, s. 117-118] w postaci następującego schematu:

- Jeffreys: prostota = szczupłość parametrów = wysokie prawdopodobieństwo początkowe;
- Popper: sprawdzalność = wysokie nieprawdopodobieństwo początkowe = szczupłość parametrów = prostota.

Intuicje Poppera zawierają prawie wszystkie elementy rozwijane później przez Akaike: a) odejście od pojmowania prostoty jedynie w kategoriach estetyki teorii i intuicyjnie pojmowanej praktyczności; b) wiązanie prostoty z pojęciem prawdopodobieństwa; c) niekonwencjonalne wiązanie prostoty z liczbą parametrów: wysuwanie przypuszczenia, że teorie matematycznie bardziej skomplikowane (ogólna teoria względności) są prostsze niż teorie o formalizmie mniej złożonym (teoria grawitacji Newtona). Popper, w odniesieniu do swojego rozumienia, prostoty cytuje ponadto tak ważną refleksję Weyla, że wymaga ona tutaj przytoczenia [13, s. 115]:

Dla przykładu załóżmy, że dwadzieścia przyporządkowanych par wartości (x,y) , należących do tej samej funkcji $y = f(x)$, leży (zgodnie z oczekiwaną dokładnością) na linii prostej po naniesieniu na wykres. Wysuniemy wówczas domniemanie, że mamy do czynienia ze ścisłym prawem przyrodniczym (...) byłoby nadzwyczaj nieprawdopodobne, by właśnie tych dwadzieścia par dowolnie wybranych obserwacji miało znajdować się na linii prostej, jeżeli prawo brzmiałoby inaczej.(...)Jest więc rzeczą zasadniczej wagi, że ową funkcję, a raczej klasę funkcji, matematycy winni zaproponować nam *a priori*, kierując się jej prostotą z matematycznego punktu widzenia.

Tę uwagę Weyla można uogólnić i pokazać, że kryje w sobie podstawowe zarzuty stawiane statystycznym szacunkom prostoty związanym z operowaniem pojęciem prawdopodobieństwa *a priori*:

- Co to znaczy, że na przykład prawa Keplera posiadają prawdopodobieństwo w świetle obserwacji?
- Co to znaczy, że uwzględniamy fakt, że to prawo ma prawdopodobieństwo przed rozpoczęciem obserwacji?

Prostota jako kryterium wyboru teorii odgrywa ważną rolę w rekonstruowaniu dynamiki nauki. Często zwraca się uwagę na znaczenie zmian, na które wpływa zastosowanie brzytwy Ockhama: rozwijanie samej teorii, jak i radykalna zmiana paradygmatu. Wspomniany J.

L. Rodríguez-Fernández [10] charakteryzuje skutki działania zasady prostoty jako swoiste obosieczne ostrze *brzytwy*, które potwierdza sukces lub przyczynia się do porażki teorii. W tym kontekście ważne jest zastrzeżenie P. Kowalca [14, s. 233], który poddaje krytycznej analizie próby zrównania prostoty i informatywności, by kryterium prostoty nie traktować jako wyłącznego a zwłaszcza finalnego w ocenie testowanych hipotez.

Bardzo ważne jest podkreślenie tego, co kryterium prostoty implikuje przede wszystkim: możliwość modelowania rzeczywistości przy użyciu ekonomicznych środków – najprostszych hipotez. Nie pociąga to natomiast za sobą wniosku, że sama natura jest prosta, lub że żyjemy w najprostszym ze światów. Kryje się tu idea *efektywności* w proponowanym wyjaśnianiu, co wiąże się naturalnie z ryzykiem uzyskania i operowania teorią nieprawdziwą. Już teraz warto, wyprzedzając tok naszych analiz, postawić tezę, że kwestia prawdziwości teorii czy modelu okaże się, w świetle analizowanych kryteriów ekonomiczności w wyjaśnianiu, zrelatywizowana jako jedna z kilku skorelowanych współrzędnych/wartości osiąganych jako cele operowania danym modelem. Niektórzy wysuwają bardziej ogólne przypuszczenie, że prostota pojawia się w kontekście uzyskanych rezultatów poznawczych jako wartość konkurencyjna lub komplementarna w stosunku do prawdy [11, s. 9]. Prace N. Cartwright [15] czy M. Morrison pokazały, że pojęcie prawdziwości teorii w hierarchii celów jej stawianych traci swoją rangę; mówi się chyba coraz częściej o teorii najlepszej, ale i zdanie tego typu nigdy nie uzyskuje cech obiektywności, lecz jest zawsze w kontekście pewnego systemu, czy nawet pewnej ontologii. W określaniu relacji między modelem a rzeczywistością szczególny nacisk jest położony na moc opisową i wyjaśniającą modelu. W ujęciu realistycznym jedną z fundamentalnych cech teorii fizycznej jest własność aproksymowania rzeczywistości [16]. Często aparat symboliczny używany w procesie konstruowania praw fizyki jest zbyt prosty, by reprezentować wszystkie aspekty badanej rzeczywistości. Nieuniknione stosowanie przybliżeń jest związane także ze stopniem złożoności sformułowań: „*The more complicated the laws becomes the greater its approximation*” [16, s. 149].

Powyższe rozważania dotyczą prostoty rozumianej jako wartość lub kryterium poznawcze. Te problemy, charakteryzowania pewnej odpowiedniości między teorią i doświadczeniem, wikłają w naszą dyskusję przynajmniej dwa stanowiska: konwencjonalizm i empiryzm. Konwencjonalizm wskazuje na to, że teoria i doświadczenie są nieizomorficzne ze sobą. Przypomnijmy w tym miejscu zgrabne ujęcie intuicyjne Einsteina: model ma być prosty, ale nie zbyt prosty; powiedzielibyśmy – prostota, ale nie za wszelką cenę. Należy pamiętać o czymś, co możnaby nazwać *zasadą adekwatności*, która wyraża się w przekonaniu, że teoria czy model winny posiadać podobny stopień złożoności, jak samo zjawisko, które chcemy opisać. Empirysta natomiast, który ogranicza wszystkie źródła wiedzy do doświadczenia, w tym fakcie widzi istotę problemu prostoty. Twierdzi, że do skutecznego uprawiania, na przykład fizyki nie potrzeba założeń filozoficznych. Czasem taki bezzałożeniowy empiryzm praktyczny nazywa się fundamentalizmem epistemologicznym [17, s. 68]. A. Grobler słusznie podkreśla tę wieloznaczność, która ujawnia się w momencie prób przypisania teorii cech prostoty. Szczególnie znamienna jest uwaga o falsyfikacjonizmie, który faworyzuje teorie prostsze w sensie liczby istotnych parametrów („hipoteza okrężnego ruchu planet jest prostsza od hipotezy orbit eliptycznych”) [17, s. 79]. Natomiast nasza analiza działania kryterium Akaike prowadzi do wniosku, w tym przypadku, zgoła innego. Być może to jest jedną z zalet *AIC* (które przecież karze model za nadmiarowe parametry), że jako prostszy wyodrębni model toru eliptycznego [por. Sekcja: „*AIC w działaniu*”].

Zdecydowanie najpoważniejszy problem, któremu trzeba stawić czoła, stosując kryteria prostoty oparte na analizach prawdopodobieństwa, pojawia się jako konsekwencja przyjęcia tezy, że dane empiryczne nie są w stanie całkowicie określić teorii. Mówimy o problemie indukcji, czyli możliwości jednoznacznego wyznaczenia prawa, które jest uogól-

nieniem danych empirycznych. Widać to już w podejmowanym problemie doboru krzywej, która jest reprezentacją wyników doświadczenia (jeśli zbiór danych przedstawimy jako zbiór punktów w układzie współrzędnych). Problem indukcji jest wyraziście eksplikowany w tzw. paradoksach potwierdzania (paradoksach konfirmacji). A. Grobler w swojej *Metodologii* poddaje je bardziej szczegółowej i interesującej analizie [17, s. 55-61]. Tutaj ich obecność jedynie sygnalizujemy, by przynajmniej ogólnie określić środowisko, w które wkracza kryterium Akaike.

3 Formalizm kryterium Akaike

Modele kosmologiczne, określone w terminach równań różniczkowych, posiadają pewne nieznanne parametry, które dofitowujemy z obserwacji - problem testowania modeli kosmologicznych (estymacji parametrów z obserwacji). Niezależnie od tego możemy porównywać różne modele z punktu widzenia ich opisu danych empirycznych. W tym kontekście ważną rolę odgrywa kryterium Akaike.

We współczesnej kosmologii obserwacyjnej mamy do czynienia z problemem degeneracji. Mianowicie wiele modeli kosmologicznych (opisujących Wszechświat) jest zgodnych z danymi empirycznymi. Aparat matematyczny, na którym bazuje się w selekcji i testowaniu modeli to metody statystyczne, a mówiąc ściślej analiza bayesowska⁶, [22, 23] w której rezygnuje się z pojęcia losowości na rzecz zdań prawdopodobnych. Rozważania prowadzi się w przestrzeni bayesowskiej:

$$(D, \Sigma(D), p(D, \theta)),$$

gdzie D – zbiór wartości danych empirycznych (wyniku obserwacji) X_i , gdzie $i = 1, \dots, N$; \tilde{D} – wektor hipotez (X_1, \dots, X_N) o wartościach w zbiorze D :

$$D = \{x_i\}, \quad i = 1, \dots, N, \quad \text{gdzie } N \text{ – liczba pomiarów}$$

$\Sigma(D)$ - σ -ciało podzbiorów D .

$p(D, \theta)$ - funkcja łącznej gęstości dla wektora obserwacji D i wektora parametrów θ :

$$p(D, \theta) = p(D|\theta) \cdot p(\theta), \quad D \subset R^n, \quad \text{gdzie } n \text{ – wymiar przestrzeni}$$

Jednym z kryteriów pozwalających na przeprowadzenie selekcji modeli, nie tylko kosmologicznych jest kryterium Akaike - *AIC* (Akaike Information Criterion) [24, 25]. Należy jednak zwrócić uwagę na to, że *AIC* jest estymatorem informacji *KL*, czyli funkcją pozwalającą oszacować, aproksymować wartość metryki Kullbacka-Leiblera. Kryterium Akaike⁷ ma zatem uzasadnienie na gruncie teorii informacji, a nie teorii bayesowskiej. Stąd nie jest ono kryterium bayesowskim⁸. Poddamy teraz *AIC* szczegółowej analizie i pokażemy jego ścisły związek z zasadą minimum entropii informacyjnej.

⁶Ujęcie statystyczne - operowanie na zmiennych losowych [18, 19, 20], ma swoich zwolenników. Należy do nich m. in. E. V. Linder [21], który preferuje fitowanie parametrów modeli przy użyciu metod statystycznych nad techniki bayesowskie selekcji modeli. Do zalet podejścia klasycznego (statystycznego) Linder zalicza m. in. (1) fitowanie parametrów nie wymaga wyboru rozkładów a priori (priorów), (2) jeżeli podzbiór danych empirycznych nie potwierdza parametrów (parametry nie fitują tego podzbioru), to może to oznaczać, że należy zbiór parametrów modelu rozszerzyć lub zmniejszyć, (3) modele mają charakter fizyczny tzn. ich parametry mają (lub powinny mieć) fizyczne znaczenie - nie są tylko pewnym sposobem reprezentowania wyników obserwacji.

⁷Podobnym kryterium do *AIC* jest *BIC* = $-2 \ln \tilde{L} + k \ln N$, z tym że *BIC* (Bayesian Information Criterion) jest kryterium bayesowskim (ma podstawy w teorii bayesowskiej). Szersze omówienie *BIC* można odnaleźć w artykułach [26, 27, 28].

⁸Bayesowską wersją *AIC* jest *EAIC* (Expected *AIC*), gdzie wyznacza się wartość oczekiwaną rozkładów a posteriori, a nie maximum funkcji wiarygodności jak w przypadku *AIC*.

3.1 Informacja Kullbacka-Leiblera

Kryterium Akaike jest estymatorem informacji (metryki, entropii) Kullbacka-Leiblera (KL), która określona jest wzorem:

$$I(f, g) = \begin{cases} \sum_i f(x_i) \ln \frac{f(x_i)}{g(x_i, \theta)}, i = 1, \dots, n & , f \text{ i } g \text{ rozkłady hipotez dyskretnych} \\ \int_{R^n} f(x) \ln \frac{f(x)}{g(x, \theta)} dx, x \in R^n & , f \text{ i } g - \text{rozkłady hipotez ciągłych} \end{cases} \quad (1)$$

Przez f rozumiemy rozkład prawdziwego - nieznanego modelu (*true model*), który jest aproksymowany przez rozkład g modelu teoretycznego M_i . Stąd f nie zależy od θ - parametrów M_i , natomiast g zależy od θ . Inaczej mówiąc, f definiuje się jako nieznaną prawdę, rzeczywistość (model prawdziwy), którą aproksymujemy modelem M_i , zaś g odzwierciedla model teoretyczny M_i . Uściślając:

$f(x)$ - rzeczywisty rozkład modelu (rozkład modelu prawdziwego). Tego rozkładu nie znamy,

$g(x, \theta)$ - rozkład modelu teoretycznego M_i , przy założeniu, że model M_i jest poprawny tzn. określone są jego parametry i rozkłady a priori (priory) tych parametrów.

Metryka KL jest stosowana do określania odległości pomiędzy dwoma rozkładami prawdopodobieństwa f i g . Wartość $I(f, g)$ przyjmuje zawsze wartości dodatnie, przy czym $I(f, g) = 0$, gdy $f = g$. Zauważmy ponadto, iż informacja Kullbacka-Leiblera nie spełnia warunku symetryczności, czyli $I(f, g) \neq I(g, f)$. Zatem nie jest to metryka w dokładnym sensie matematycznym. Entropia KL stanowi rozszerzenie entropii Shannona.

Definicja

Entropia Shannona (średnia ilość informacji) określona jest wzorem:

$$H(A) = \begin{cases} \sum_a p(a) I(a) = - \sum_a p(a) \log_2 p(a) & , A \text{ ma rozkład dyskretny} \\ \int_R p(x) I(x) = - \int_R p(x) \log_2 p(x) & , A \text{ ma rozkład ciągły} \end{cases}$$

gdzie A jest zmienną losową⁹ oraz

$$I(\circ) = \begin{cases} I(a) = \log_2 \frac{1}{p(a)} & , A \text{ ma rozkład dyskretny} \\ I(x) = \log_2 \frac{1}{p(x)} & , A \text{ ma rozkład ciągły} \end{cases}$$

$I(\circ)$ - ilość informacji potrzebna do scharakteryzowania zdarzenia \circ , którego miara probabilistyczna (prawdopodobieństwo) wystąpienia wynosi $p(a)$ lub $p(x)$.

Entropię KL rozumie się jako informację traconą, gdy rzeczywistość (model prawdziwy) aproksymujemy modelem M_i (mówiąc ściślej: gdy g jest używane do aproksymacji f). Biorąc to pod uwagę, metryki Kullbacka-Leiblera nie można używać do selekcji modeli, ponieważ zależy ona od nieznanego rozkładu f (nieznanej nam prawdy). Stąd konieczność stosowania estymatorów Kullbacka-Leiblera, do których zaliczana jest funkcja AIC (kryterium Akaike). W przypadku metryki KL zachodzi następująca zależność: im mniejsza wartość (mniejsza odległość pomiędzy f i g) $I(f, g)$ tym lepszy model teoretyczny M_i (model teoretyczny M_i lepiej aproksymuje model prawdziwy). Czyli najlepszy model teoretyczny to taki, który najlepiej aproksymuje model prawdziwy - odległość $I(f, g)$ jest najmniejsza. Ujmując to inaczej, im mniejsza jest odległość pomiędzy nieznaną prawdą, a modelem teoretycznym, tym bardziej jest on z nią zgodny.

⁹W podejściu bayesowskim hipotezą.

W kontekście naszych rozważań dotyczących selekcji modeli kosmologicznych skupimy się na postaci całkowitej metryki Kullbacka-Leiblera (założenie o jednostajnym (ciągłym) rozkładzie parametrów θ_i modelu M_i):

$$I(f, g) = \int f(x) \ln \frac{f(x)}{g(x, \theta)} dx \quad (2)$$

Bazując na (2) informację KL można zapisać w postaci:

$$I(f, g) = \int f(x) \ln f(x) dx - \int f(x) \ln g(x, \theta) dx = E_f[\ln f(x)] - E_f[\ln g(x, \theta)], \quad (3)$$

gdzie E_f - wartość oczekiwana obliczona dla funkcji $f(x)$.

Jeżeli $E_f[\ln f(x)] = const.$, czyli prawda jest ustalona, to minimalizacja $I(f, g)$ jest równoważna maksymalizacji $E_f[\ln g(x, \theta)]$ ¹⁰ - nieznanej prawdy. Korzystając z wzoru (3) i informacji zawartych w ostatnim zdaniu przedstawimy zarysowo dowód kryterium Akaike¹¹ przy następujących założeniach: (a) model prawdziwy to jeden z modeli z klasy rozważanych modeli teoretycznych, czyli $f = g(x, \theta)$, (b) nie znamy parametrów¹² modelu teoretycznego (aproksymującego), (c) szukamy estymatora $I(f, g)$ dla konkretnego modelu z (b) tzn. modelu o nieznanym parametrach:

Nasze zadanie polega na wyznaczeniu maksimum i minimum $E_f[\ln g(x, \theta)]$: nie wchodząc w zawiłe szczegóły dowodowe¹³ [31] podajmy, iż nieobciążony estymator $[E_f[\ln g(x, \theta)]]_{\theta=\hat{\theta}}$ dla danego modelu może przyjąć postać $\hat{E}_f = \frac{1}{N}(\ln \tilde{L} - k)$ tzn.

$$E[\hat{E}_f] = [E_f[\ln g(x, \theta)]]_{\theta=\hat{\theta}} \quad ,$$

gdzie \tilde{L} - maximum funkcji wiarygodności, N - liczba danych empirycznych, k - liczba parametrów modelu, $\hat{\theta}$ - estymator wektora θ . Przy ustalonym N : $\max[E_f[\ln g(x, \theta)]]_{\theta=\hat{\theta}} = \max(\ln \tilde{L} - k) = \ln \tilde{L} - k$. Wykazaliśmy zatem, że entropia KL przyjmuje wartość minimalną dla ustalonej liczby obserwacji, gdy $\max[E_f[\ln g(x, \theta)]]_{\theta=\hat{\theta}} = \ln \tilde{L} - k$.

Przy danych $g(x, \theta)$ oraz $f(x)$ ¹⁴ istnieje taka wartość θ , że metryka K-L przyjmuje minimalną wartość - **zasada minimalnej entropii**. Podamy teraz fakt [32], który ma decydujące znaczenie przy wyprowadzeniu kryterium Akaike (zarys dowodu znajduje się powyżej): Przy maksymalnej gęstości prawdopodobieństwa estymatora $\hat{\theta}$ dla wektora θ zachodzi warunek: wektor θ ma taką wartość, że $I(f, g)$ osiąga wartość minimalną. Czyli ujmując to prościej kryterium Akaike jest wyprowadzone przez estymację (aproksymację) minimalnej entropii Kullbacka-Leiblera.

¹⁰Na mocy prawa wielkich liczb za estymator wartości oczekiwanej funkcji można przyjąć wartość średnią z wartości tej funkcji dla posiadanych obserwacji. Stąd estymatorem $E_f[\ln g(x, \theta)]$ - nieznanej prawdy jest $\hat{E} = \frac{1}{n} \sum_{k=1}^n \ln g(x_k, \theta)$ - wielkość znana, n - liczba danych empirycznych. Zatem rozważania matematyczne prowadzimy w oparciu o estymator $E_f[\ln g(x, \theta)]$.

¹¹Podajmy, iż dowód AIC często przeprowadza się w sposób analogiczny, ale przy założeniu, że dane empiryczne to zmienne losowe. Wówczas $\hat{\theta}$ to statystyka, a więc entropia K-L to też statystyka. W takim podejściu analizie podlega $E_f[I(f, g)]$, a nie $I(f, g)$.

¹²Aby je określić klasę modeli zawięzamy do jednego modelu, a następnie wykonujemy dosyć zawiłe zabiegi matematyczne, które pomijamy.

¹³Dowód, w którym rezygnuje się z założenia (a), zaproponował T. Takeuchi [29]. Uzasadnienie AIC może być również przeprowadzone w formalizmie bayesowskim [30].

¹⁴Jest to wielkość nieznaną, zatem nie możemy skorzystać z niej w rozważaniach matematycznych. W związku z tym, żeby uzyskać taką możliwość możemy: (1) włączyć model prawdziwy do zbioru modeli aproksymujących (teoretycznych) tzn. $f(x) = g(x, \theta)$ lub (2) wyznaczyć estymator $f(x)$ (o ile da się to zrobić).

3.2 Różne postacie AIC

3.2.1 W terminach funkcji wiarygodności

$$AIC = -2 \ln \tilde{L} + 2k, \tilde{L} = \max\{L(\theta_i, D) \geq 0, \quad (4)$$

gdzie $L(\theta_i, D)$ - funkcja wiarygodności dla parametrów modelu θ_i i danych empirycznych D (mówiąc ściślej zbioru wartości danych empirycznych \tilde{D}),

k - liczba składowych wektora parametrów θ modelu (liczba parametrów modelu).

Zauważmy, iż we wzorze na AIC występują dwa wyrażenia. Podamy teraz ich funkcję:

\tilde{L} - określa stopień fitowania modelu (zgodność modelu z danymi empirycznymi),

$2k$ - określa stopień złożoności (skomplikowalności) modelu.

Zatem $2k$ interpretujemy jako *czynnik karzący* (*penalty term*) lub używając terminologii filozoficznej *czynnik Ockhama* (*Ockham factor*), czyli *wyrażenie karzące modele za posiadanie dodatkowych (nadmiarowych) parametrów*.

Uściślając mamy do czynienia z sytuacją, gdy model scharakteryzowany przez parametry istotne jest zgodny z danymi empirycznymi (jak najmniejsza wartość AIC), natomiast model zawierający oprócz parametrów istotnych parametry dodatkowe (nieistotne) gorzej fituje wyniki obserwacji (dane empiryczne) tzn. AIC osiąga stosunkowo dużą wartość. Odwołując się do formalizmu matematycznego zachodzi następujący warunek:

Dla modeli określonych za pomocą parametrów istotnych funkcja AIC jest malejąca, natomiast dla modeli zawierających oprócz parametrów istotnych parametry dodatkowe (nieistotne) funkcja AIC jest rosnąca.

Czynnik karzący może być zapisany w terminach *evidencji* (*evidence*)¹⁵, przy założeniu, że $p(\theta|D, M)$ jest nie znormalizowaną gęstością¹⁶ dla wektora parametrów θ . Wtedy:

$$E = p(D|M) = \int p(\theta|D, M)d\theta = \tilde{L} \cdot (2\pi)^{\frac{d}{2}} \sqrt{\det C} \cdot p(\hat{\theta}|M) = \tilde{L} \cdot OF, \quad (5)$$

gdzie OF - *penalty term* (*Ockham factor*), $\hat{\theta} = \theta_{MOD}$, \tilde{L} - maximum funkcji wiarygodności, d - liczba składowych θ , C - macierz kowariancji¹⁷, $\hat{\theta}$ - estymator θ , $p(\hat{\theta}|M)$ - gęstość a priori estymatora θ , θ_{MOD} - moda $p(\theta|D, M)$, czyli wektor, w którym gęstość $p(\theta|D, M)$ osiąga maximum. Według ewidencji E , określonej wzorem (5), bardziej złożone modele (o większej liczbie parametrów istotnych) zajmują większą objętość przestrzeni wektora parametrów θ niż modele mniej złożone (o mniejszej liczbie parametrów istotnych).

Jeżeli θ ma jedną składową (θ jest hipotezą jednowymiarową) o gęstości a priori $p(\theta|M) = \frac{1}{\Delta\theta}$, gdzie $\Delta\theta$ to przedział dozwolonych wartości dla rozważanego parametru, to $OF = \frac{2\pi\sigma}{\Delta\theta} = \frac{V_{posterior}}{V_{prior}}$, czyli jest to stosunek objętości zajmowanej przez gęstość a posteriori θ - $p(\theta|D, M)$ do objętości zajmowanej przez gęstość a priori θ - $p(\theta|M)$ w przestrzeni wektora θ . Ponadto zachodzi zależność $\log \frac{V_{posterior}}{V_{prior}} \equiv I$, gdzie I interpretuje się jako ilość informacji o parametrach modelu po zgromadzeniu danych empirycznych¹⁸. Zauważmy, iż im więcej mamy informacji o parametrach modelu na podstawie obserwacji, tym większy jest $OF = \frac{V_{posterior}}{V_{prior}}$ ($V_{posterior}$ rośnie, a V_{prior} maleje), a tym samym większa ewidencja ($E = \tilde{L} \cdot OF$) dla tego modelu. Stosunek $\frac{V_{posterior}}{V_{prior}}$ można również rozumieć jako objętość

¹⁵Wyprowadzenie można znaleźć w artykule [33].

¹⁶tzn. $\int_{\mathbb{R}^n} p(\theta|D, M)d\theta \neq 1$ - θ rozkład ciągły lub $\sum_i p(\theta_i|D, M) \neq 1$ - θ rozkład skokowy.

¹⁷Na przekątnej znajdują się wariancje odpowiednich składowych, natomiast na pozostałych miejscach kowariancje odpowiednich składowych.

¹⁸Stąd $\log OF \approx I$.

w przestrzeni parametrów nie wykorzystywaną (wasted - marnowaną) przez model, czyli niepotwierdzoną przez obserwacje. Zatem im mniejszy jest OF , tym więcej objętości jest nie wykorzystywanej, a stąd ewidencja dla modelu staje się coraz mniejsza ($E = \tilde{L} \cdot OF$ maleje, bo OF maleje).

3.2.2 W terminach funkcji χ^2

Celem podania innej postaci kryterium Akaike musimy zdefiniować błędy gaussowskie. Z dokonywaniem obserwacji (pomiarów) dowolnej wielkości fizycznej przyjmującej wartości rzeczywiste wiąże się występowanie błędów pomiarowych, które rozumiemy wtedy jako hipotezy o rozkładzie normalnym. Wprowadźmy zatem pojęcie błędów gaussowskich [34]: Niech e będzie hipotezą oznaczającą błąd obserwacji (eksperymentu) dowolnej wielkości fizycznej przyjmującej wartości rzeczywiste. Hipotezę e można zapisać jako sumę hipotez e_i (modelować jako sumę mniejszych błędów (wkładów) e_i):

$$e = \sum_i e_i$$

gdzie e_i ma rozkład normalny (Gaussa) oraz e_i spełniają centralne twierdzenie graniczne (CTG) tzn. zachodzi warunek:

$$\bigwedge_{x \in R} P\left(\frac{e_i - E(e_i)}{\sqrt{Var(e_i)}} \leq x\right) \longrightarrow \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt$$

Jeśli założymy gaussowski rozkład błędów, to zachodzi następująca zależność:

$$\tilde{L} \propto \exp\left(-\frac{\chi^2}{2}\right) \Rightarrow \chi^2 \propto -2 \ln \tilde{L} \Leftrightarrow \chi^2 = b[-2 \ln \tilde{L}], b > 0 \quad (6)$$

Stąd AIC przyjmuje postać:

$$AIC = \chi^2 + 2k, \chi^2 - \text{efektywna funkcja } \chi^2 \quad (7)$$

gdzie $\chi^2 = \sum_{i=1}^N X_i^2$, X_i - niezależne składowe wektora obserwacji (hipotezy) o takim samym rozkładzie $N(0, \sigma)$. Zatem AIC ma rozkład Gaussa, czyli jest opisywany przez krzywą Gaussa. Gdy $k = 0$, czyli model nie jest określony przez żaden parametr, to $AIC = \chi^2$.

Wartość χ^2 odgrywa istotną funkcję w estymacji parametrów i ocenie jakości dofitowania. Służy ona do wyboru parametrów modelu w taki sposób, aby jak najlepiej fitował on dane empiryczne. Innymi słowy wartość χ^2 odpowiada za znajdowanie najlepszych parametrów modelu, ale nie można jej stosować do selekcji modeli. Zauważmy, iż dla różnych rozkładów danych empirycznych otrzymujemy różne postacie funkcji wiarygodności¹⁹, a stąd różne wzory na χ^2 ($\chi^2 \propto -2 \ln \tilde{L}$).

3.2.3 W terminach krzywych dofitowanych do danych empirycznych

Kryterium Akaike można zapisać w innej postaci, mianowicie w terminach krzywych (funkcji) dopasowanych do danych empirycznych. W tym celu omówiony zostanie problem fitowania krzywych (curve fitting problem) [36]. Najbardziej ogólna forma tego zagadnienia występuje w wielu kontekstach. W tym przypadku przedstawimy je w prostej postaci²⁰:

¹⁹Można się z nimi zapoznać w artykule [35].

²⁰W takim sensie, iż N - liczba składowych wektora obserwacji wynosi 2. Czyli rozważamy dwie dane empiryczne.

Dysponujemy pewnymi danymi empirycznymi x i y ²¹, które umieszczamy w układzie kartezjańskim o osiach x i y . Formułujemy hipotezę o postaci funkcji jednej zmiennej określającej związek pomiędzy tymi danymi w rozpatrywanym układzie kartezjańskim. Rozważmy przykłady takich hipotez:

(1) związek pomiędzy danymi ma charakter liniowy tzn. $y = a_1x + a_0 + N(0, \sigma^2)$, $a_1, a_0 = const.$ - parametry krzywej, $N(0, \sigma^2)$ - błąd pomiaru.

(2) relacja pomiędzy danymi jest kwadratowa tzn. $y = a_2x^2 + a_1x + a_0 + N(0, \sigma^2)$, $a_1, a_0, a_2 = const.$ - parametry krzywej, $N(0, \sigma^2)$ - błąd pomiaru.

Zauważmy, iż bez podania konkretnych wartości stałych w funkcji określającej związek pomiędzy danymi mamy do czynienia z rodziną krzywych (family of curves) - w naszych przykładach rodziną linii prostych i rodziną parabol. Uogólnienie rozpatrywanego problemu (problemu fitowania krzywych) polega na rozważeniu N składowych wektora obserwacji o wartościach w zbiorze N elementowym. Wtedy funkcja opisująca relację pomiędzy obserwacjami będzie funkcją $N - 1$ zmiennych. Metodą pozwalającą na dopasowanie rodziny krzywych (w szczególności krzywej) do danych jest metoda najmniejszych kwadratów polegająca na mierzeniu odległości pomiędzy rodziną krzywych a obserwacjami - im mniejsza jest ta odległość, tym lepiej rodzina krzywych fituje dane. Uogólnieniem tej metody jest metoda największej wiarygodności, która może być stosowana przy dwóch założeniach: (1) dane empiryczne obarczone są błędem, (2) hipoteza dotycząca postaci funkcji określającej relację pomiędzy danymi jest poprawna. Metodę tą można wyrazić w następujący sposób: krzywa najlepiej fitująca dane empiryczne to taka krzywa, która jest najbardziej prawdopodobna.

AIC jako kryterium służące do selekcji modeli w terminach krzywych dopasowanych do obserwacji rozumie się jako selekcję rodziny krzywych na podstawie danych empirycznych. Mianowicie:

$$AIC = -2 \ln L(F) + 2k, \quad (8)$$

gdzie F - rodzina krzywych, $L(F)$ - maksimum funkcji wiarygodności dla F , k - liczba parametrów (stałych w funkcjach wyrażających relację pomiędzy danymi) rodziny krzywych, $2k$ - określa stopień złożoności rodziny krzywych - wyrażenie karzące rodzinę krzywych za posiadanie dużej liczby parametrów (im większe k tym większy stopień złożoności). AIC można określić jako funkcję karzącą maksymalnego prawdopodobieństwa (*penalised maximum likelihood function*). Najlepsza rodzina krzywych to taka, dla której AIC osiąga minimalną wartość.

3.3 Funkcja AIC

Kryterium Akaike [37] stosuje się do odróżniania istotnych parametrów modelu, dzięki którym fituje on dobrze dane empiryczne od parametrów dodatkowych (*extra parameters*) zwiększających jedynie ogólność modelu, ale nie mających znaczącego wpływu na poprawę dopasowania modelu. Innymi słowy, AIC służy do filtracji nadmiarowych parametrów, które w nie istotny sposób poprawiają dopasowanie modelu do danych empirycznych.

Najlepszy model to taki, który minimalizuje AIC (AIC osiąga najmniejszą wartość) - minimalizuje informację Kullbacka-Leiblera (patrz wzór (1)) tzn. minimalizuje ilość informacji utraconych, gdy model prawdziwy - nieznaną prawdą, rzeczywistość jest aproksymowana przez model najlepszy. Zatem kryterium Akaike pozwala stwierdzić, który z rozważanych modeli najlepiej aproksymuje rzeczywistość - jest najbardziej zgodny z danymi empirycznymi (najlepiej fituje dane empiryczne).

²¹Uściślając są to wartości wektora obserwacji złożonego z 2 składowych.

AIC może być stosowane, gdy dysponujemy odpowiednio dużą liczbą wyników obserwacji. Najczęściej przyjmuje się, iż liczba danych empirycznych musi być duża²² w porównaniu z liczbą parametrów najbardziej złożonego (skomplikowanego) modelu z klasy rozważanych modeli. Ponadto wykazuje się²³, iż kryterium Akaike ma tę własność, że gdy do zbioru rozpatrywanych modeli włączymy model prawdziwy, czyli najlepszy (najbardziej poprawny), to nawet dla bardzo dużej liczby obserwacji *AIC* nie zawsze go wskaże jako najlepszy. *AIC* zatem nie jest kryterium konsystentnym (niesprzecznym, spójnym)²⁴.

AIC daje słabe wyniki, gdy²⁵:

- (1) model słabo fituje dane empiryczne bez względu na liczbę parametrów.
- (2) dane empiryczne są zbyt słabe, żeby ograniczyć (wylimitować) parametry nadmiarowe (w nie istotny sposób poprawiające fitowanie modelu) modelu.

Kryterium Akaike odwołuje się do *brzytwy Ockhama (Occam razor)*. Mianowicie gdy dwa modele fitują dane empiryczne w równym stopniu (równie dobrze), czyli $\tilde{L}_1 = \tilde{L}_2$, to preferowany jest model prostszy tzn. z mniejszą liczbą parametrów (mniejsza wartość *AIC*):

$$\begin{aligned} AIC_1 &= -2 \ln \tilde{L}_1 + 2k_1, \\ AIC_2 &= -2 \ln \tilde{L}_2 + 2k_2, \end{aligned}$$

W przypadku, gdy $k_1 < k_2$, to $AIC_1 < AIC_2$.

Widać stąd, iż *AIC* rozpatruje się w kontekście prostoty²⁶ [por. [38]] - modele prostsze (o mniejszej liczbie parametrów) są bardziej zgodne z danymi empirycznymi niż modele bardziej złożone (o większej liczbie parametrów).

Dla *AIC* można zdefiniować następującą wielkość²⁷:

$$\begin{aligned} \Delta AIC_i &\equiv AIC_i - AIC_{\min}, \quad i = 1, \dots, K, \\ \Delta AIC_i &= 0 \quad \text{dla najlepszego modelu} \quad (AIC_i = AIC_{\min}) \\ \Delta AIC_i &> 0 \quad \text{dla pozostałych modeli} \quad (AIC_i > AIC_{\min}) \end{aligned} \quad (9)$$

gdzie

AIC_i - wartość funkcji *AIC* dla *i*-tego modelu,

$AIC_{\min} = \min\{AIC_i, i = 1, \dots, K\}$ - najlepszy model wybrany spośród *K*-modeli.

ΔAIC_i można interpretować jako ilość utraconej informacji, gdy nieznaną rzeczywistość (model prawdziwy) aproksymujemy modelem *i*, a nie najlepszym modelem z rozważanego zbioru modeli. Można ją zastosować do selekcji modeli, mianowicie im większe ΔAIC_i dla *i*-tego modelu, tym mniejsze prawdopodobieństwo, iż jest on aproksymacją modelu prawdziwego. Ujmując to dokładniej:

$\Delta AIC_i \in (0, 2)$ - model *i* ma prawie takie same potwierdzenie empiryczne jak model najlepszy (ważność *i*-tego modelu względem modelu najlepszego jest znacząca).

$\Delta AIC_i \in (2, 4)$ lub $\Delta AIC_i \in (4, 7)$ - model *i* ma mniejsze potwierdzenie empiryczne od

²²Próbka jest stosunkowo mała dla $\frac{N}{k} \leq 40$, *N* - liczba danych empirycznych.

²³Mówiąc ściślej symulacje komputerowe.

²⁴Konsystentna wersja *AIC* to *CAIC* (sformułowana przez Bozdogana), mianowicie: $CAIC = -2 \ln \tilde{L} + k(\ln N + 1)$.

²⁵Te warunki odnoszą się także do *BIC*.

²⁶Pojęcie prostoty jest niejednoznaczne. Omówienie różnych jego ujęć w filozofii nauki można odnaleźć w pozycji [11].

²⁷Interpretuje się ją jako *quick 'strenght of evidence' for considered model with respect to the best one* czyli szybkość fitowania (siły dowodu - siły ewidencji) modelu *i* w odniesieniu (w stosunku) do modelu najlepszego (AIC_{\min}) Inaczej mówiąc jest to pewna skala mówiąca nam o ile dany model jest gorszy od modelu najlepszego.

modelu najlepszego (ważność i -tego modelu względem modelu najlepszego jest widocznie mniejsza).

$\Delta AIC_i > 10$ - model i jest praktycznie nie dopasowany do danych empirycznych (brak ważności i -tego modelu względem modelu najlepszego).

W przypadku gdy N - liczba składowych wektora danych empirycznych (liczba danych empirycznych) jest mała: $\frac{N}{k} \leq 40$ kryterium Akaike przyjmuje postać (Sugiura 1978) [39]:

$$AIC_c = AIC + \frac{2k(k+1)}{N-k-1} \quad (10)$$

Zauważmy, iż dla dużych N ($N \gg k$): $AIC_c = AIC$. Natomiast dla małych N ($N \sim k$) czynnik karzący, czyli $2k + \frac{2k(k+1)}{N-k-1}$ jest większe niż czynnik karzący dla AIC .

Względna siła ewidencji (dowodu) modeli²⁸ (dopasowanie do danych empirycznych) może być modelowana jako maksimum z następującej funkcji wiarygodności [32]:

$$\tilde{L}_{wA} = \max\{L(g_i, D)\} \propto \exp\{-\frac{1}{2}\Delta AIC_i\}, \quad g_i\text{- rozkład modelu teoretycznego } M_i \quad (11)$$

gdzie \tilde{L}_{wA} nazywane są *wagami Akaike (Akaike weights)* - w_i . Odpowiednik wag Akaike w podejściu bayesowskim (formalizmie bayesowskim) to $p(M_i|D)$, przy założeniu, że wszystkie rozważane modele są jednakowo prawdopodobne, mianowicie:

$$p(M_i|D) = \frac{\exp\{-\frac{1}{2}\Delta AIC_i\}}{\sum_{r=1}^K \exp\{-\frac{1}{2}\Delta AIC_r\}} \equiv w_i \quad (12)$$

Z ostatniego wzoru widać, iż model minimalizujący AIC to model o największym $p(M_i|D)$, natomiast im większe ΔAIC_i , tym mniejsze $p(M_i|D)$.

Używając AIC_c wzór na wagi Akaike przyjmuje postać [28]:

$$w_i = \frac{\exp\{-\frac{1}{2}\Delta AIC_{c,i}\}}{\sum_{r=1}^K \exp\{-\frac{1}{2}\Delta AIC_{c,r}\}} \quad (13)$$

gdzie K - liczba modeli, $\Delta AIC_{c,i} = AIC_c - AIC_i$, $\Delta AIC_{c,r} = AIC_c - AIC_r$, AIC_i - wartość AIC dla i -tego modelu, AIC_r - wartość AIC dla r -tego modelu.

Istotną wielkością jest *savvy prior*, który wprowadza się jeśli chcemy interpretować wagi Akaike za pomocą bayesowskich posteriorów. Wielkość ta zależy od ilości danych oraz liczby parametrów i określana jest wzorem:

$$p(M_i) = C \exp\{\frac{1}{2}\Delta BIC_i - \frac{1}{2}\Delta AIC_i\} = \frac{\exp[k_i(\frac{1}{2}\ln N - 1)]}{\sum_{j=1}^K \exp[k_j(\frac{1}{2}\ln N - 1)]}, \quad (14)$$

gdzie

$$C = \frac{1}{\sum_{k=1}^K \exp\{\frac{1}{2}\Delta BIC_i\} \exp\{-\frac{1}{2}\Delta AIC_i\}},$$

gdzie $\Delta BIC_i = BIC_i - BIC_{\min}$, $\Delta AIC_i = AIC_i - AIC_{\min}$, $BIC = -2\ln \tilde{L} + k \ln N = \chi^2 + k \ln N$, k - liczba parametrów modelu, N - liczba danych empirycznych. Najlepszy model to taki, dla którego BIC osiąga najmniejszą wartość.

Dla ustalonej liczby danych empirycznych $p(M_i)$ zależy tylko od liczby parametrów M_i . Prior jest większy dla modeli o większej liczbie parametrów - preferowane są a priori

²⁸relative strenght of evidence for models.

modele bardziej złożone²⁹. Dla *savvy prior* zachodzi następująca zależność: im większa liczba danych, tym więcej informacji, z których może skorzystać model z większą liczbą parametrów. Zatem ma on uzasadnienie w teorii informacji. Najlepszy model powinien zmieniać się wraz ze zmianą liczby obserwacji.

3.4 Krytyka *AIC*

Podamy główne trudności związane z kryterium Akaike:

3.4.1 Problemy z $\tilde{L} = \max\{L(\theta_i, D)\}$ - maximum funkcji wiarygodności

Nie każdy problem można rozwiązać stosując selekcję modeli. Wiele zagadnień wiąże się z testowaniem modeli - estymacją ich parametrów (przybliżaniem, aproksymowaniem wartości tych parametrów). Dla tych kwestii odpowiednikiem kryterium Akaike jest metoda największej wiarygodności polegająca na wyznaczaniu maximum funkcji wiarygodności (maximum likelihood) - *ML*. Z *ML* związana jest następująca trudność: często zawodzi w oddzielaniu istotnych informacji, które niosą ze sobą dane od szumu - błędów pomiarowych. Występowanie tej tendencji zaznacza się wyraźnie w 3 przypadkach: (1) liczba danych empirycznych w stosunku do liczby parametrów estymowanych jest mała, (2) liczba danych empirycznych jest mała, (3) liczba estymowanych parametrów jest duża tj. rośnie wraz z liczbą danych empirycznych.

Rozważmy przykład wspomnianej wyżej trudności: niech $f(x)$ będzie jednowymiarową gęstością Gaussa³⁰ o średniej μ i odchyleniu standardowym σ . Wtedy funkcja wiarygodności przyjmuje postać:

$$\begin{aligned} L &= \prod_{i=1}^N \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x_i - \mu)^2}{2\sigma^2}\right\} \Rightarrow \\ &\Rightarrow \ln L = -\sum_{i=1}^N \left(-\frac{1}{2} \ln 2\pi - \ln \sigma - \frac{1}{2} \left[\frac{(x_i - \mu)}{\sigma}\right]^2\right) \end{aligned}$$

Stąd:

$$\frac{\partial \ln L}{\partial \mu} = 0 \Leftrightarrow \hat{\mu}_{ML} = \frac{\sum_{i=1}^N X_i}{N} = \bar{X} \quad \text{oraz} \quad \frac{\partial \ln L}{\partial \sigma} = 0 \Leftrightarrow \hat{\sigma}_{ML}^2 = \sum_{i=1}^N \frac{(X_i - \bar{X})^2}{N}$$

gdzie $\hat{\sigma}_{ML}^2$ jest estymatorem obciążonym tzn. $E[\hat{\sigma}_{ML}^2] \neq \sigma^2$. Dla dużej liczby danych empirycznych obciążenie jest małe, natomiast dla małej liczby wyników obserwacji obciążenie jest znaczące.

Do maximum funkcji wiarygodności nawiązuje problem Neymana-Scotta [40]. Mianowicie mówi on jak sprawdzają się estymatory *ML* w sytuacji gdy liczba danych w stosunku do estymowanych parametrów jest mała³¹. Uściślając: (X_n) - wektor obserwacji, $n = 2N$, gdzie dla każdego n : $X_n = (X_k, X_l), k, l = 1, \dots, N$ oraz $\{x_{i1}, x_{i2} : i = 1, \dots, N\}$ to zbiór wartości (X_n) . Zakładamy, iż x_{i1} oraz x_{i2} pochodzą z populacji o średniej μ_i i odchyleniu standardowym σ (niezależnym od i). Problem Neymana-Scotta dotyczy estymacji μ_i i σ . Z trudnością Neymana-Scotta wiąże się zagadnienie dokładności predykcji (*predictive accuracy*). Przed jego zarysowym omówieniem należy wprowadzić wielkość zwaną estymatorem

²⁹Nie jest to zgodne z brzytwą Ockhama.

³⁰Rozkład Gaussa ma swoje kołowe (angular) odpowiedniki - kołowy rozkład von Misesa (von Mises circular distribution) i sferyczny rozkład von Misesa i Fishera (von Mises-Fisher spherical distribution).

³¹Liczba estymowanych parametrów rośnie wraz z liczbą wyników obserwacji.

MEKLD (Minimum Expected Kullback-Leibler Distance), który określa się również jako rozkład predyktywny (predictive distribution). Estymator ten wyznacza się maksymalizując funkcję wiarygodności (a tym samym logarytm funkcji wiarygodności) nie odwołując się tylko do dostępnych wyników obserwacji, ale też do danych, które możemy potencjalnie uzyskać z rozważanego źródła. Problem dokładności predykcji w kontekście estymatora *MEKLD* można przedstawić w następujący sposób: jeżeli dysponujemy pewnymi danymi empirycznymi w problemie Neymana-Scotta i naszym celem jest minimalizacja błędów związanych z pobraniem nowych danych, to należy zastosować estymator *MEKLD*. Innymi słowy: jeżeli chcemy maksymalizować dokładność predykcji, to należy minimalizować rozkład predyktywny.

Wykazuje się [41], że *AIC* zawodzi w kwestii dokładności predykcji. Uściślając: w problemie Neymana-Scotta tzn. w sytuacji gdy liczba danych jest mała w porównaniu do estymowanych parametrów mamy: $\hat{\sigma}_{ML}^2 \rightarrow \frac{1}{2}\sigma^2$ oraz $\hat{\sigma}_{MEKLD}^2 \rightarrow \frac{3}{2}\sigma^2$. Oznacza to, że estymator *AIC* (na gruncie estymacji parametrów miejsce *AIC* zajmuje metoda największej wiarygodności) $\hat{\sigma}_{ML}^2$ i estymator *MEKLD* $\hat{\sigma}_{MEKLD}^2$ nie są zbieżne do tej samej wielkości, a stąd *AIC* nie dostarcza dokładnych predykcji. Ponadto *ML* i *MEKLD* są niekonsystentne³² (nie spójne, sprzeczne), gdyż *ML* ponad-fituje dane (over-fits the data) z $\hat{\sigma}_{ML}^2 n \rightarrow \frac{1}{2}\sigma^2$ ($\hat{\sigma}_{ML}^2$ nie dąży do σ^2), natomiast *MEKLD* pod-fituje dane (under-fits the data) z $\hat{\sigma}_{MEKLD}^2 n \rightarrow \frac{3}{2}\sigma^2$ ($\hat{\sigma}_{ML}^2$ nie dąży do σ^2).

3.4.2 Problemy z czynnikiem karzącym (czynnikiem Ockhama)

(a) jednowymiarowy wielomian regresji (univariate polynomial regression)³³

Zagadnienie³⁴ to polega na wyborze stopnia wielomianu aproksymującego nieznaną funkcję. Dokładniej: niech $t(x)$ będzie pewną nieznaną funkcją, $x_n, y_n : n = 1, \dots, N$ zbiorem wartości wektora obserwacji, gdzie $x \in [-1, 1]$ i $y_n = t(x_n) + \varepsilon_n, n = 1, \dots, N, \varepsilon_n$ - błędy pomiarowe mające rozkład Gaussa o średniej równej zero i nieznannej wariancji. Naszym zadaniem jest skonstruowanie pewnej funkcji wielomianowej $f(d, x)$ ³⁵ stopnia d , której używa się do przewidywania wartości nieznannej funkcji $t(x)$, gdzie $x \in [-1, 1]$. Miarą sukcesu tej aproksymacji jest *ESPE* (Expected Squared Prediction Error), czyli średnia wartość $[f(d, x) - t(x)]^2$:

$$ESPE[f(d, x)] = \frac{\sum_{m=1}^M [f(d, x_m) - t(x_m)]^2}{M} \quad (15)$$

gdzie $\{x_m : m = 1, \dots, M\}$ i $x_m \in [-1, 1]$.

(b) problemy *AIC* z gapps models.

Przypomnijmy, iż przeważnie zakłada się, że modele mają rozkład jednostajny (są jednakowo prawdopodobne), którego funkcja gęstości ma postać:

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{dla } x \in [a, b] \\ 0, & \text{dla } x \notin [a, b] \end{cases}$$

³²Estymator jest konsystentny (spójny, niesprzeczny), gdy przy $n \rightarrow \infty$ (n - liczba składowych wektora obserwacji) dąży do tej samej wielkości co prawdziwy rozkład danych empirycznych (rozkład uwzględniający wszystkie wyniki obserwacji - wyniki obserwacji dążą do nieskończoności).

³³Rozważa się przypadki wielowymiarowe.

³⁴Z jednowymiarowego wielomianu regresji korzysta się w ekonometrii, mianowicie w ekonometrycznej autoregresji [36].

³⁵W zależności od wyników obserwacji *AIC* wybiera różne stopnie funkcji wielomianowych.

Gappy model to taki model, dla którego przedział $[a, b]$ zapisuje się jako sumę rozłącznych przedziałów. Niech: (a) dysponujemy N danymi empirycznymi (wektor obserwacji zawiera N składowych), (b) średnia odległość pomiędzy kolejnymi przedziałami wynosi $\frac{1}{N}$, (c) największa odległość pomiędzy kolejnymi przedziałami wynosi $\frac{\ln N}{N}$, (d) $b = a + \frac{\ln N}{N}$. Przy zachodzeniu warunków (a) - (d) można pokazać, iż *AIC* preferuje gapps models nad modele, dla których interwału $[a, b]$ nie rozbijamy na sumę rozłącznych przedziałów.

D. L. Dowe, S. Gardner, G. Oppy [36] pokazują, że z analizowanymi trudnościami lepiej niż *AIC* radzą sobie estymatory *MML* (Minimum Message Length) - patrz Criticism of *AIC*. Mówiąc ściślej wykazuje się to dla estymatorów wyprowadzonych na podstawie aproksymacji jednego z estymatorów *MML* - *SMML* (Strict Minimum Message Length). Wyznaczanie estymatorów *MML* bazuje na zasadzie *MML*. Według niej należy przyjąć taką teorię, która opisuje dane empiryczne w najkrótszym dwuczęściowym przekazie, gdzie pierwsza część tego przekazu dowodzi teorii a druga koduje dane empiryczne przy założeniu, że dowiedziona teoria jest prawdziwa. Podamy definicję *SMML*:

Rozważmy czwórkę H, X, f, p :

H - przestrzeń parametrów (zakłada się, że jest ona generowana przez σ - ciało jej podzbiorów).

X - wektor hipotez przyjmujących wartości ze zbioru $x_i : i \in N$.

f - dany rozkład a priori funkcji gęstości spełniający warunek $\int_H f(h)dh = 1, h \in H$.

p - znana warunkowa gęstość prawdopodobieństwa: $p : (X, H) \rightarrow [0, 1]$ i $p(x, h) = p(x|h)$, gdzie $\sum_i p(x_i|h) = 1$ dla wszystkich $h \in H$.

Estymatorem *SMML* nazywamy funkcję $m : X \rightarrow H$ taką, że $m(x) = h$. Funkcja m przy pewnych danych wartościach pochodzących ze zbioru $x_i : i \in N$ mówi nam którą teorię należy na ich podstawie przyjąć.

Niewątpliwie do słabości *AIC* należy to, że ma zastosowanie tylko do zagadnień, w których parametry mają rozkład ciągły.

3.5 Aplikacje statystyczne

Podamy przykłady zastosowania metod statystycznych w formalizmie kryterium Akaike:

(1) Hipoteza χ^2 .

Niech D_i - dane empiryczne, $y(x_i|\alpha)$ - model z parametrami α . Na przykład jeśli model jest funkcją liniową, to α oznacza tangens kąta nachylenia (współczynnik kierunkowy prostej) i punkt przecięcia prostej z osią współrzędnych. Wtedy hipoteza χ^2 ma postać:

$$\chi^2 = \sum_i \omega_i [D_i - y(x_i|\alpha)]^2, \omega_i = \frac{1}{\sigma_1^2} = const. \quad (16)$$

Parametry najbardziej zgodne z danymi empirycznymi (parametry najlepiej fitujące dane empiryczne) to takie, dla których χ^2 osiąga minimalną wartość. Minimalną wartość wyznacza się poprzez rozwiązanie równania: $\frac{\partial \chi^2}{\partial \alpha_i} = 0$.

Rozważmy szczególny przypadek: (a) błędy pomiarowe mają rozkład Gaussa, (b) model jest funkcją liniową. Wtedy χ^2 przyjmuje minimalną wartość dla $\nu = n - m$ stopni swobody (sumy kwadratów danych empirycznych), gdzie n - liczba danych empirycznych i m - liczba parametrów modelu (stałych występujących we wzorze określającym rozważaną funkcję liniową). Przez Q będziemy rozumieć miarę probabilistyczną (prawdopodobieństwo), czyli wartość liczbowa zgodności modelu z wynikami obserwacji (goodness of fit), gdy parametry tego modelu najlepiej fitują dane empiryczne tzn. χ^2 osiąga minimalną wartość. Q określone jest wzorem:

$$Q = 1 - P\left(\frac{\nu}{2}, \frac{\hat{\chi}^2}{2}\right), \quad (17)$$

gdzie $\hat{\chi}^2$ - estymator χ^2 , czyli funkcja podająca wartość χ^2 z pewnym przybliżeniem. Na przykład w przypadku gdy Q ma małą wartość, to: (a) model jest zły i należy go odrzucić, (b) błędy pomiarowe są większe niż powinny być, (c) błędy pomiarowe nie mają rozkładu Gaussa.

(2) ograniczenie na $\tilde{L} = \max\{L(\alpha, D)\}$, gdzie D - zbiór wartości D_i , α - parametry modelu. Gdy priory³⁶ (rozkłady a priori) parametrów modelu są niezależne, to taki model nazywamy akceptowalnym (acceptable) jeżeli zachodzi warunek:

$$-2 \ln \left[\frac{L(\alpha)}{\tilde{L}} \right] \leq c, \quad (18)$$

gdzie c - próg (threshold), $L(\alpha)$ - wartość funkcji wiarygodności dla pewnych wartości D w przestrzeni modelu (model space), dla której rozpatruje się całkę $\int_R P(\alpha|D)d\alpha$.

(3) estymacja parametrów modelu³⁷.

Dane uzyskane w wyniku przeprowadzenia pewnej obserwacji (eksperymentu) można zazwyczaj wytłumaczyć za pomocą modelu zawierającego określoną liczbę parametrów, mianowicie: niech $D = \{y_1, \dots, y_n\}$ będzie zbiorem wartości wektora obserwacji Y , m prawdziwą wartością mierzonej wielkości Y , a ε_i błędem pomiarowym³⁸. Wtedy zachodzi zależność $y_i = m + \varepsilon_i \Rightarrow \varepsilon_i = y_i - m, i = 1, \dots, n$. Rozkład $\varepsilon_i - g(y_i, \alpha)$, gdzie α to wektor parametrów rozkładu ε_i definiuje rozkład $y_i - m$. Gdy składowe α są znane, to model będzie opisany za pomocą m . W sytuacji, gdy rozważamy dwie wielkości mierzone Y i Z oraz zależność pomiędzy nimi ma postać $Z = f(Y, \lambda)$ mamy: $z_i = f(Y, \lambda) + \varepsilon_i$, a rozkład $z_i - f(Y, \lambda)$ jest wyznaczony przez rozkład ε_i .

Podamy przykładowe sposoby estymacji parametrów:

(3a) Wnioskowanie statystyczne bazujące na uśrednianiu modeli.

Jeśli chcemy estymować parametry modelu, a z analizy prawdopodobieństw wynikowych $p(M_i|D)$ wynika, że kilka modeli dobrze opisuje dane, to wnioski dotyczące parametru, czyli posterior dla tego parametru lepiej oprzeć na wszystkich modelach, niż na pojedynczym. Działanie takie nazywamy wnioskowaniem statystycznym bazującym na uśrednianiu modeli, mianowicie:

Niech κ będzie parametrem, który poddajemy analizie. Ponadto κ jest zdefiniowane dla każdego modelu z klasy rozważanych modeli M_i . Wtedy posterior κ - wnioski na temat κ określony jest wzorem:

$$p(\kappa|D) = \sum_{i=1}^K p(\kappa|D, M_i) \cdot p(M_i|D), \quad (19)$$

gdzie $p(\kappa|D, M_i) = \int p(\kappa|D, \alpha_i, M_i) \cdot p(\alpha_i|D, M_i)d\alpha_i$, gdzie α_i - parametry modeli. Na mocy wzoru (19) widać, że wnioskowanie oparte na jednym modelu ma sens, gdy posterior tego modelu $p(M_i|D) \approx 1$, a posteriory pozostałych rozważanych modeli są bliskie zeru (zaniedbywalne).

(3b) Metody Monte Carlo.

Podamy dwie przykładowe wypowiedzi metod Monte Carlo:

(I) Niech α_{true} - prawdziwy zbiór parametrów, którego nie znamy, D_1, D_2, \dots - dane empiryczne, $\alpha_1, \alpha_2, \dots, \alpha_0$ - parametry, które dofitowujemy z wyników obserwacji D_1, D_2, \dots

³⁶Ze względu na występowanie prioru w rozważaniach bayesowskich analiza bayesowska poddawana jest krytyce, mianowicie: (1) wniosek - rozkłady a posteriori bazujący (oparty) na tych rozważaniach zależy od wyboru prioru - wniosek nie jest obiektywny, (2) wniosek - rozkłady a posteriori zależy od przyjętej parametryzacji modelu, z tego względu że prior nie jest niezmienniczy (inwariantny) na zmianę parametryzacji (reparametryzację).

³⁷Sposób dokładnego wyznaczania wartości parametrów modelu został podany przez Fishera (1935) [42].

³⁸Przeważnie zakłada się, że ma on rozkład Gaussa o nieznannej wariancji.

Nasze zadanie polega na wywnioskowaniu postaci rozkładu $\alpha_i - \alpha_{true}$ bez znajomości rozkładu α_{true} . Ten cel osiągamy w następujący sposób: przyjmujemy, że na przykład α_0 jest parametrem prawdziwym. Wtedy rozkład hipotez $\alpha_0 - \alpha_i$, który potrafimy wyznaczyć utożsamiamy z rozkładem hipotez $\alpha_{true} - \alpha_i$. W ten sposób w ogólnym przypadku otrzymujemy pewien wielowymiarowy rozkład $\alpha_i^S - \alpha_0$.

(II) W przypadku gdy rozkład a posteriori $P(\alpha|x)$, gdzie α - parametry modelu, x - dane empiryczne jest skomplikowany generujemy z niego próbkę $\{\alpha_i\}, i = 1, \dots, N$. Wartość parametrów modelu może być oszacowana poprzez wybór z $\{\alpha_i\}$ takich parametrów, dla których rozważany posterior $P(\alpha|x)$ osiąga wartość maksymalną - oszacowanie mody rozkładu lub $\hat{\alpha} = \frac{1}{N} \sum_{i=1}^N \alpha_i$ - oszacowanie wartości oczekiwanej rozważanego posterioru $P(\alpha|x)$. Z metodami Monte Carlo wiąże się MCMC (Monte Carlo Markov Chains) [42, 43]. MCMC są metodami stosowanymi do wyznaczania posteriorów. W szczególności można wyznaczyć posterior dla parametrów $P(\alpha|x)$:

$$P(\alpha|x) = \frac{P(x|\alpha) \cdot P(\alpha)}{\int P(x|\alpha') \cdot P(\alpha') d\alpha'} \quad (20)$$

4 AIC zakres ważności

W sekcji pierwszej – bardzo ogólnej – podaliśmy różne konotacje prostoty, którą umieszczano w wielu, często nieprzystających kontekstach i próbowano oszacować różnymi *wskaźnikami*. Wymieńmy te najczęściej używane: efektywność, ilość założeń bazowych, jakość przesłanek, prostota konceptualna, koherencja, dogodność, ekonomiczność. Często badane przez nas kryterium było po prostu kojarzone z estetyką, elegancją teorii, modelu czy hipotezy. Nasza praca stanowi próbę odejścia od tego typu intuicyjnych interpretacji prostoty i podejmuje kwestię możliwości podejść ilościowych.

Pierwszym celem tej części pracy, poświęconej już tylko kryterium Akaike, jest podanie zastrzeżeń w odniesieniu do tego, czy prostotę można wyjaśniać (czy wręcz *liczyć*) w kategoriach ilości informacji. Czy w ogóle można podać miarę liczbową prostoty? Najważniejszym jednak zadaniem będzie podanie tych obszarów zastosowań *AIC*, w których to kryterium pokazuje swoją największą moc i skuteczność. Chcemy mocno podkreślić, że *AIC* nie jest kryterium uniwersalnym i dającym się wszędzie stosować; podobnie jak prostota nie da się wyodrębnić jako wartość ogólna i niezależna. Raczej jest rodzajem relacji przynajmniej dwuargumentowej. Można chyba powiedzieć o teorii, że jest prosta, ale tylko intuicyjnie; ściśle relację prostoty wyrazimy w zdaniu: „teoria T_1 jest prostsza niż teoria T_2 ”.

Podstawowym nieporozumieniem, dotyczącym zakresu stosowalności *AIC*, jest to, że używa się go jako kryterium, które stwierdza, że dana hipoteza jest prawdziwa lub fałszywa. *AIC* nie wydaje generalnych i izolowanych ocen jednego modelu. *AIC* nie jest narzędziem weryfikacyjnym w sensie ścisłym; ono mówi jedynie o bliskości (*odległości* mierzonej przez metrykę K-L) do prawdy. Zresztą samo mówienie o prawdzie czy prawdziwości hipotezy domaga się precyzji. Na przykład realista pyta rzeczywiście o teorie prawdziwe; empirysta raczej o te, które są najbardziej adekwatne empirycznie. Należy też pamiętać, jak ryzykowne i najczęściej chybione jest wprowadzanie bezwzględnej liczbowej miary stopnia prostoty teorii czy prawa nauki, jak i równoważenie kryterium prostoty z zawartością informacyjną teorii czy prawa [14, s. 230-233].

Tradycyjnie mówi się, że działanie badawcze osiąga, przede wszystkim, dwa cele: znajdowanie tych teorii, które najlepiej aproksymują badaną rzeczywistość [44] i konstruowanie hipotez dostarczających najlepszych przewidywań. Takie ujęcie zbiega się z ideą selekcji

modeli, która także nakierowana jest na wybór modelu najbliższego prawdzie i tego, który charakteryzuje się najlepszą mocą predykcji. W tym kontekście konkurują ze sobą dwa kryteria selekcji modeli: prostota z jednej strony i dokładność fitowania danych z drugiej. Ten efekt można ująć jeszcze inaczej w postaci problemu: czy celem modelu jest przede wszystkim predyktywne wykraczanie poza uzyskane dane, czy raczej interpolacja między danymi. W tym sensie prostota zbiega się z informatywnością teorii czy modelu. Sam model dysponuje pewną informacją, ale dzięki nowym danym empirycznym (informacji dodatkowej – z zewnątrz) jest w stanie odpowiedzieć na interesujące badacza pytania. Występuje tu rodzaj zależności odwrotnej: im bardziej informatywny model tym mniej informatywnie teorie.

Konkurujące ze sobą hipotezy najczęściej ocenia się zestawiając dwa kryteria:

- *AIC*, które maksymalizuje wartość $2 \ln \tilde{L} - 2k$
- *BIC*, które maksymalizuje wartość $2 \ln \tilde{L} - k \ln N$.

W tym zestawieniu widoczne stają się różne cele, które osiąga się stosując oba kryteria: *AIC* maksymalizuje dokładność predykcji; *BIC* szacuje maksymalne zbliżenie modelu teoretycznego do prawdziwego modelu. Naturalnie można zastanawiać się nad problemem, dlaczego dany model nie dokładnie fituje dane. Albo model jest po prostu zły, albo sam proces zbierania danych jest obciążony błędem. I nie jest to bynajmniej banalna kwestia. W odpowiednim jej ujęciu i rozwiązaniu leży prawdopodobnie największa siła kryterium Akaike: uchwycenie pewnego balansu między zbyt dokładnym odzwierciedleniem danych empirycznych przez model (na przykład krzywą dopasowaną do punktów pomiarowych) – tzw. *overfitting problem* z jednej strony, a koniecznym do założenia błędem pomiaru (pewnym nadmiarem informacji) – *noise in the data*. W tej niezwykle prostej intuicji kryje się istota skuteczności sformalizowanej procedury badawczej, którą opisaliśmy w sekcji drugiej. M. Forster i E. Sober ujęli ją w postaci swoistego paradoksu, którego nie sposób chyba wyrazić bardziej dosłownie [45, s. 5-9]:

Even though a hypothesis with more adjustable parameters would fit the data better, scientists seem to be willing to sacrifice goodness-of-fit if there is a compensating gain in simplicity.(...)

Since we assume that observation is subject to error, it is overwhelmingly probable that the data we obtain will not fall exactly on that true curve.(...) Since the data points do not fall exactly on the true curve, such a best-fitting curve will be *false*. If we think of the true curve as the 'signal' and the deviation from the true curve generated by errors of observation as 'noise', then fitting the data perfectly involves confusing the noise with the signal. It is overwhelmingly probable that any curve that fits the data perfectly is false.

Akaike [46] discovered a way of estimating the size of the overfitting factor.

Na podstawie lektury wyników prac Sobera i Forstera [47, 45, 48] można sformułować, mimo posiadanych różnych słabości, w miarę bezpieczny wstęp do *podręcznika użytkownika* kryterium Akaike:

1. Założenie 1. By dokładność przewidywań była dokładnością matematyczną, definiuje się odległość między fitowanym modelem a prawdziwym przy pomocy metryki K-L.
2. Założenie 2. Nowe dane tkwią w tej samej rzeczywistości co stare:
 - funkcja, która wiąże zmienne zależne i niezależne jest taka sama;

- stare i nowe dane są wzięte z tego samego rozkładu.
3. Założenie 3. Powtarzane szacunki danego parametru istotnego w modelu generują rozkład normalny.
 4. *AIC* jest kryterium względnym, tzn. odnosi się tylko do oceny określonego aspektu działania konkurujących ze sobą modeli.
 5. Celem *AIC* jest przede wszystkim pokazanie, który model nadaje się (lepiej niż inny) do wysuwania predykcji. Celem nie jest otrzymanie najlepszego obrazu fitowanych danych, ale przewidywanie nowych, a dokonuje się to kosztem prawdziwości modelu. Można to ująć jeszcze mocniej: celem nie jest sprawdzenie jak prawdziwa jest hipoteza. Dlatego nieporozumieniem jest zestawianie *AIC* i *BIC* jako kryteriów konkurujących, ponieważ każde z nich ocenia co innego.
 6. W dokonywaniu selekcji przy pomocy *AIC* podstawową kwestią jest wyodrębnienie parametrów istotnych. Można oszacować wartość tych parametrów przez znalezienie takiej wartości parametru, która maksymalizuje prawdopodobieństwo danych.
 7. Dokładność predykcji wiąże się ściśle z ilością otrzymanych danych [por. Sekcja: Krytyka *AIC*].

Przy ocenie zastosowania metod statystycznych w badaniu naukowym, zawsze napotykamy na zarzuty dotyczące już samego pojęcia prawdopodobieństwa, problemów w probabilistyczną ekstrapolacją własności poszczególnego indywiduum, koniecznością przyjęcia zależności statystycznej między wynikami powtarzanych obserwacji. Zauważmy jednak, jak mocno kryterium Akaike uzależnia własną koncepcję prostoty od założeń w rzeczy samej empirycznych (na przykład jednorodność natury, założenie o tym, że rozkład jest normalny). Jeśli uzna się dodatkowo, że *AIC* nie rości sobie pretensji do oceny prawdziwości pojedynczej hipotezy, to przy wielu ograniczeniach jest niezwykle przydatnym narzędziem w pracy badawczej.

5 *AIC w działaniu* [49]

Poniższy przykład ma na celu pokazanie, że kryterium Akaike nie działa jedynie na zasadzie odrzucania bardziej skomplikowanych modeli na rzecz prostszych. Dość sugestywnie odpowie na pytanie, dlaczego teorie prostsze intuicyjnie, nie zawsze są bliższe prawdy. Niezależnie od sposobu jego wyprowadzenia, można intuicyjnie spojrzeć na końcową formułę - zawiera ona w sobie informację o jakości dopasowania danych do testowanych modeli jak i o skomplikowaniu, które mierzone jest w dosyć prosty sposób - ilością dopasowywanych parametrów. Dla wielu pozostaje jednak niejasnym, czy nie zachodzi tu przekłamanie na rzecz zbytnej prostoty lub idealnej zgodności z punktami pomiarowymi. Jest to niestety kwestia dosyć subiektywna, gdyż *AIC* nie jest jedynym kryterium na rynku, ani też „prostota” czy „elegancja” nie przekładają się jednoznacznie na matematyczne formuły. Można jednak przeprowadzić anachroniczny eksperyment, aby przetestować sam test, i użyć *AIC* do problemów dawno już rozwiązanych, tego typu, w którym bardziej skomplikowana teoria zastąpiła bez cienia wątpliwości poprzednią. Mówiąc inaczej, można zadać pytanie, czy gdyby dawni uczeni dysponowali tym kryterium, zatrzymaliby się ze swoimi teoriami na prymitywnym etapie, tylko dlatego, że prostota górowałaby nad jakością dopasowania?

Idealnym przykładem jest tutaj kształt orbity ziemskiej – z grubsza rzecz biorąc jest to okrąg, więc czy potrzebne jest komplikowanie sobie życia i dokładanie do niego mimośrodu i orientacji elipsy w płaszczyźnie ekliptyki? Wyobraźmy sobie, że doświadczenie polega na codziennym pomiarze odległości Ziemia - Słońce. Oczywiście to po części eksperyment myślowy, więc będziemy musieli zaniedbać pewne praktyczne kwestie. W szczególności wyznaczenie jednostki astronomicznej w kilometrach wymaga wykorzystania np. takich zjawisk jak przejście Wenus na tle tarczy słonecznej, ale można uznać, że interesuje nas tylko względna zmiana promienia wodzącego licząc od danego dnia. Wtedy, zakładając, że akceptujemy stałość prędkości polowej na orbicie, możemy wyznaczyć stosunki odległości ze stosunków prędkości kątovej Słońca na niebie. Dodatkowo, należy pamiętać, że jeżeli prawdą jest hipoteza o eliptyczności orbity, jeden dzień nie oznacza zmiany o taki sam kąt promienia wodzącego. Możemy uznać, że błąd ten jest niewielki - jak zobaczymy dalej, całkowity błąd i tak będzie oszacowany arbitralnie - albo też, dokonywać pomiarów w różnych porach, kierując się jako zmienną niezależną właśnie położeniem kątowym Słońca na sferze niebieskiej. W efekcie dostaniemy 365 (lub 366) punktów pomiarowych, z wartościami odległości w umownych jednostkach. Jeżeli orbita jest kołowa, to bezbłędny pomiar powinien dać 365 „jedynek” - możemy jedynie sprawdzić prawdopodobieństwo *a posteriori* takiej sytuacji, zauważając, że nie ma tu właściwie żadnego dopasowanego parametru - średnia odległość powinna po prostu wynieść 1. Dla elipsy z kolei - jeżeli uda się ją wyłowić wśród błędów pomiarowych - musimy wyznaczyć mimośród oraz np. moment przejścia Ziemi przez peryhelium - to dwa parametry. Tu także potrzebne jest prawdopodobieństwo *a posteriori*, ale dla każdego punktu trzeba wziąć pod uwagę, że odległość zależy istotnie od kąta. Na koniec można dodatkowo założyć, że w praktyce każdego dnia wykonamy wiele pomiarów i operować będziemy średnimi wraz z ich błędami.

Jeżeli idzie o błędy, to nasz eksperyment myślowy polega również na zbadaniu, jak wielkie mogą one być, aby sens miał w ogóle taki pomiar i stosowanie *AIC*. Załączony rysunek pokazuje przykładowy rozkład (zasymulowanych) punktów pomiarowych z gausowskimi błędami, dopasowany okrąg, oraz elipsę. Jest to dosyć ekstremalny przypadek, gdyż błędy pojedynczego pomiaru mają tutaj wartość odchylenia standardowego równą 0.5, czyli połowie badanej odległości! Można chyba spokojnie założyć, że badania orbity podejmuje się, gdy sprzęt pozwala na „nieco” lepszą dokładność.

Zaskakujący (dla niektórych) wniosek otrzymujemy patrząc na różnice *AIC* dla porównywanych modeli - elipsa, pomimo skomplikowania o dwa dodatkowe parametry, jest faworyzowana przez to kryterium. Średnio rzecz biorąc (tzn. uśredniając po 100 zestawach symulacji) *AIC* jest wyższe o 9 dla okręgu. Już dziesięciokrotnie mniejszy błąd owocuje różnicą rzędu 1000. Liczby mówią same za siebie.

6 Zakończenie

Zbadaliśmy w pracy funkcje i cele kryterium prostoty sformalizowanego w terminach statystyki klasycznej (niebayesowskiej) – kryterium Akaike. Jego atrakcyjność w naukowej praktyce badawczej polega głównie na tym, że mówi ono nam, że w dysponowanym modelu uwzględnienie kolejnego parametru zjawiska daje nam coraz mniejszy wkład do wyjaśnienia zjawiska. Użyteczność tego kryterium polega na tym, że można je stosować w nauce wszędzie tam, gdzie w wyjaśnianiu zjawisk posługujemy się modelami, a parametry modelu odpowiadają za jego zmienność. Podkreślaliśmy w tekście, że współczesna naukowa praktyka badawcza w ogólności koncentruje swoją uwagę bardziej na badaniu modeli niż budowaniu teorii.

Uczeni koncentrują swoje badania na modelach, a teorie służą im w następnym etapie ich

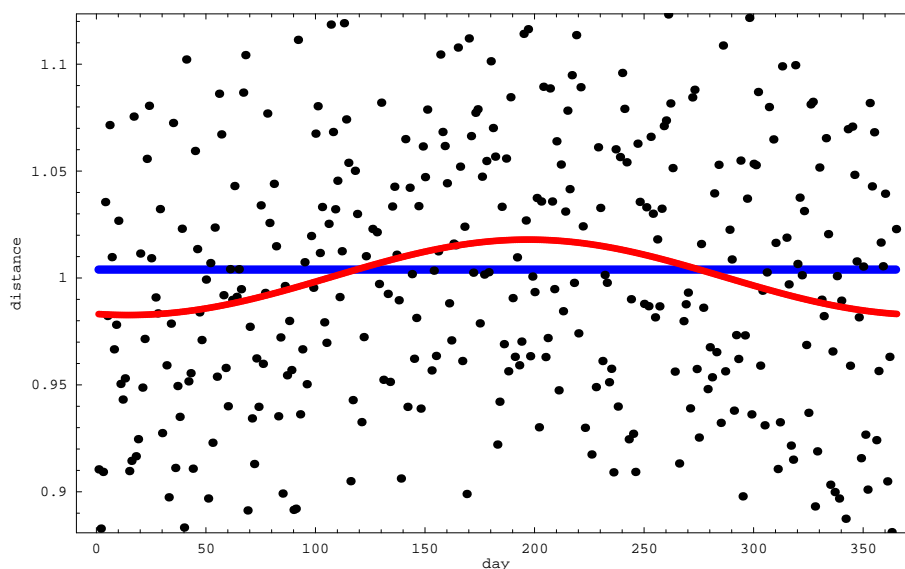


Figure 1: Rysunek przedstawia punkty pomiarowe na każdy dzień (czyli średnie z 50 dziennych pomiarów), wraz z dopasowaną orbitą kołową (niebieska) i eliptyczną (czerwona).

konstrukcji. W momencie, kiedy same modele zostały już sformułowane, badania przesuwają się w kierunku określenia charakterystyki i wartości samego modelu. W tym momencie stają się ważne kryteria oceny tych modeli, które są elementami skrzynki narzędziowej w sensie N. Cartwright.

W pracy staraliśmy się pokazać ścisły związek kryterium prostoty z koncepcją brzytwy Ockhama. W tym kontekście zwróciliśmy uwagę na ważne zastrzeżenie P. Kawalca, by kryterium prostoty nie traktować jako ostatecznie rozstrzegającego w ocenie testowanych hipotez. Oczywiście powyższe zastrzeżenie posiada bezpośrednie zastosowanie do kryterium Akaike.

W definicji kryterium Akaike występuje człon, które „karze” model za posiadanie dodatkowych (nadmiarowych) parametrów (przypomnijmy folklorystyczne powiedzenie G. Gamowa, że mając pięć parametrów, możemy dofitować nawet słonia w butelce). To mogłoby sugerować, że AIC działa po prostu w kierunku wyeliminowania bardziej skomplikowanego modelu – o większej liczbie parametrów – niejako od samego początku. Nic bardziej błędnego; przykład z sekcji 5 opracowany i udostępniony nam przez T. Stachowiaka [cytowanie j.w.] pokazuje, że AIC faworyzuje eliptyczne orbity Keplerowskie w układzie grawitacyjnym Ziemia – Słońce, a nie orbity kołowe.

Niewątpliwie zaletą badanego przez nas kryterium, podkreślaną przez wielu autorów, jest jego niebayesowski charakter, co sprawia, że AIC nie jest uwikłane w założenia o charakterze filozoficznym [por. [50]]. Z drugiej strony zwróćmy uwagę, że jest ono ściśle powiązane z pojęciem prostoty, brzytwą Ockhama, etc. W pracy zwróciliśmy uwagę na związek prostoty w rozumieniu K. Poppera z jej operacjonalizacją w postaci AIC , podkreślając, że *de facto* AIC zawiera i rozwija najważniejsze intuicje Poppera.

We współczesnej praktyce badawczej sytuacji problemowe wymagają stosowania zabiegu selekcji modeli, będącymi autonomicznymi narzędziami tej praktyki. Z jednej strony posiadamy dane empiryczne (obarczone błędami), a z drugiej model. W ten sposób dysponujemy opisem tego samego zjawiska przy pomocy wielu modeli (*multiple explanation problem*). Model lepszy wg. kryterium Akaike, to ten, który uwzględnia istotne parametry

modelu, dające się wyłowić przy dostępnej jakości danych. Czyli zawsze nasza selekcja będzie zrelatywizowana do danych empirycznych, którymi dysponujemy. Lepsza ich jakość może się okazać skuteczna w wyłapaniu nowego parametru, jako istotnego – mierzącego nowy stopień swobody układu.

Spis literatury

- [1] Lieu R 2007 (*Preprint* 0705.2462)
- [2] Cartwright N, Shomar T and Suárez M 1996 *Theories and Models in Scientific Processes* ed Herfel W E e a (Amsterdam: Editions Rodopi) p 137–149
- [3] Bennett C L *et al.* 2003 *Astrophys. J. Suppl.* **148** 1 (*Preprint* astro-ph/0302207)
- [4] Hinshaw G *et al.* 2003 *Astrophys. J. Suppl.* **148** 135 (*Preprint* astro-ph/0302217)
- [5] Spergel D N *et al.* (WMAP) 2003 *Astrophys. J. Suppl.* **148** 175 (*Preprint* astro-ph/0302209)
- [6] Tegmark M *et al.* (SDSS) 2004 *Phys. Rev.* **D69** 103501 (*Preprint* astro-ph/0310723)
- [7] Riess A G *et al.* 2007 *Astrophys. J.* **659** 98–121 (*Preprint* astro-ph/0611572)
- [8] Padmanabhan T and Choudhury T R 2003 *Mon. Not. Roy. Astron. Soc.* **344** 823–834 (*Preprint* astro-ph/0212573)
- [9] Padmanabhan T 2006 (*Preprint* astro-ph/0603114)
- [10] Rodríguez-Fernández J L 1999 *Endeavour* **23** 3 121–125
- [11] Strawiński W 1991 *Prostota, redukcja, jedność nauki* (Warszawa: Wydawnictwo FEA)
- [12] Rissanen J 1978 *Automatica* **14** 465–471
- [13] Popper K R 2002 *Logika odkrycia naukowego* (Warszawa: PWN)
- [14] Kawalec P 2006 *Przyczyna i wyjaśnianie, Studium z filozofii i metodologii nauk* (Lublin: Wydawnictwo KUL)
- [15] Cartwright N 1997 *Phil. Sci.* **64** 292–303
- [16] Morrison M 2006 *Poznań Studies in the Philosophy of the Sciences and the Humanities* **86** 145–172
- [17] Grobler A 2006 *Metodologia nauk* (Kraków: Aureus, ZNAK)
- [18] Fisz M 1954 *Rachunek prawdopodobieństwa i statystyka matematyczna* (Warszawa: PWN)
- [19] Kryszicki W, Bartos J, Dyczka W, Królikowski K and Wasilewski W 1997 *Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach* (Warszawa: PWN)
- [20] Jakubowski J and Sztencel R 2001 *Wstęp do teorii prawdopodobieństwa* (Warszawa: SCRIPT)
- [21] Linder E V and Miquel R 2007 (*Preprint* astro-ph/0702542)

- [22] Osiewalski J 2001 *Ekonometria bayesowska w zastosowaniach* (Kraków: Wydawnictwo Akademii Ekonomicznej)
- [23] Grabski F and Jaźwiński J 2001 *Metody bayesowskie w niezawodności i diagnostyce* (Warszawa: WKL)
- [24] Akaike H 1974 *IEEE Trans. Auto. Control* **19** 716–723
- [25] Akaike H 1981 *J. Econometr.* **16** 3–14
- [26] Szydłowski M, Kurek A and Krawiec A 2006 *Phys. Lett.* **B642** 171–178 (*Preprint astro-ph/0604327*)
- [27] Szydłowski M and Kurek A 2006 *AIP Conf. Proc.* **861** 1031–1036 (*Preprint astro-ph/0603538*)
- [28] Liddle A R 2007 *Mon. Not. Roy. Astron. Soc. Lett.* **377** L74–L78 (*Preprint astro-ph/0701113*)
- [29] Takeuchi T 2000 *Astrophysics and Space Science* **271** 213–226
- [30] Burnham K P and Anderson D R 2004 *Soc. Methodes and Research* **33** 2 261
- [31] Burnham K P and Anderson D R 2002 *Model Selection and Multimodel Inference: A Practical Information – Theoretical Approach* 2nd ed (New York: Springer-Verlag)
- [32] Biesiada M 2007 *JCAP* **0702** 003 (*Preprint astro-ph/0701721*)
- [33] Szydłowski M and Kurek A 2008 (*Preprint 0801.0638*)
- [34] D’Agostini G 1999 *CERN Yellow Report* **99** 03
- [35] Kurek A and Szydłowski M 2007 (*Preprint arXiv:0710.2125 [astro-ph]*)
- [36] Dowe D L Gardner S and Oppy G 2007 *Brit. J. Phil. Sci.* **58** 709–754
- [37] Szydłowski M 2005 *Roczniki Filozoficzne* **51** 2 217–235
- [38] Kiesepä I A 1997 *Brit. J. Phil. Sci.* **48** 21–48
- [39] Sugiura N 1978 *Commun. Stat. A-Theor.* **A7** 13
- [40] Neyman J and Scott E 1948 *Econometrika* **16** 1–32
- [41] Wallace C S 2005 *Statistical and Inductive Inference by Minimum Message Length* (Berlin: Springer)
- [42] Verde L 2007 (*Preprint astro-ph/0712.3028*)
- [43] D’Agostini G 2003 *Reports on Progress in Physics* **66** 9 1383–1419(37)
- [44] Smith P 1998 *Brit. J. Phil. Sci.* **49** 253–277
- [45] Forster M and Sober E 1994 *Brit. J. Phil. Sci.* **45** 1–35
- [46] Akaike H 1973 *2nd International Symposium on Information Theory* ed Petrov B N and Csaki F (Budapest: Akademiai Kiado)

- [47] Sober E 2002 *Phil. Sci.* **69** 112–123
- [48] Forster M 2000 *J. Math. Psy.* **44** 205–231
- [49] Stachowiak T *Praca w przygotowaniu*
- [50] Kawalec P 2003 *Roczniki Filozoficzne* **51 1** 113–142