

Metoda największej wiarygodności i informacja Fisher'a w fizyce
i ekonofizyce

Jacek Syska

Uniwersytet Śląski, Instytut Fizyki

Skrypt dla studentów ekonofizyki sfinansowany w ramach projektu:
“Uniwersytet partnerem gospodarki opartej na wiedzy”.
(wersja pierwsza)

Spis treści

0.1	Wstęp	4
1	Metoda największej wiarygodności	8
1.1	Wnioskowanie w MNW	10
1.1.1	Weryfikacja hipotez z wykorzystaniem ilorazu wiarygodności	13
1.2	MNW w analizie regresji	14
1.2.1	Dewiancja jako miara dobroci dopasowania. Przykład rozkładu Poissona.	15
1.2.2	Analiza regresji Poissona.	17
1.2.2.1	Podobieństwo dewiancji do SKR analizy częstościowej	21
1.3	Zasada niezmienniczości ilorazu funkcji wiarygodności	23
2	Entropia względna i informacja Fishera	24
2.1	Obserwowana i oczekiwana informacja Fishera	24
2.1.1	Wartość oczekiwana i wariancja funkcji wynikowej	26
2.2	Wstęp do geometrii różniczkowej na przestrzeni statystycznej i α -koneksja	28
2.3	Dolne ograniczenie Rao-Cramera	29
2.3.1	Dowód TRC (wersja dowodu dla przypadku skalarne)	30
2.3.2	Przypadek wieloparametrowy DORC	31
2.3.2.1	Przykład wektorowego DORC	32
2.4	Entropia informacyjna	35
2.4.1	Entropia Shannona i jej własności	35
2.4.1.1	Interpretacja entropii Shannona	35
2.4.1.2	Przypadek ciągłego rozkładu prawdopodobieństwa	36
2.4.2	Entropia względna	37
2.5	Geometria przestrzeni rozkładów prawdopodobieństwa i metryka Fishera-Rao	39
2.6	Informacja Fishera	43
2.6.1	Informacja Fishera jako entropia	43
2.7	Pojęcie kanału informacyjnego	47
2.7.1	Pojemność informacyjna dla zmiennej losowej położenia	48
2.7.2	Podstawowe założenie fizyczne podejścia Friedena-Soffera	51
3	Zasady informacyjne	54
3.1	Estymacja w statystyce klasycznej a estymacja fizyczna. Postawienie problemu	54
3.1.1	Strukturalna zasada informacyjna. Metoda EFI	55
3.1.1.1	I oraz Q dla parami niezależnych zmiennych położeniowych próby	57
3.2	Przepływ informacji	58

3.3	Kinematyczna postać informacji Fishera	61
3.3.0.2	Postać kinematyczna pojemności zapisana w prawdopodobieństwie	64
3.4	Równania master	65
3.5	Podsumowanie rozważań	65
4	Kryteria informacyjne w teorii pola	67
4.1	Informacja Fishera i klasyfikacja modeli	67
4.1.1	Podział modeli ze względu na N oraz kategorie I	68
4.1.1.1	Dowód podziału na dwie kategorie I	68
4.1.2	Podział modeli ze skończonym I na podklasy z różnym N	69
4.1.3	Konkluzje i konsekwencje podziału modeli na kategorie I	69
4.2	Równania zasad informacyjnych dla równań modeli fizycznych	70
4.3	TFI oraz specyficzne formy Q w teorii pola	73
4.3.1	Skalary Kleina-Gordona	75
4.3.2	TFI równania Kleina-Gordona dla pól rangi N	76
4.3.2.1	Wstępna foliacja \mathcal{S} oraz pochodna kowariantna	76
4.3.2.2	Równanie ruchu Diraca dla pola swobodnego rangi N	77
4.3.3	Końcowe uwagi o wkładzie Q w zasadę strukturalną	81
5	Przykłady z fizyki statystycznej i ekonofizyki oraz efekt EPR-Bohm'a	82
5.1	Wyprowadzenie klasycznej fizyki statystycznej z informacji Fishera	82
5.1.1	Fizyczne sformułowanie zagadnienia	82
5.1.2	Informacja kinetyczna i strukturalna oraz sformułowanie zasad informacyjnych	83
5.1.3	Rozkład Boltzmana dla energii	85
5.1.4	Model Aoki-Yoshikawy dla produktywności branż	91
5.1.5	Rozkład Maxwella-Boltzmana dla prędkości	96
5.1.6	Informacja Fishera jako ograniczenie dla wzrostu entropii	101
5.1.6.1	Wynik dla strumienia cząstek ze spinem $1/2$	103
5.2	Zastosowanie wprowadzonego formalizmu do analizy paradoksu EPR	105
5.2.1	Warunki brzegowe	106
5.2.2	Pojemność informacyjna dla zagadnienia EPR-Bohm'a	108
5.2.3	Informacja strukturalna. Amplituda prawdopodobieństwa	111
5.2.4	Niepewność wyznaczenia kąta	118
5.2.4.1	Wpływ zaszumienia pomiaru	119
5.2.5	Informacja Q jako miara splątania	119
6	Zakończenie	121
7	Dodatki	123
7.1	Dodatek: Zasada nieoznaczoności Heisenberga	123
7.2	Dodatek: Równanie Schrödingera	126
7.3	Dodatek: Rezultaty EFI dla elektrodynamiki Maxwella oraz teorii grawitacji	128
7.3.1	Dodatek: Pole cechowania Maxwella	128
7.3.2	Dodatek: Metoda EFI dla teorii grawitacji	130

7.4	Dodatek: Informacyjny odpowiednik drugiej zasady termodynamiki: Twierdzenie <i>I</i> . Temperatura Fishera	132
-----	--	-----

0.1 Wstęp

Tematem skryptu jest metoda największej wiarygodności (MNW) oraz informacja Fishera (IF) w fizyce i statystyce. Temat dotyczy bardzo aktualnego sposobu konstrukcji modeli fizycznych wywodzącego się ze statystycznego opisu zjawisk, którego formalizm pozwala na opis całego spektrum różnych teorii pola, klasycznych i kwantowych. Kluczowe w tym podejściu pojęcie (oczekiwanej) IF wprowadził Fisher na gruncie jego rozważań związanych z oszacowaniem parametrów modeli, podlegających badaniu statystycznemu w ramach ogólnej metody ekstremalnej wartości funkcji wiarygodności L . IF opisuje lokalne własności funkcji wiarygodności $L \equiv P(\mathbf{y}|\Theta)$, która formalnie jest łącznym prawdopodobieństwem (lub łączną gęstością prawdopodobieństwa) danych \mathbf{y} , lecz jest rozumiana jako funkcja zbioru parametrów Θ , który z kolei tworzy współrzędne w przestrzeni statystycznej. Analiza statystyczna modeli fizycznych idzie jak dotąd dwoma nurtami.

Pierwszy z nich, geometryczny, próbuje opisać metodę statystyczną wprowadzoną przez Fishera w latach 20 poprzedniego wieku [1], twórcy podstaw techniki statystycznej otrzymywania dobrych estymatorów MNW, w jak się okazało naturalnym dla niej środowisku geometrii różniczkowej. Rozwijając MNW, już w 1945 roku C.R. Rao [2] zauważył, że macierz informacji Fishera określa metrykę Riemanna i badał strukturę modeli statystycznych z punktu widzenia geometrii Riemannowskiej. Z kolei B. Efron [3] badając jedno-parametrowe modele i analizując ich asymptotyczne własności dla procedur estymacyjnych, wprowadził i odkrył użyteczność pojęcia statystycznej krzywizny. A.P. Dawid [4] wprowadził pojęcie koneksji na przestrzeni wszystkich dodatnio określonych rozkładów prawdopodobieństwa, pokazując, że ze względu na tę koneksję statystyczna krzywizna jest krzywizną zewnętrzną (embedding curvature). Jednak problemem Dawida był nieskończony wymiar przestrzeni rozkładów. W roku 1980 S. Amari [5] opublikował systematyczne ujęcie teorii Dawida dla modeli skończenie wymiarowych i podał spójne określenie α -koneksji (wprowadzonej wcześniej poza kontekstem statystycznej estymacji przez N.N. Chentsova). W 1982 S. Amari wraz z H. Nagaoka [12] wykazali dualność płaskich przestrzeni modeli eksponencjalnych z e -koneksją i modeli mieszanych z m -koneksją.

Procedury estymacyjne statystycznego opisu mechaniki kwantowej (falowej) poszły dwoma drogami. Pierwsza z nich związana jest z naturalnym dla mechaniki kwantowej formalizmem macierzy gęstości, druga z konstrukcją zasad informacyjnych (entropijnych). W przypadku formalizmu macierzy gęstości, ich zbiór $S = \bigcup_{r=1}^k S_r$ ($S_r \cup S_i = \emptyset$, $i \neq r$) dla przypadku skończenie wymiarowych przestrzeni Hilberta \mathcal{H} , tworzy zbiór wypukły. Dla stanów czystych, podzbiór S_1 tego zbioru tworzą punkty ekstremalne, a przestrzeń stanów czystych związana z nim może być utożsamiona z zespoloną przestrzenią rzutową CP^{k-1} , ($k = \dim \mathcal{H}$). Na przestrzeni tej można wprowadzić (z dokładnością do stałej) metrykę Riemannowską nazywaną metryką Fubiniiego-Study, która jest kwantową wersją metryki Fishera. Statystyczną estymacją w modelach dla stanów czystych zajmowali się między innymi Fujiwara, Nagaoka i Matsumoto [6]. Natomiast w przypadku podzbioru S_k zbioru S dualna struktura z metryką może być traktowana jako kwantowy analog metryki Fishera z $\pm\alpha$ -koneksją.

Drugim nurtem, który wyłonił się w ostatnich kilkunastu latach i którym szedł rozwój zastosowań MNW oraz pojęcia (oczekiwanej) IF w fizyce jest formalizm ekstremalnej fizycznej informacji (EFI) opracowany przez Friedena i jego współpracowników, w szczególności Soffera [7]. Konstrukcję modeli fizycznych z wykorzystaniem informacji Fishera zapoczątkował Frieden, podając metodę wyprowadzenia z informacji Fishera członu kinetycznego modeli fizycznych. Na-

stepnie zapostulował wprowadzenie dwóch zasad informacyjnych służących do ustalenia związku pomiędzy informacją Fishera I oraz informacją strukturalną Q , tzn. poprzez zapostulowaną nową strukturalną zasadę informacyjną skonstruował on człony strukturalne rozważanych przez siebie modeli. W odróżnieniu od Friedena stosujemy jednak inne [9], bardziej fizyczne a mniej informacyjne, podejście do konstrukcji podstawowych zasad informacyjnych, posługując się pojęciem całkowitej fizycznej informacji K a nie wprowadzonym przez Friedena pojęciem zmiany fizycznej informacji. Różnica ta chociaż nie powoduje zasadniczo rachunkowych zmian w sposobie wyprowadzenia równań ruchu bądź równań stanu układu dla rozważanych do tej pory problemów, jednak zmieniając pojęcie informacji fizycznej oraz jej rozkładu na kinetyczne i strukturalne stopnie swobody, idzie w linii prowadzonych ostatnio badań nad konstrukcją zasady ekwipartycji dla entropii. To inne niż Friedenowskie podejście do pojęcia fizycznej informacji powoduje również zmiany w pojmowaniu istoty przekazu informacji (entropii) w procesie pomiaru przy jej przekazie od strukturalnych do kinetycznych stopni swobody. Pomimo różnic samą metodę będziemy dalej nazywać podejściem Friedenowskim. Gdyby pominąć chwilowo proces pomiaru, w metodzie Friedena próbkowanie przestrzeni jest wykonywane przez układ, który poprzez właściwe dla niego pole (i związane z nim amplitudy) o randze N będącej wielkością próby, próbkuje jego kinetycznymi (Fisherowskimi) stopniami swobody dostępną mu przestrzeń konfiguracyjną. Następnie, ponieważ IF jest (z dokładnością do znaku) infinitezymalnym typem entropii Kulbacka-Leiblera, to zauważając, że entropia Kulbacka-Leiblera jest wykorzystywana w statystyce do przeprowadzania testów wyboru modeli, pojawia się przypuszczenie, że IF może poprzez narzucenie na nią odpowiednich dodatkowych ograniczeń, zapostulowanych w postaci wspomnianych dwóch zasad informacyjnych, skalarnej (wariacyjnej) oraz wewnętrznej (strukturalnej), doprowadzić do wyprowadzenia równań ruchu bądź równań stanu układów fizycznych, najlepszych z punktu widzenia owych informacyjnych zasad. Na tym zasadza się Friedenowska idea estymacji fizycznych modeli.

Zasady informacyjne mają uniwersalną postać, jednak ich konkretne realizacje zależą od fizyki rozważanego zagadnienia. Pierwsza z zasad informacyjnych (wariacyjna) prowadzi do otrzymania właściwej relacji dyspersyjnej dla układu. Druga, strukturalna opisuje wewnętrzne charakterystyki układu związane np. z jego spinem. Ciekawą sprawą jest, że wiele rachunków można wykonać dla przypadku, dla którego całkowita informacja układu dzieli się na dwie równe (lub z czynnikiem $1/2$) części, tzn. informację Fishera oraz informację strukturalną, mając swoją całkowitą wartość równą zero. Frieden podał informacyjne wyprowadzenie równania Kleina-Gordona dla ogólnego modelu pola z rangą N , z szczególnym uwzględnieniem przypadku pola skalarnego z $N=2$. Dla pola spinorowego z $N=8$ otrzymał równanie Diraca a dla oraz $N=4$ równania Maxwella. Procedura jest na tyle ogólna, że umożliwia opis pól Rarity-Schwingera, ogólnej teorii względności oraz wprowadzenie transformacji cechowania [7]. W oparciu o wprowadzone zasady informacyjne Frieden podał również informacyjne wyprowadzenie zasady nieoznaczoności Heisenberga oparte ze statystycznego punktu widzenia o twierdzenie Rao-Cramera dla informacji Fishera oraz jej relację z pojemnością informacyjną układu zapisaną w reprezentacji pędowej, czyli po dokonaniu transformacji Fouriera. Transformacja Fouriera pełni zresztą w całym formalizmie Friedenowskim rolę wyjątkową, będąc jednym z typów samosplątania wewnątrz przestrzeni konfiguracyjnej układu, o czym wspomnimy nieco poniżej. Frieden podał również wyprowadzenie klasycznej fizyki statystycznej, tzn. jej podstawowych rozkładów, Boltzmannna dla energii oraz Maxwella-Boltzmannna dla pędu jak również pewnych rozkładów, które zinterpretował jako

odpowiadające przypadkom nierównowagowym. Kolejną sprawą było wyprowadzenie górnego ograniczenia na tempo zmiany entropii układu dla przykładów klasycznego strumienia cząstek, gęstości rozkładu ładunku, czteropotencjału elektrodynamicznego oraz strumienia cząstek o spinie $1/2$ [7]. Podał również opis teorii pomiaru z szumem wykorzystując wariacyjny formalizm EFI pozwalający na opis redukcji funkcji falowej w trakcie pomiaru urządzeniem dającym swój własny szum. Mianowicie po dokonaniu ekstremalizacji sumy informacji fizycznej K niemierzonego układu oraz funkcjonału opisującego własności układu pomiarowego (a będącego splotem funkcji log-wiarygodności dla funkcji przyrządu splecionej nieliniowo z rozkładem układu), otrzymał równanie ruchu, które (po przejściu do nierelatywistycznej granicy Schrödingera) daje równanie typu Feynmana-Mensky’ego z nieliniowym członem opisującym kolaps funkcji falowej w pomiarze. Ciekawe jest to, że w tym przypadku w pełni ujawnia się traktowanie czasu na równi ze zmiennymi przestrzennymi, czyli jako zmiennej losowej z rozkładem prawdopodobieństwa. Zaproponowana przez nas postać zasad informacyjnych daje formalnie te same równania ewolucji funkcji falowej układu oplecionej funkcją pomiarową przyrządu, jednak otrzymana interpretacja jest zdecydowanie bardziej spójna niż Friedenowska, pozwalając na jednoznaczne rozróżnienie układu poza pomiarem od układu w pomiarze [10].

Przyjęta postać wewnętrznej zasady informacyjnej $K = I + \kappa Q = 0$ [9] (gdzie tzw. współczynnik efektywności $\kappa = 1$ lub $\frac{1}{2}$), postulowana za Friedenem dla każdego omawianego problemu, została ostatnio wyprowadzona dla przypadku $\kappa = 1$ [10]. Sugeruje ona splątanie przestrzeni danych obserwowanych z nieobserwowaną przestrzenią konfiguracyjną układu [10]. Zatem informacja Q nie reprezentuje jedynie informacji strukturalnej układu ale i informację o splątaniu widocznym w korelacji danych w przestrzeni pomiarowej, a EFI może być wykorzystywana jako mechanizm w estymacji stanów splątanych. Np. w przypadku problemu EPR-Bohmy, splątanie zachodzi pomiędzy rzutem spinu obserwowanej cząstki i nieobserwowaną konfiguracją łączną układu, a w przypadku relatywistycznych równań ruchu otrzymujemy splątanie kinetycznych i strukturalnych (masa) stopni swobody, czego wyrazem jest związek stanu obserwowanej cząstki w czasoprzestrzeni z jej własnym stanem w przestrzeni energetyczno-pędowej. Ten drugi przypadek jest przykładem wspomnianego samosplątania opisanego transformatą Fouriera. Ponieważ Q związane jest tu z masą cząstki, zatem w podejściu informacyjnym można wyciągnąć również wniosek, że samosplątanie powinno pomóc w odczytaniu struktury wewnętrznej cząstek. W końcu pojęcie informacji Fishera i jej reinterpretacja przez Friedena jako członu kinetycznego teorii, pozwoliła na przeprowadzenie informacyjnego dowodu [9] o niewyprowadzalności mechaniki kwantowej (falowej) oraz każdej teorii pola, dla której ranga pola N jest skończona, z mechaniki klasycznej.

Temat skryptu dotyczy więc fundamentalnego zagadnienia związanego z określeniem statystycznej procedury estymacji modeli fizycznych. Jego realizacja wymaga znajomości problemów związanych z stosowaniem statystycznej MNW oraz fizycznej EFI dla konstrukcji modeli fizycznych, jak również podstaw metod geometrii różniczkowej.

Na koniec uwaga słownikowa i podsumowanie treści metody EFI. Pojęcie “likelihood function” zostało wprowadzony przez Fishera jako mające związek z prawdopodobieństwem. Również słownikowo powinno być ono przetłumaczone jako “funkcja możliwości”. Zastosowano jednak tłumaczenie “funkcja wiarygodności”. Jako posumowanie istoty przedstawionej metody, powiedzmy, że jest ona wyrazem *zastosowania informacji Fishera w teorii pola w ujęciu Friedena, którego*

inspiracja pochodzi z obszaru optyki.

Temat skryptu związany jest z dociekaniem, które dane mi było mieć wspólnie ze Sławomirem Manią, Dorotą Mroziakiewicz, Janem Sładkowskim, Robertem Szafronem i Sebastianem Zającem, którym za te dociekania i rozmowy dziękuję.

Rozdział 1

Metoda największej wiarygodności

Z powodu możliwości zastosowania MNW do rozwiązania wielu bardzo różnych problemów estymacyjnych metoda ta stała się obecnie zarówno metodą podstawową ale również punktem wyjścia dla różnych metod analizy statystycznej. Jej wszechstronność związana jest z możliwością przeprowadzenia analizy statystycznej dla małej próbki, opisu zjawisk nieliniowych oraz zmiennych losowych posiadających zasadniczo dowolny rozkład prawdopodobieństwa, oraz z powodu szczególnych własności otrzymywanych przez nią estymatorów, które okazują się być zgodne, asymptotycznie nieobciążone, efektywne oraz dostateczne [11]. MNW zasada się na intuicyjnie jasnym postulatcie przyjęcia za prawdziwe takich wartości parametrów funkcji rozkładu prawdopodobieństwa (pdf) zmiennej losowej, które maksymalizują funkcję wiarygodności (realizacji konkretnej próbki).

Rozważmy zmienną losową Y , która przyjmuje wartości \mathbf{y} zgodnie z rozkładem prawdopodobieństwa $p(\mathbf{y}|\Theta)$, gdzie $\Theta = (\theta_1, \theta_2, \dots, \theta_p)^T \equiv (\theta_i)_1^p$, jest zbiorem p parametrów tego rozkładu (T oznacza transpozycję). Gdy $p > 1$ wtedy Θ nazywamy parametrem wektorowym. W szczególnym przypadku $p = 1$, tzn. $\Theta = \theta$, mówimy, że parametr jest skalarny.

Centralnym pojęciem MNW jest funkcja wiarygodności $L(y; \Theta)$ pojawienia się zbioru danych $\mathbf{y} \equiv (\mathbf{y}_n)_{n=1}^N = y_1, y_2, \dots, y_N$ otrzymanych w N obserwacjach zmiennej losowej Y , tzn. (pojawienia się) próbki pobranej z populacji charakteryzowanej przez parametr wektorowy Θ . Przez wzgląd na zapis stosowany w fizyce, wprowadzimy oznaczenie $P(y|\Theta) \equiv L(y; \Theta)$, zapis który podkreśla, że formalnie funkcja wiarygodności jest funkcją łącznego rozkładu prawdopodobieństwa wielu zmiennych losowych $\mathbf{y} \equiv (\mathbf{y}_n)_{n=1}^N$:

$$P(y|\Theta) = \prod_{n=1}^N p_n(y_n|\Theta) . \quad (1.1)$$

Zakładamy, że próba jest prosta w znaczeniu niezależności rozkładów $p_n(y_n|\Theta)$ dla poszczególnych pomiarów n w N elementowej próbie. W przypadku analizy jednej zmiennej losowej Y rozkłady te spełniają dodatkowo związek:

$$p_n(y_n|\Theta) = p(\mathbf{y}|\Theta) . \quad (1.2)$$

Definicja (1.1) może być uogólniona na przypadek, gdy zmienne losowe Y_n próby (Y_1, Y_2, \dots, Y_N) (z realizacjami $(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N)$) mają różne parametry rozkładu dla każdego $n \in N$. Wtedy rozkłady $p_n(y_n|\Theta)$ zmiennych Y_n , chociaż są tego samego typu, jednak na ogół nie spełniają już

związku (1.2), jak to ma miejsce np. w analizie regresji (Rozdział 1.2).

W skrócie funkcję wiarygodności oznaczamy będziemy oznaczali jako $P(\Theta)$. Zapis ten podkreśla, że funkcja ta jest również funkcją nieznanymi parametrów $\Theta = (\theta_i)_1^p$ populacji, które chcemy oszacować. Zbiór wartości parametrów $\Theta = (\theta_i)_1^p$ tworzy współrzędne rozkładu prawdopodobieństwa rozumianego jako punkt w podprzestrzeni przestrzeni statystycznej \mathcal{S} [12]. Temat ten rozwinie dalej.

Będąc funkcją $\Theta = (\theta_i)_1^p$, funkcja wiarygodności służy do konstrukcji estymatorów $\hat{\Theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p)^T \equiv (\hat{\theta}_i)_1^p$ parametrów $\Theta = (\theta_1, \theta_2, \dots, \theta_p)^T \equiv (\theta_i)_1^p$, tzn. wybieramy takie $(\hat{\theta}_i)_1^p$, dla których funkcja wiarygodności przyjmuje maksymalną wartość i nazywamy je estymatorami MNW. Zatem warunek konieczny na otrzymanie estymatorów MNW sprowadza się do znalezienia rozwiązania układu p tzw. równań wiarygodności:

$$S(\Theta) \equiv \frac{\partial}{\partial \Theta} \ln P(y|\Theta) = 0 \quad (1.3)$$

gdzie zagadnienie maksymalizacji sprowadzono do (na ogół) analitycznie równoważnego problemu maksymalizacji logarytmu funkcji wiarygodności $\ln P(y|\Theta)$. Funkcję $S(\Theta)$ będącą gradientem logarytmu funkcji wiarygodności:

$$S(\Theta) \equiv \frac{\partial}{\partial \Theta} \ln P(y|\Theta) = \begin{pmatrix} \frac{\partial \ln P(y|\Theta)}{\partial \theta_1} \\ \vdots \\ \frac{\partial \ln P(y|\Theta)}{\partial \theta_p} \end{pmatrix} \quad (1.4)$$

nazywamy funkcją wynikową.

Po otrzymaniu (wektora) estymatorów $\hat{\Theta}$, maksymalizowaną wartość funkcji wiarygodności definiujemy jako numeryczną wartość funkcji wiarygodności powstałą przez podstawienie w $P(y|\Theta)$ wartości oszacowanej $\hat{\Theta}$ w miejsce parametru Θ .

Przykład: Rozważmy problem estymacji (skalarne) parametru, tzn. $\Theta = \theta$ dla zmiennej losowej y opisanej rozkładem dwumianowym (Bernoulliego)

$$P(y|\theta) = \binom{m}{y} \theta^y (1-\theta)^{m-y} \quad (1.5)$$

Estymacji parametru θ dokonamy na podstawie *pojedynczej* obserwacji (długość próby $n = 1$) zmiennej y , nazywanej czasami częstością. Parametr m charakteryzuje rozkład zmiennej Bernoulliego y (i nie ma związku z długością n próby). Zatem ponieważ $y \equiv (y_1)$, więc $P(y|\theta)$ jest funkcją wiarygodności dla $n = 1$ wymiarowej próby. Jej logarytm wynosi

$$\ln P(y|\theta) = \ln \binom{m}{y} + y \ln \theta + (m-y) \ln (1-\theta) \quad (1.6)$$

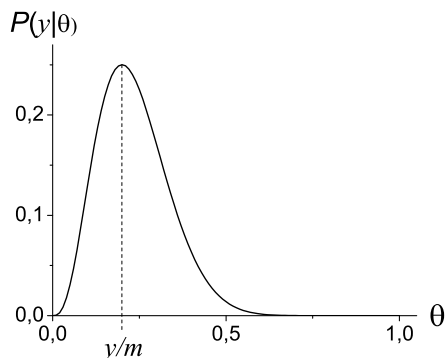
Równanie wiarygodności (1.3) przyjmuje postać:

$$S(\theta) = \frac{1}{\theta} y - \frac{1}{1-\theta} (m-y) \Big|_{\theta=\hat{\theta}} = 0 \quad (1.7)$$

a rozwiązanie go daje estymator MNW parametru θ :

$$\hat{\theta} = \frac{y}{m} \quad (1.8)$$

dla rozkładu dwumianowego. Dobrą ilustracją powyższej procedury znajdowania wartości estymatora parametru θ jest poniższy rysunek (dla $m = 5$)



Rysunek 1.1: Graficzna ilustracja metody największej wiarygodności dla $P(y|\theta)$ określonego wzorem (1.5). Przyjęto zaobserwowaną wartość stosunku $y/m = 1/5$. Maksimum $P(y|\theta)$ przypada na wartość θ równą punktowemu oszacowaniu $\hat{\theta} = y/m = 1/5$ tego parametru. Maksymalizowana wartość funkcji wiarygodności jest równa $P(y|\hat{\theta})$.

1.1 Wnioskowanie w MNW

Konstrukcja punktowego oszacowania parametru w MNW oparta jest o postulat maksymalizacji funkcji wiarygodności przedstawiony powyżej. Jest on wstępem do statystycznej procedury wnioskowania. Kolejnym krokiem jest konstrukcja przedziału wiarygodności. Jest on odpowiednikiem przedziału ufności, otrzymywanego w częstotliwościowym podejściu statystyki klasycznej do procedury estymacyjnej. Do jego konstrukcji niezbędna jest znajomość rozkładu prawdopodobieństwa estymatora parametru, co (dzięki “porządnym” granicznym własnościom stosowanych estymatorów) jest możliwe niejednokrotnie jedynie asymptotycznie, tzn. dla wielkości próby dążącej do nieskończoności. Znajomość rozkładu estymatora jest też niezbędna we wnioskowaniu statystycznym odnoszącym się do weryfikacji hipotez.

W sytuacji, gdy nie dysponujemy wystarczającą ilością danych, potrzebnych do przeprowadzenia skutecznego częstotliwościowego wnioskowania, Fisher [13] zaproponował do określenia niepewności dotyczącej parametru θ wykorzystanie maksymalizowanej wartości funkcji wiarygodności. Wiarygodnościowy przedział ufności jest zdefiniowany jako zbiór wartości parametru θ , dla których wartość funkcji wiarygodności osiąga (umownie) wystarczająco wysoką wartość, tzn.:

$$\left\{ \theta, \frac{P(y|\Theta)}{P(y|\hat{\Theta})} > c \right\} \quad (1.9)$$

dla pewnego parametrem odcięcia c , nazywanego poziomem wiarygodności. Iloraz wiarygodności $P(y|\Theta)/P(y|\hat{\Theta})$ reprezentuje pewien typ unormowanej wiarygodności i jako taki jest wielkością skalarną. Jednak z powodu niejasnego znaczenia określonej wartości parametru obcięcia c pojęcie to wydaje się być na pierwszy rzut oka za słabe aby dostarczyć taką precyzję wypowiedzi jaką daje analiza częstotliwościowa.

Istotnie, wartość c nie odnosi się do żadnej wielkości obserwowanej, tzn. na przykład 1%-we obcięcie ($c = 0,01$) nie ma ścisłego probabilistycznego znaczenia. Inaczej ma się sprawa dla częstotliwościowych przedziałów ufności. W tym przypadku wartość współczynnika $\alpha = 0,01$

oznacza, że gdybyśmy rozważyli realizację przedziału ufności na poziomie ufności $1 - \alpha = 0,99$, to przy pobraniu nieskończonej (w praktyce wystarczająco dużej) liczby próbek, 99% wszystkich wyznaczonych przedziałów ufności pokryłoby prawdziwą wartość parametru θ w populacji. Pomimo tej słabości analizy związanej z MNW zobaczymy, że rozbudowanie analizy stosunku wiarygodności okazuje się być istotne w wnioskowaniu statystycznym analizy doboru modeli, i to aż po konstrukcję równań teorii pola.

Przykład rozkładu normalnego z jednym estymowanym parametrem: Istnieje jednak wyjątek pozwalający na prostą interpretację wiarygodnościowego przedziału ufności. Dotyczy on zmiennej Y posiadającej rozkład Gaussa oraz sytuacji gdy interesuje nas estymacja parametru θ będącego wartością oczekiwaną $E(Y)$ zmiennej Y . Omówimy go poniżej. Częstotliwościowe wnioskowanie o nieznanym parametrze θ wymaga określenia rozkładu jego estymatora. W ogólności dla dużych próbek staramy się o ile to możliwe skorzystać z twierdzeń granicznych dotyczących rozkładu ilorazu wiarygodności, jednak w przypadku rozkładu normalnego możliwa jest konstrukcja skończonego wymiarowa. Niech y_1, \dots, y_N będzie realizacją próby prostej dla zmiennej Y o rozkładzie normalnym $N(\theta, \sigma^2)$ ze znaną wariancją σ^2 . Logarytm funkcji wiarygodności dla $N(\theta, \sigma^2)$ ma postać:

$$\ln P(y; \theta) = -\frac{N}{2} \ln 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - \theta)^2 \quad (1.10)$$

skąd ponieważ estymatorem wartości oczekiwanej jest średnia arytmetyczna $\hat{\theta} = \bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i$, zatem po nieskomplikowanej analizie wyrażenia $\sum_{i=1}^N (y_i - \bar{y})^2 = \sum_{i=1}^N ((y_i - \theta) + (\theta - \bar{y}))^2$ otrzymujemy:

$$\ln \frac{P(y|\theta)}{P(y|\hat{\theta})} = -\frac{N}{2\sigma^2} (\bar{y} - \theta)^2 \quad (1.11)$$

gdzie po prawej stronie otrzymujemy wyrażenie kwadratowe. Ponieważ \bar{Y} jest nieobciążonym estymatorem parametru θ , co oznacza, że wartość oczekiwana $E(\bar{Y}) = \theta$, zatem \bar{Y} ma rozkład normalny $N\left(\theta, \frac{\sigma^2}{N}\right)$. Wynika z tego, że tzw. statystyka ilorazu wiarygodności Wilka:

$$W \equiv 2 \ln \frac{P(y|\hat{\theta})}{P(y|\theta)} \sim \chi_1^2 \quad (1.12)$$

ma w tym przypadku rozkład χ^2 z jednym stopniem swobody [13].

Wykorzystując ten fakt możemy wykonać wyskalowanie wiarygodności oparte o możliwość powiązania przedziału wiarygodności z jego częstotliwościowym odpowiednikiem. Zgodnie z (1.12) wnioskujemy, że dla ustalonego (choć nieznanego parametru θ) prawdopodobieństwo, że iloraz wiarygodności znajduje się w wyznaczonym dla parametru obciążenia c , wiarygodnościowym przedziale ufności, wynosi:

$$P\left(\frac{P(y|\theta)}{P(y|\hat{\theta})} > c\right) = P\left(2 \ln \frac{P(y|\hat{\theta})}{P(y|\theta)} < -2 \ln c\right) = P(\chi_1^2 < -2 \ln c) \quad (1.13)$$

Zatem jeśli dla jakiegoś $0 < \alpha < 1$ wybierzemy parametr obciążenia

$$c = e^{-\frac{1}{2}\chi_{1,(1-\alpha)}^2} \quad (1.14)$$

gdzie $\chi_{1,(1-\alpha)}^2$ jest kwantylem rzędu $100(1 - \alpha)\%$ rozkładu χ -kwadrat, to spełnienie przez θ związku

$$P\left(\frac{P(y|\theta)}{P(y|\hat{\theta})} > c\right) = P\left(\chi_1^2 < \chi_{1,(1-\alpha)}^2\right) = 1 - \alpha \quad (1.15)$$

oznacza, że przyjęcie wartości c zgodnej z (1.14) daje zbiór możliwych wartości parametru θ :

$$\left\{\theta, \frac{P(y|\theta)}{P(y|\hat{\theta})} > c\right\} \quad (1.16)$$

nazywany $100(1 - \alpha)\%$ -owym wiarygodnościowym przedział ufności. Jest on odpowiednikiem wyznaczonego na poziomie ufności $(1 - \alpha)$ częstotliwościowego przedziału ufności dla θ . Zatem przypadek rozkładu normalnego z estymacją skalarnego parametru oczekiwanego poziomu zjawiska daje po skorzystaniu z wzoru (1.14) wartość parametru obciążenia równego $c = 0.15$ lub $c = 0.04$ dla odpowiednio 95%-owego ($1 - \alpha = 0.95$) bądź 99%-owego ($1 - \alpha = 0.99$) przedziału ufności.

Zwróćmy uwagę, że chociaż konstrukcje częstotliwościowego i wiarygodnościowego przedziału ufności są różne, to ich losowość wynika w obu przypadkach z rozkładu prawdopodobieństwa estymatora $\hat{\theta}$.

Ćwiczenie: W oparciu o powyższe rozważania wyznaczyć, korzystając z (1.11) ogólną postać przedziału wiarygodności dla parametru θ jednowymiarowego rozkładu normalnego.

Dla zmiennych o innym rozkładzie niż rozkład normalny, statystyka Wilka W ma w ogólności inny rozkład niż χ^2 [13]. Jeśli więc zmienne nie mają dokładnie rozkładu normalnego, lub dysponujemy za małą próbką by móc odwoływać się do ich rozkładów granicznych wtedy formuła (1.12) więc i (1.14) dają jedynie przybliżone wyskalowanie przedziału wiarygodności. Jednakże w przypadkach wystarczająco regularnych, zdefiniowanych jako takie, w których możemy zastosować przybliżenie kwadratowe:

$$\ln \frac{P(y|\hat{\theta})}{P(y|\theta)} \approx -\frac{1}{2} \text{iF}(\hat{\theta}) (\hat{\theta} - \theta)^2 \quad (1.17)$$

wtedy powyższe rozumowanie jest słuszne w przybliżeniu. Wielkość $\text{iF}(\hat{\theta})$, która pojawiła się powyżej jest (obserwowaną) informacją Fishera, a powyższa formuła stanowi poważne narzędzie w analizie doboru modeli. Można powiedzieć, że cały skrypt koncentruje się na analizie zastosowania (wartości oczekiwanej) tego wyrażenia i jego uogólnień. Do sprawy wrócimy dalej.

Ćwiczenie: Rozważmy przypadek parametru skalarnego θ w jednym eksperymencie ze zmienną y posiadającą rozkład Bernoulliego z $m = 15$. W wyniku pomiaru zaobserwowaliśmy wartość $y = 3$. Prosta analiza pozwala wyznaczyć wiarygodnościowy przedział ufności dla parametru θ . Ponieważ przestrzeń V_θ parametru θ wynosi $V_\theta = (0, 1)$, zatem łatwo pokazać, że dla $c = 0,01$,

$c = 0,1$ oraz $c = 0,5$ miałyby on realizację odpowiednio $(0,019; 0,583)$, $(0,046; 0,465)$ oraz $(0,098; 0,337)$. Widać, że wraz ze wzrostem wartości c przedział wiarygodności zacieśnia się wokół wartości oszacowania punktowego $\hat{\theta} = y/m = 1/5$ parametru θ , i nic dziwnego, bo wzrost wartości c oznacza akceptowanie jako możliwych do przyjęcia tylko takich modelowych wartości parametru θ , które gwarantują wystarczająco wysoką wiarygodność próbki.

Powyższy przykład pozwala nabyć pewnej intuicji co do sensu stosowania ilorazu funkcji wiarygodności. Mianowicie w świetle otrzymania w pomiarze określonej wartości y/m oszacowującej parametr θ , jesteśmy skłonni preferować model z taką wartością parametru θ , która daje większą wartość (logarytmu) ilorazu wiarygodności $P(y|\theta)/P(y|\hat{\theta})$. Zgodnie z podejściem statystyki klasycznej nie oznacza to jednak abymy uważali, że parametr θ ma jakiś rozkład. Jedyne wobec niewiedzy co do modelowej (populacyjnej) wartości parametru θ preferujemy ten model, który daje większą wartość ilorazu wiarygodności w próbce.

1.1.1 Weryfikacja hipotez z wykorzystaniem ilorazu wiarygodności

Funkcję wiarygodności można również wykorzystać w celu weryfikacji hipotez statystycznych, drugiego typu wnioskowania statystycznego. Rozważmy prostą hipotezę zerową $H_0 : \theta = \theta_0$ wobec złożonej hipotezy alternatywnej $H_1 : \theta \neq \theta_0$. W tym celu wprowadźmy unormowaną funkcję wiarygodności:

$$\frac{P(y|\theta_0)}{P(y|\hat{\theta})} \tag{1.18}$$

skonstruowaną przy założeniu prawdziwości hipotezy zerowej. Hipotezę zerową odrzucamy na rzecz hipotezy alternatywnej, jeśli jej wiarygodność jest „za mała”. Sugeruje to, że złożona hipoteza alternatywna zawiera pewną hipotezę prostą, która jest lepiej poparta przez otrzymane w próbce dane niż hipoteza zerowa.

Jak o tym wspomnieliśmy powyżej, np. 5%-owy poziom wiarygodności, nic sam w sobie nie mówi o frakcji liczby przedziałów wiarygodności pokrywających nieznaną wartość szacowanego parametru. Potrzebne jest wyskalowanie ilorazu wiarygodności. Również dla weryfikacji hipotez skalowanie wiarygodności jest istotne. Stwierdziliśmy, że takie skalowanie jest możliwe wtedy gdy mamy do czynienia z jednoparametrowym przypadkiem rozkładu Gaussa, a przynajmniej z przypadkiem wystarczająco regularnym.

W przypadku jednoparametrowego, regularnego problemu, skalowanie poprzez wykorzystanie statystyki Wilka służy otrzymaniu empirycznego poziomu istotności p . Z wzoru (1.12) otrzymujemy wtedy przybliżony (a dokładny dla rozkładu normalnego) empiryczny poziom istotności:

$$p = P(\chi_1^2 > -2 \ln c) . \tag{1.19}$$

Definicja ta oznacza, że istnieje typowy związek pomiędzy stosunkiem wiarygodności a empirycznym poziomem istotności, podobnie jak to jest dla relacji pomiędzy poziomem ufności $1 - \alpha$ a poziomem istotności α w analizie częstotliwościowej. Zatem w przypadku jednoparametrowego rozkładu normalnego możemy wykorzystać empiryczny poziom istotności p do określenia interesującej nas wartości parametru obciążenia c .

Jednakże podobne skalowanie ilorazu wiarygodności okazuje się być znacznie trudniejsze już chociażby tylko w przypadku dwu-parametrowego rozkładu normalnego, gdy obok θ estymujemy σ^2 [13]. Wtedy określenie co oznacza sformułowanie „zbyt mała” wartość c jest dość dowolne i

zależy od rozważanego problemu lub wcześniejszej wiedzy wynikającej z innych źródeł niż prowadzone statystyczne wnioskowanie. Wybór dużego parametru obciążenia c spowoduje, że istnieje większe prawdopodobieństwo popełnienia błędu pierwszego rodzaju polegającego na odrzuceniu hipotezy zerowej w przypadku, gdy jest ona prawdziwa. Wybór małego c spowoduje zwiększenie popełnienia błędu drugiego rodzaju, tzn. przyjęcia hipotezy zerowej w sytuacji gdy jest ona błędna.

1.2 MNW w analizie regresji

Analiza zawarta w całym Rozdziale 1.2 oparta jest na przedstawieniu metody MNW w analizie regresji klasycznej w pracy Kleinbauma, Kuppera, Mullera i Nizama [14] oraz eleganckim opracowaniu Mroziakiewicz [15]. W metodzie regresji klasycznej, estymatory parametrów strukturalnych modelu regresji są otrzymane arytmetyczną metodą najmniejszych kwadratów (MNK). Zmienne objaśniające x_i , $i = 1, \dots, N$ w regresji klasycznej nie mają charakteru stochastycznego, co oznacza, że eksperyment jest kontrolowany ze względu na te zmienne. MNW polega na minimalizacji sumy kwadratów odchyłek obserwowanych wartości zmiennej objaśnianej od ich wartości teoretycznych spełniających równanie regresji. MNK ma znaczenie probabilistyczne tylko w przypadku analizy standardowej, gdy zmienna objaśniana ma rozkład normalny. Jej estymatory pokrywają się wtedy z estymatorami MNW. Istotnie, łatwo zauważyć, że tak się sprawy mają: Załóżmy, że zmienne (Y_1, Y_2, \dots, Y_N) odpowiadające kolejnym wartościom zmiennej objaśnianej są względem siebie niezależne, mają rozkład normalny ze średnią $\mu_i = E(Y | x_i) = E(Y_i)$ zależną od wariantu zmiennej objaśniającej x_i , oraz (na ogół) taką samą wariancją $\sigma^2(Y_i) = \sigma^2(Y)$. Funkcja wiarygodności (dla próbki (y_1, y_2, \dots, y_N)) dla normalnego klasycznego modelu regresji ma postać:

$$\begin{aligned} P(y) &= \prod_{i=1}^n f(y_i, \mu_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2} (y_i - \mu_i)^2\right\} \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu_i)^2\right\}. \end{aligned} \quad (1.20)$$

Widać, że maksymalizacja $P(\mu_i)$ ze względu na $(\mu_i)_{i=1}^N$ pociąga za sobą minimalizację sumy kwadratów reszt (SKR)¹

$$SKR = \sum_{i=1}^n (y_i - \mu_i)^2, \quad (1.21)$$

gdzie $\mu_i = E(Y | x_i)$ jest postulowanym modelem regresji. Zatem w standardowej, klasycznej analizie regresji, estymatory MNW pokrywają się z estymatorami MNK. Widać, że procedura maksymalizacji dla SKR prowadzi do liniowej postaci estymatorów $\hat{\mu}_i$ parametrów μ_i .

Jednak rozwiązanie układu równań wiarygodności (1.3) jest zazwyczaj nietrywialne. Jest tak, gdy otrzymany w wyniku ekstremizacji układ algebraicznych równań wiarygodności na estymatory jest nieliniowy. W konsekwencji możemy nie otrzymać zwartej analitycznej postaci estymatorów. Przykładem może być analiza regresji Poissona, w której do rozwiązania równań wiarygodności wykorzystujemy metody iteracyjne. W takich sytuacjach na ogół odwołujemy się

¹SSE w literaturze angielskiej.

do programów komputerowych dla analizy statystycznej (np. SAS). Po podaniu postaci funkcji wiarygodności program komputerowy dokonuje jej maksymalizacji rozwiązując układ (1.3) np. metodą R-Newtona [13, 15] i licząc numerycznie wartości estymatorów parametrów modelu.

Funkcja (logarytmu) ilorazu wiarygodności jest również wykorzystywana w analizie regresji dla przeprowadzania testów statystycznych dla weryfikacji hipotez o istotności rozszerzenia modelu mniej złożonego, tzw. “niższego”, o mniejszej liczbie parametrów, do bardziej złożonego modelu “wyższego”, posiadającego większą liczbę parametrów. Statystyka wykorzystywana do tego typu testów (związana z pojęciem dewiancji D [14, 13, 16]) ma postać:

$$-2 \ln \frac{P(y|\hat{\Theta}_1)}{P(y|\hat{\Theta}_2)} \quad (1.22)$$

gdzie $P(y|\hat{\Theta}_1)$ jest maksymalizowaną wartością funkcji wiarygodności dla modelu mniej złożonego, a $P(y|\hat{\Theta}_2)$ dla modelu bardziej złożonego. Przy prawdziwości hipotezy zerowej H_0 o braku konieczności rozszerzania modelu niższego do wyższego, statystyka (1.22) ma asymptotycznie rozkład χ^2 z liczbą stopni swobody równą różnicy liczby parametrów modelu wyższego i niższego.

Maksymalizowana wartość funkcji wiarygodności zachowuje się jak współczynnik determinacji R^2 [14, 15], tzn. rośnie wraz ze wzrostem liczby parametrów w modelu, zatem wielkość pod logarytmem należy do przedziału $(0, 1)$ i statystyka (1.22) przyjmuje wartości z przedziału $(0, +\infty)$, stąd (asymptotycznie) zbiór krytyczny dla H_0 jest prawostronny. Im lepiej zatem model wyższy dopasowuje się do danych empirycznych w stosunku do modelu niższego, tym większa jest wartość statystyki ilorazu wiarygodności (1.22) i większa szansa, że wpadnie ona w przedział odrzuceń dla hipotezy zerowej H_0 , który leży w prawym ogonie wspomnianego rozkładu χ^2 [14, 15].

1.2.1 Dewiancja jako miara dobroci dopasowania. Przykład rozkładu Poissona.

Rozważmy zmienną losową Y posiadającą rozkład Poissona. Rozkład ten jest wykorzystywany do modelowania zjawisk związanych z rzadko zachodzącymi zdarzeniami, jak na przykład z liczbą rozpadających się niestabilnych jąder w czasie t czy liczbą wypadków samochodowych mających miejsce w pewnej okolicy w ciągu roku. Ma on postać:

$$p(Y = y|\mu) = \frac{\mu^y e^{-\mu}}{y!}, \quad \text{oraz } y = 0, 1, \dots, \infty, \quad (1.23)$$

gdzie μ jest parametrem rozkładu. Zmienna losowa podlegająca rozkładowi Poissona może przyjmować tylko nieujemną wartość całkowitą. Rozkład ten można wyprowadzić z rozkładu dwumianowego, bądź wykorzystując rozkłady Erlanga i wykładniczy [11]. Rozkład Poissona posiada pewną interesującą właściwość statystyczną, mianowicie jego wartość oczekiwana, wariancja i trzeci moment centralny są równe parametrowi rozkładu μ :

$$E(Y) = \sigma^2(Y) = \mu_3 = \mu. \quad (1.24)$$

Aby pokazać dwie pierwsze równości w Eq.(1.24) skorzystajmy bezpośrednio z definicji odpowiednich momentów, otrzymując:

$$\begin{aligned} E(Y) &= \sum_{y=0}^{\infty} y \cdot p(Y = y; \mu) = \sum_{y=0}^{\infty} y \cdot \frac{\mu^y e^{-\mu}}{y!} = e^{-\mu} \sum_{y=1}^{\infty} \frac{\mu^y}{(y-1)!} \\ &= e^{-\mu} \mu \sum_{y=1}^{\infty} \frac{\mu^{y-1}}{(y-1)!} = e^{-\mu} \mu \sum_{l=0}^{\infty} \frac{\mu^l}{l!} = e^{-\mu} \mu e^{\mu} = \mu, \end{aligned}$$

oraz, korzystając z powyższego:

$$\begin{aligned} \sigma^2(Y) &= E(Y^2) - [E(Y)]^2 = E(Y^2) - \mu^2 = \sum_{y=0}^{\infty} y^2 \cdot p(Y = y; \mu) - \mu^2 \\ &= \sum_{y=0}^{\infty} y^2 \cdot \frac{\mu^y e^{-\mu}}{y!} - \mu^2 = e^{-\mu} \sum_{y=1}^{\infty} y \frac{\mu^y}{(y-1)!} - \mu^2 = e^{-\mu} \mu \sum_{l=0}^{\infty} (l+1) \frac{\mu^l}{l!} - \mu^2 \\ &= e^{-\mu} \mu \left[\sum_{l=0}^{\infty} l \frac{\mu^l}{l!} + e^{\mu} \right] - \mu^2 = e^{-\mu} \mu [e^{\mu} \mu + e^{\mu}] - \mu^2 = (\mu^2 + \mu) - \mu^2 = \mu. \end{aligned}$$

Ćwiczenie: Pokazać Eq.(1.24) dla trzeciego momentu.

Rozważmy model regresji dla zmiennej objaśnianej Y (tzw. odpowiedzi) posiadającej rozkład Poissona. Zmienne Y_i , $i = 1, 2, \dots, N$ posiadają więc również rozkład Poissona i zakładamy, że są parami wzajemnie niezależne. Zatem $(Y_i)_{i=1}^N$ tworzy próbę prostą. Niech X jest zmienną objaśniającą (tzw. czynnikiem) kontrolowanego eksperymentu, w którym X nie jest zmienną losową, ale jej zmiana, jest rozważana jako możliwa przyczyna warunkująca nielosową zmianę wartości zmiennej Y . Oczywiście czynników X_1, X_2, \dots, X_k może być więcej.

Ignorując zależność zmiennej Y od czynników X_1, X_2, \dots, X_k , rozważamy tzw. *model podstawowy*. Dla rozkładu (1.23) i próby $y \equiv (Y_i)_{i=1}^N$, funkcja wiarygodności ma postać:

$$P(y|\mu) = \prod_{i=1}^N \frac{\mu_i^{Y_i} e^{-\mu_i}}{Y_i!} = \frac{\left(\prod_{i=1}^N \mu_i^{Y_i} \right) \exp\left(-\sum_{i=1}^N \mu_i\right)}{\prod_{i=1}^N Y_i!}, \quad (1.25)$$

tzn. jest wyrażona jedynie jako ogólna funkcja zbioru parametrów $\mu \equiv (\mu_1, \mu_2, \dots, \mu_N)$, gdzie $\mu_i = E(Y_i)$. Rozważmy układ równań MNW:

$$\frac{\partial}{\partial \mu_i} [\ln P(y|\mu)] = 0, \quad i = 1, 2, \dots, N. \quad (1.26)$$

Dla funkcji wiarygodności (1.25) otrzymujemy:

$$\ln P(y|\mu) = \sum_{i=1}^N Y_i \ln \mu_i - \sum_{i=1}^N \mu_i - \sum_{i=1}^N \ln Y_i!. \quad (1.27)$$

Zatem rozwiązanie układu (1.26), daje:

$$\mu_i = \hat{\mu}_i = Y_i, \quad i = 1, 2, \dots, N, \quad (1.28)$$

jako estymatory modelu podstawowego. Zatem funkcja wiarygodności (1.25) modelu podstawowego przyjmuje w punkcie μ zadanym przez estymatory (1.28) wartość maksymalną:

$$P(y|\hat{\mu}) = \frac{\left(\prod_{i=1}^N Y_i^{Y_i} \right) \exp\left(-\sum_{i=1}^N Y_i\right)}{\prod_{i=1}^N Y_i!}, \quad (1.29)$$

gdzie zastosowano oznaczenie $\hat{\mu} = (\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_N)$.

1.2.2 Analiza regresji Poissona.

Niech zmienna zależna Y reprezentuje liczbę zliczeń badanego zjawiska (np. przypadków awarii określonego zakupionego sprzętu), otrzymaną dla każdej z N podgrup (np. klienckich). Każda z tych podgrup wyznaczona jest przez komplet wartości zmiennych objaśniających $x \equiv (X_1, X_2, \dots, X_k)$ (np. wiek, poziom wykształcenia, cel nabycia sprzętu). Zmienna Y_i określa liczbę zliczeń zjawiska w i -tej podgrupie, $i = 1, 2, \dots, N$. W konkretnej próbie $(Y_i)_{i=1}^N = (y_i)_{i=1}^N$. Niech $x_i = (X_{1i}, X_{2i}, \dots, X_{ki})$ oznacza zbiór wartości jednego wariantu zmiennych (X_1, X_2, \dots, X_k) , tzn. dla jednej konkretnej podgrupy i .

Rozważmy następujący model regresji Poissona²:

$$\mu_i \equiv E(Y_i) = \ell_i r(x_i, \beta), \quad i = 1, 2, \dots, N, \quad (1.30)$$

opisujący zmianę wartości oczekiwanej liczby zdarzeń Y_i (dla rozkładu Poissona) wraz ze zmianą wariantu x_i . Funkcja regresji po prawej stronie (1.30) ma dwa czynniki. Pierwszy funkcyjny czynnik funkcji regresji, $r(x_i, \beta)$, opisuje *tempo zdarzeń* określanych mianem porażek w i -tej podgrupie (tzn. jest częstością tego zjawiska), a co za tym idzie $r(x_i, \beta) > 0$, gdzie $\beta \equiv (\beta_0, \beta_1, \dots, \beta_k)$ jest zbiorem nieznanymi parametrów tego modelu regresji. Natomiast drugi czynnik ℓ_i jest współczynnikiem określającym *dla każdej i -tej podgrupy* (np. klientów) *skumulowany czas prowadzenia badań kontrolnych dla wszystkich jednostek tej podgrupy*. Ponieważ funkcja regresji³ $r(x_i, \beta)$ przedstawia typową liczbę porażek na jednostkę czasu, zatem nazywamy ją *ryzykiem*.

Uwaga o postaci funkcji regresji: Funkcję $r(x_i, \beta)$ można zamodelować na różne sposoby [13]. Wprowadzając oznaczenie:

$$\lambda_i^* \equiv \beta_0 + \sum_{j=1}^k \beta_j X_{ji}, \quad (1.31)$$

$r(x_i, \beta)$ ma różną postać w zależności od typu danych. Może mieć ona postać charakterystyczną dla regresji liniowej (wielokrotnej), $r(x_i, \beta) = \lambda_i^*$, którą stosujemy szczególnie wtedy gdy zmienna Y ma rozkład normalny. Postać $r(x_i, \beta) = 1/\lambda_i^*$ jest stosowana w analizie z danymi pochodzącymi z rozkładu exponentialnego, natomiast $r(x_i, \beta) = 1/(1 + \exp(-\lambda_i^*))$ w modelowaniu regresji logistycznej dla opisu zmiennej dychotomicznej [14, 13].

W regresji Poissona użyteczna jest następująca postać funkcji regresji:

$$r(x_i, \beta) = \exp(\lambda_i^*), \quad \lambda_i^* = \beta_0 + \sum_{j=1}^k \beta_j X_{ji}. \quad (1.32)$$

Ogólniej mówiąc analiza regresji odnosi się do modelowania wartości oczekiwanej zmiennej zależnej (objaśnianej) jako funkcji pewnych czynników. Postać funkcji wiarygodności stosowanej do estymacji współczynników regresji β odpowiada założeniom dotyczącym rozkładowi zmiennej zależnej. Tzn. zastosowanie konkretnej funkcji regresji $r(x_i, \beta)$, np. jak w (1.32), wymaga określenia postaci funkcji częstości $r(x_i, \beta)$, zgodnie z jej postacią dobraną do charakteru losowej

²W ogólności zachodzi związek: $\mu_i \equiv E(Y_i) = \ell_i r(x_i, \beta) = \ell_i \Delta t \lambda(x_i, \beta)$, gdzie Δt jest przedziałem czasu, a $\lambda(x_i, \beta)$ tzw. funkcją intensywności. W przedstawionej analizie przyjęto $\Delta t = 1$ i dlatego utożsamiamy funkcję ryzyka $r(x_i, \beta)$ z funkcją intensywności.

³Czynnik $r(x_i, \beta)$ nazywany dalej funkcją regresji, chociaż właściwie nazwa ta odnosi się do całej $E(Y_i)$.

zmiennej przy której generowane są dane w badanym zjawisku. Na ogół przy konstrukcji $r(x_i, \beta)$ pomocna jest uprzednia wiedza dotycząca relacji między rozważanymi zmiennymi.

Ponieważ Y_i ma rozkład Poissona (1.23) ze średnia μ_i , $p(Y_i|\mu_i) = \frac{\mu_i^{Y_i}}{Y_i!} e^{-\mu_i}$, $i = 1, 2, \dots, N$, zatem dane $Y_i = 0, 1, \dots, \infty$ dla $i = 1, 2, \dots, N$ są generowane z rozkładów warunkowych:

$$p(Y_i|\beta) = \frac{[\ell_i r(x_i, \beta)]^{Y_i}}{Y_i!} e^{-\ell_i r(x_i, \beta)}, \quad (1.33)$$

wokół funkcji regresji, (1.30), $\mu_i = \ell_i r(x_i, \beta)$, dla $i = 1, 2, \dots, N$. Funkcja wiarygodności dla analizy regresji Poissona ma więc postać:

$$\begin{aligned} P(y|\beta) &= \prod_{i=1}^N p(Y_i|\beta) = \prod_{i=1}^N \frac{(\ell_i r(x_i, \beta))^{Y_i} e^{-\ell_i r(x_i, \beta)}}{Y_i!} \\ &= \frac{\prod_{i=1}^N (\ell_i r(x_i, \beta))^{Y_i} \exp\left[-\sum_{i=1}^N \ell_i r(x_i, \beta)\right]}{\prod_{i=1}^N Y_i!}. \end{aligned} \quad (1.34)$$

Aby w praktyce posłużyć się funkcją regresji $r(x_i, \beta)$ będącą określoną funkcją zmiennej $\lambda_i^* = \beta_0 + \sum_{j=1}^k \beta_j X_{ji}$, parametry $\beta_0, \beta_1, \dots, \beta_k$ muszą być oszacowane. Estymatory MNW, $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$, tych parametrów otrzymuje się rozwiązując $k + 1$ równań wiarygodności:

$$\frac{\partial}{\partial \beta_j} \ln P(y|\beta) = 0, \quad j = 0, 1, 2, \dots, k. \quad (1.35)$$

W przypadku regresji Poissona $P(y|\beta)$ jest określona zgodnie z (1.34). Zauważmy, że dla rozkładu Poissona zachodzi zgodnie z (1.24) oraz (1.30), $\sigma^2(Y_i) = E(Y_i) = \ell_i r(x_i, \beta)$, co oznacza, że wariancja $\sigma^2(Y_i)$ zmiennej objaśnianej nie jest stała lecz zmienia się jako funkcja ℓ_i oraz x_i , wchodząc w analizę z różnymi wagami wraz ze zmianą wariantu i . Ponieważ układ (1.35) jest na ogół rozwiązywany iteracyjnymi metodami numerycznymi [14], a wariancja $\sigma^2(Y_i)$ jest również funkcją β , zatem na każdym kroku procesu iteracyjnego wagi te zmieniają się jako funkcja zmieniających się składowych estymatora $\hat{\beta}$. Algorytmy takiej analizy określa się ogólnym mianem *algorytmów najmniejszych kwadratów⁴ iteracyjnie ważonych* (IRLS) Różne programy do analiz statystycznych, w tym SAS wykorzystujący procedurę PROC GENMOD, mogą być użyte do znajdowania estymatorów $\hat{\beta}$ MNW dla funkcji wiarygodności (1.34). Również macierz kowariancji MNW estymatorów⁵ $\hat{V}(\hat{\beta})$ oraz miary dobroci dopasowania modelu, takie jak omówiona poniżej dewiancja, mogą być otrzymane przy użyciu tego typu programów.

Maksymalna wartość funkcji wiarygodności $P(y|\mu)$ wyznaczona w oparciu o (1.29) będzie, dla każdego zbioru danych i dla liczby parametrów $k + 1 < N$, większa niż otrzymana przez maksymalizację funkcji wiarygodności (1.34). Jest tak, ponieważ w wyrażeniu (1.29) na funkcję

⁴Należy jednak pamiętać, że zwrotu “najmniejszych kwadratów” nie należy tu brać dosłownie, gdyż metoda najmniejszych kwadratów ma sens jedynie wtedy, gdy rozkład zmiennej Y jest normalny.

⁵Która jest odwrotnością (macierzy) obserwowanej informacji Fishera (2.3) [13]:

$$\hat{V}(\hat{\beta}) = \mathbf{IF}^{-1}(\hat{\beta}). \quad (1.36)$$

wiarygodności modelu podstawowego *nie narzuca się żadnych ograniczeń na postać μ_i* , natomiast (1.34) wymaga aby $\mu_i = \ell_i r(x_i, \beta)$.

Zgodnie z powyższym zdaniem, analizę doboru modelu regresji można rozpocząć od postawienia hipotezy zerowej wobec alternatywnej. W hipotezie zerowej wyróżnimy proponowany model regresji. Wybór modelu badanego oznacza wybór funkcji wiarygodności (1.34) z nim związanej. Stawiamy więc hipotezę zerową:

$$H_0 : \mu_i = \ell_i r(x_i, \beta), i = 1, 2, \dots, N, \quad (1.37)$$

która odpowiada wyborowi modelu z funkcją wiarygodności (1.34), wobec hipotezy alternatywnej:

$$H_A : \mu_i \text{ nie ma ograniczonej postaci, } i = 1, 2, \dots, N, \quad (1.38)$$

która odpowiada wyborowi modelu podstawowego zawierającego tyle parametrów μ_i ile jest punktów pomiarowych, tzn. N , z funkcją wiarygodności (1.29).

Niech więc $P(y, \hat{\beta})$ jest maksymalną wartością funkcji wiarygodności określoną jak w (1.34). Oznacza to, że w miejsce miejsce parametrów $\beta = (\beta_0, \beta_1, \dots, \beta_k)$ podstawiono ich estymatory $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k)$ wyznaczone przez MNW, jako te które maksymalizują funkcję wiarygodności (1.34). Podobnie rozumiemy funkcję wiarygodności $P(y|\hat{\mu})$ modelu podstawowego.

Ponieważ celem każdej analizy jest otrzymanie możliwie najprostszego opisu danych, model $\mu_i = \ell_i r(x_i, \beta)$ zawierający $k + 1$ parametrów β , będzie uznany za dobry, jeśli maksymalna wartość funkcji wiarygodności wyznaczona dla niego, będzie prawie tak duża, jak funkcji wiarygodności dla nie niosącego żadnej informacji modelu podstawowego, z liczbą parametrów μ_i równą licznie punktów pomiarowych N . Sformułowanie „prawie tak duża” oznacza, że wartość funkcji wiarygodności $P(y|\hat{\beta})$ nie może być istotnie statystycznie mniejsza od $P(y|\hat{\mu})$. Zasadniczo powinno to oznaczać, że musimy podać miary pozwalające na określenie statystycznej istotności przy posługiwaniu się intuicyjnym parametrem obciążenia c (Rozdział 1.1). Okazuje się, że dla dużej próby miary typu (1.39) nabywają cech pozwalającego na budowanie wiarygodnościowych obszarów krytycznych nabywających charakteru standardowego (częstotliwościowego).

Określmy *statystykę typu ilorazu wiarygodności*:

$$D(\hat{\beta}) = -2 \ln \left[\frac{P(y|\hat{\beta})}{P(y|\hat{\mu})} \right] \quad (1.39)$$

nazywaną *dewiancją* (deviance) dla modelu regresji, w tym przypadku dla modelu Poissona z określoną postacią $\mu_i = \ell_i r(x_i, \beta)$. Służy ona do badania dobroci dopasowania modelu z zadaną postacią $\mu_i = \ell_i r(x_i, \beta)$ w stosunku do modelu podstawowego, bez narzuconej postaci na μ_i , tzn. do stwierdzenia, czy $P(y|\hat{\beta})$ jest istotnie *mniejsza* od $P(y|\hat{\mu})$, sugerując w ten sposób istotnie statystyczny brak dopasowania badanego modelu $\mu_i = \ell_i r(x_i, \beta)$, do danych empirycznych. Jak pokażemy poniżej dewiancja może być rozumiana jako *miara zmienności reszt* (tzn. odchylenia wartości obserwowanych w próbie od wartości szacowanych przez model) *wokół linii regresji*.

Przy prawdziwości hipotezy $H_0 : \mu_i = \ell_{ir}(x_i, \beta)$, rozkład dewiancji $D(\hat{\beta})$ dla regresji Poissona, można asymptotycznie przybliżyć rozkładem chi-kwadrat (por. dyskusja w [13, 14]) z $N - k - 1$ stopniami swobody.

Zatem statystyczny test dobroci dopasowania, tzn. nie występowania braku dopasowania badanego modelu $H_0 : \mu_i = \ell_{ir}(x_i, \beta)$, w stosunku do modelu podstawowego, przebiega więc w modelu Poissona następująco: Porównujemy otrzymaną w próbie wartość statystyki $D(\hat{\beta})$ z wartością krytyczną leżącą w prawym ogonie rozkładu chi-kwadrat (o $N - k - 1$ stopniach swobody). Przyjęcie przez $D(\hat{\beta})$ wartości równej lub większej od krytycznej skutkuje odrzuceniem hipotezy zerowej.

Podana liczba stopni swobody dewiancji $D(\hat{\beta})$ wynika z następującego rozumowania. Zapiszmy (1.39) w postaci:

$$D(\hat{\beta}) + 2 \ln P(y|\hat{\beta}) = 2 \ln P(y|\hat{\mu}) , \quad (1.40)$$

co po skorzystaniu z (1.34) dla $\beta = \hat{\beta}$ ma postać:

$$D(\hat{\beta}) + 2 \sum_{i=1}^N \ell_{ir}(x_i, \hat{\beta}) = 2 \ln P(y|\hat{\mu}) + 2 \ln \left(\prod_{i=1}^N Y_i! \right) - 2 \ln \left(\prod_{i=1}^N (\ell_{ir}(x_i, \hat{\beta}))^{Y_i} \right) . \quad (1.41)$$

Można zauważyć, że prawa strona tego równania ma N -stopni swobody. Istotnie, ze względu na (1.28)⁶, $\hat{\mu} \equiv (\hat{\mu}_i) = (Y_i)$, $i = 1, 2, \dots, N$, liczba niezależnych zmiennych po prawej strony powyższego równania, których wartości trzeba określić z eksperymentu wynosi N . Natomiast drugi składnik po lewej stronie ma liczbę stopni swobody równą $k + 1$, co jest liczbą estymatorów parametrów strukturalnych $\hat{\beta}$ modelu regresji, których wartości trzeba określić z eksperymentu. Ponieważ liczba stopni swobody po prawej i lewej stronie równania musi być taka sama, zatem liczba stopni swobody dewiancji $D(\hat{\beta})$ wynosi $N - k - 1$.

Dewiancje dla hierarchicznych klas modeli mogą służyć do budowy testów stosunku wiarygodności. Zwróćmy szczególnie uwagę na funkcję wiarygodności (1.34) zawierającą zbiór parametrów $\beta = (\beta_0, \beta_1, \dots, \beta_k)$ z dewiancją $D(\hat{\beta})$ daną wyrażeniem (1.39). Przypuśćmy, że chcemy zweryfikować hipotezę o tym, że $k - r$ (gdzie $0 < r < k$) ostatnich parametrów będących składowymi wektora β jest równych zero.

Hipoteza zerowa, o nieistotności rozszerzenia modelu wyższego do niższego, ma wtedy postać:

$$H_0 : \beta_{r+1} = \beta_{r+2} = \dots = \beta_k = 0 , \quad (1.42)$$

Hipoteza alternatywna H_A mówi, że przynajmniej jeden z parametrów strukturalnych $\beta_{r+1}, \beta_{r+2}, \dots, \beta_k$ jest różny od zera.

Funkcja wiarygodności przy prawdziwości hipotezy zerowej H_0 , (1.42), ma postać taką jak w (1.34), tyle, że zastąpiono w niej parametr β parametrem $\beta_{(r)}$:

$$\beta_{(r)} \equiv (\beta_0, \beta_1, \dots, \beta_r; 0, 0, \dots, 0) \text{ gdzie liczba zer wynosi } k - r . \quad (1.43)$$

⁶W przyjętym przedstawieniu danych jak dla diagramu punktowego, N jest ogólnie liczbą punktów pomiarowych (równą liczbie wariantów czy komórek). Tylko dla modelu podstawowego jest N również liczbą parametrów.

Oznaczmy funkcje wiarygodności tego modelu jako $P(y|\beta_{(r)})$ a $\hat{\beta}_{(r)}$ jest estymatorem MNW wektorowego parametru $\beta_{(r)}$, wyznaczonym przez rozwiązanie odpowiadającego mu układu równań wiarygodności (oczywiście dla niezerowych parametrów $\beta_0, \beta_1, \dots, \beta_r$). Estymator $\hat{\beta}_{(r)} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_r; 0, 0, \dots, 0)$ maksymalizuje funkcję wiarygodności $P(y|\beta_{(r)})$.

Test ilorazu wiarygodności dla weryfikacji hipotezy H_0 przeprowadzamy posługując się statystyką *statystyką ilorazu wiarygodności*:

$$-2 \ln \left[\frac{P(y|\hat{\beta}_r)}{P(y|\hat{\beta})} \right], \quad (1.44)$$

która przy prawdziwości hipotezy zerowej ma asymptotycznie rozkład chi-kwadrat z $k - r$ stopniami swobody, co widać, zapisując (1.44) jako różnicę dewiancji:

$$-2 \ln \left[\frac{P(y|\hat{\beta}_{(r)})}{P(y|\hat{\beta})} \right] = -2 \ln \left[\frac{P(y|\hat{\beta}_{(r)})}{P(y|\hat{\mu})} \right] + 2 \ln \left[\frac{P(y|\hat{\beta})}{P(y|\hat{\mu})} \right] = D(\hat{\beta}_{(r)}) - D(\hat{\beta}), \quad (1.45)$$

i korzystając z podobnej analizy jak dla (1.41).

Zatem, przy prawdziwości hipotezy zerowej (1.42), którą można zapisać jako $H_0 : \beta_{r+1} = \beta_{r+2} = \dots = \beta_k = 0$, różnica $D(\hat{\beta}_{(r)}) - D(\hat{\beta})$ ma dla dużej próby w przybliżeniu rozkład chi-kwadrat z $k - r$ stopniami swobody.

Wniosek: Zatem, jeśli używany regresji Poissona do analizowania danych empirycznych, modele tworzące hierarchiczne klasy mogą być porównywane między sobą poprzez wyznaczenie statystyki ilorazu wiarygodności (1.44), lub co na jedno wychodzi, poprzez wyznaczenie różnicy między parami dewiancji dla tych modeli, pamiętając o wniosku jaki już znamy z analizy dewiancji, że *im model gorzej dopasowuje się do danych empirycznych tym jego dewiancja jest większa*.

1.2.2.1 Podobieństwo dewiancji do SKR analizy częstotliwościowej

Warunkowe wartości oczekiwane $\mu_i \equiv E(Y_i) = \ell_i r(x_i, \beta)$, $i = 1, 2, \dots, N$, (1.30), są w analizie regresji przyjmowane jako teoretyczne przewidywania modelu regresji dla wartości zmiennej objaśnianej Y_i , zwanej odpowiedzią (układu). W próbie odpowiadają im oszacowania (oznaczone jako) \hat{Y}_i , które w i -tej komórce są następująco:

$$\hat{Y}_i = \ell_i r(x_i, \hat{\beta}), \quad i = 1, 2, \dots, N, \quad (1.46)$$

zgodnie z wyestymowaną postacią modelu regresji. Wykorzystując (1.46) możemy zapisać dewiancję modelu (1.39) następująco:

$$\begin{aligned} D(\hat{\beta}) &= -2 \ln \left[\frac{P(y|\hat{\beta})}{P(y|\hat{\mu})} \right] = -2 \ln \left[\frac{\prod_{i=1}^N \hat{Y}_i^{Y_i} \exp\left(-\sum_{i=1}^N \hat{Y}_i\right)}{\prod_{i=1}^N Y_i^{Y_i} \exp\left(-\sum_{i=1}^N Y_i\right)} \right] \\ &= -2 \left[\sum_{i=1}^N Y_i \ln \hat{Y}_i - \sum_{i=1}^N \hat{Y}_i - \sum_{i=1}^N Y_i \ln Y_i + \sum_{i=1}^N Y_i \right] \end{aligned} \quad (1.47)$$

tzn:

$$D(\hat{\beta}) = 2 \sum_{i=1}^N \left[Y_i \ln \left(\frac{Y_i}{\hat{Y}_i} \right) - (Y_i - \hat{Y}_i) \right]. \quad (1.48)$$

Taka postać dewiancji oznacza, że $D(\hat{\beta})$ zachowuje się jak suma kwadratów reszt $SKR = \sum_{i=1}^N (Y_i - \hat{Y}_i)^2$ w standardowej wielorakiej regresji liniowej.

Podobieństwo to jest następujące: Gdy dopasowywany model dokładnie przewiduje obserwowane wartości, tzn. $\hat{Y}_i = Y_i$, $i = 1, 2, \dots, N$ wtedy, jak SKR w analizie standardowej, tak $D(\hat{\beta})$ w analizie wiarygodnościowej jest równe zero [14, 17]. Z drugiej strony wartość $D(\hat{\beta})$ jest tym większa im większa jest różnica między wartościami obserwowanymi Y_i i przewidywanymi \hat{Y}_i przez oszacowany model.

Uwaga: W analizowanym modelu Y_i , $i = 1, 2, \dots, N$ są niezależnymi zmiennymi Poissona (np. zmiennymi częstości), natomiast wartości \hat{Y}_i są ich przewidywaniami. Gdy przewidywane wielkości mają rozsądną wartość⁷, np. $\hat{Y}_i > 3$, $i = 1, 2, \dots, N$, to statystyka (1.48) może być (dla dużej próby) przybliżona statystyką chi-kwadrat, o postaci:

$$\chi^2 = \sum_{i=1}^N \frac{(Y_i - \hat{Y}_i)^2}{\hat{Y}_i}, \quad (1.49)$$

z $N - k - 1$ stopniami swobody [14].

⁷Zauważmy, że statystyka (1.49) może mieć myląco dużą wartość gdy wielkości \hat{Y}_i są bardzo małe.

1.3 Zasada niezmienniczości ilorazu funkcji wiarygodności

Z powyższych rozważań wynika, że funkcja wiarygodności reprezentuje niepewność dla ustalonego parametru. Nie jest ona jednak funkcją gęstości prawdopodobieństwa rozkładu dla tego parametru. Pojęcie takie byłoby całkowicie obce statystyce klasycznej (nie włączając procesów stochastycznych). Inaczej ma się sprawa w tzw. statystyce Bayesowskiej.

Aby zrozumieć różnicę pomiędzy podejściem klasycznym i Bayesowskim rozważmy transformację parametru.

Rozważmy eksperyment, w którym dokonujemy jednokrotnego pomiaru zmiennej o rozkładzie dwumianowym (1.5). Funkcja wiarygodności ma więc postać $P(\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$, z parametrem $n = 12$. W pomiarze otrzymano $x = 9$. Testujemy model, dla którego $\theta = \theta_1 = 3/4$ wobec modelu z $\theta = \theta_2 = 3/10$. Stosunek wiarygodności wynosi:

$$\frac{P(\theta_1 = 3/4)}{P(\theta_2 = 3/10)} = \frac{\binom{n}{x} \theta_1^9 (1-\theta_1)^3}{\binom{n}{x} \theta_2^9 (1-\theta_2)^3} = 173.774 \quad (1.50)$$

Dokonajmy hiperbolicznego wzajemnie jednoznacznego przekształcenia parametru:

$$\psi = \frac{1}{\theta}. \quad (1.51)$$

Funkcja wiarygodności po transformacji parametru ma postać $\tilde{P}(\psi) = \binom{n}{x} (1/\psi)^x (1-1/\psi)^{n-x}$.

Wartości parametru ψ odpowiadające wartościom θ_1 and θ_2 wynoszą odpowiednio $\psi_1 = 4/3$ oraz $\psi_2 = 10/3$.

Łatwo sprawdzić, że transformacja (1.51) nie zmienia stosunku wiarygodności, tzn.:

$$\frac{\tilde{P}(\psi_1 = 4/3)}{\tilde{P}(\psi_2 = 10/3)} = \frac{P(\theta_1 = 3/4)}{P(\theta_2 = 3/10)} = 173.774. \quad (1.52)$$

Zatem widać, że stosunek wiarygodności jest niezmienniczy ze względu na wzajemnie jednoznaczną transformację parametru. Gdyby transformacja parametru była np. transformacją “logit” $\psi = \ln(\theta/(1-\theta))$ lub paraboliczną $\psi = \theta^2$, to sytuacja również nie uległaby zmianie. Również w ogólnym przypadku transformacji parametru własność *niezmienniczości stosunku wiarygodności* pozostaje słuszna. Oznacza to, że informacja zawarta w próbie jest niezmiennicza ze względu na wybór parametryzacji, tzn. powinniśmy być w takiej samej sytuacji niewiedzy niezależnie od tego jak zamodelujemy zjawisko, o ile różnica w modelowaniu sprowadza się jedynie do transformacji parametru. W omawianym przykładzie powinniśmy równie dobrze móc stosować parametr θ , jak $1/\theta$, θ^2 , czy $\ln(\theta/(1-\theta))$.

Natomiast sytuacja ma się zupełnie inaczej w przypadku Bayesowskiego podejścia do funkcji wiarygodności, w którym funkcja wiarygodności uwzględnia (Bayesowski) rozkład prawdopodobieństwa $f(\theta|x)$ parametru θ . Oznacza to, że Jakobian transformacji $\theta \rightarrow \psi$ parametru, modyfikując funkcję rozkładu parametru, zmienia również funkcję wiarygodności. Zmiana ta zależy od wartości parametru, różnie zmieniając licznik i mianownik w (1.50), co niszczy *intuicyjną* własność niezmienniczości ilorazu wiarygodności ze względu na transformację parametru [13].

Rozdział 2

Entropia względna i informacja Fishera

Niech $\hat{\Theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p)$ jest estymatorem MNW wektorowego parametru $\Theta = (\theta_1, \theta_2, \dots, \theta_p)$, otrzymanym po rozwiązaniu układu równań wiarygodności (1.3). Rozwiązanie to, jako maksymalizujące funkcję wiarygodności, musi wypełniać warunek ujemnej określoności formy kwadratowej

$$\sum_{i=1}^p \sum_{j=1}^p \frac{\partial^2 \ln P}{\partial \theta_i \partial \theta_j} \Big|_{\Theta=\hat{\Theta}} \Delta \theta_i \Delta \theta_j, \quad (2.1)$$

gdzie przyrosty $\Delta \theta_i, \Delta \theta_j$ nie zerują się jednocześnie. W przypadku skalarnym (tzn. jednego parametru θ) warunek ten oznacza ujemność drugiej pochodnej logarytmu funkcji wiarygodności w punkcie $\theta = \hat{\theta}$. Większa wartość $-\frac{\partial^2}{\partial \theta^2} \ln P(y|\theta) \Big|_{\theta=\hat{\theta}}$ oznacza węższe maksimum $\ln P$ w punkcie $\theta = \hat{\theta}$, tzn. większą krzywiznę funkcji wiarygodności, a co za tym idzie mniejszą niepewność określenia parametru θ .

2.1 Obserwowana i oczekiwana informacja Fishera

Ponieważ Θ może być w ogólności parametrem wektorowym, więc jako uogólnienie przypadku skalarnego zdefiniujemy $p \times p$ -wymiarową macierz:

$$\mathbf{iF}(\Theta) \equiv -\frac{\partial^2}{\partial \Theta \partial \Theta^T} \ln P(y|\Theta) \equiv -\left(\frac{\partial^2 \ln P}{\partial \theta_i \partial \theta_j} \right)_{p \times p}. \quad (2.2)$$

Jej wartość w punkcie $\Theta = \hat{\Theta}$ (oznaczoną po prostu jako \mathbf{iF}):

$$\mathbf{iF} = \mathbf{iF}(\hat{\Theta}), \quad (2.3)$$

nazywamy *obserwowaną informacją Fishera*¹. W teorii wiarygodności odgrywa ona kluczową rolę. Jako statystyka, czyli funkcja próby $y \equiv (y_1, y_2, \dots, y_N)$, jej realizacja w próbie jest macierzą liczbową. Z faktu wyznaczenia \mathbf{iF} w punkcie estymatora MNW oznacza, że jest ona dodatnio

¹We współczesnych wykładach \mathbf{iF} jest zazwyczaj zapisane jako:

$$\mathbf{iF} = \left(\frac{\partial \ln P(\Theta)}{\partial \theta_{n'}} \frac{\partial \ln P(\Theta)}{\partial \theta_n} \right). \quad (2.4)$$

Obie definicje, tzn. (2.4) oraz (2.2), prowadzą do tych samych konkluzji, o ile $\int d^N \mathbf{y} P(\Theta) (\partial \ln P(\Theta) / \partial \theta_n) = 0$. Zasadniczą zaletą zdefiniowania \mathbf{iF} poprzez (2.4) jest to, że bardzo naturalne staje się wtedy wprowadzenie tzw. α -koneksji na przestrzeni statystycznej \mathcal{S} [12].

określona, natomiast z (2.2) widać, że \mathbf{iF} jest macierzą symetryczną.

Jako przykład ilustrujący związek wielkości informacji Fishera (IF) z niepewnością oszacowania parametru, rozważmy realizację próby prostej y dla zmiennej Y posiadającej rozkład $N(\theta, \sigma^2)$. Załóżmy, że wariancja σ^2 jest znana a estymowanym parametrem jest wartość oczekiwana $\theta = E(Y)$. Logarytm funkcji wiarygodności ma postać:

$$\ln P(y|\Theta) = -\frac{N}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - \theta)^2 \quad (2.5)$$

skąd funkcja wynikowa (1.4) jest równa:

$$S(\theta) = \frac{\partial}{\partial \theta} \ln P(\vec{y}; \theta) = \frac{1}{\sigma^2} \sum_{i=1}^N (y_i - \theta) = \frac{N}{\sigma^2} (\bar{y} - \theta) . \quad (2.6)$$

Rozwiązując (jedno) równanie wiarygodności otrzymujemy postać estymatora parametru θ :

$$\hat{\theta} = \bar{y} \quad (2.7)$$

natomiast obserwowana informacja Fishera (2.3) jest równa:

$$\mathbf{iF}(\hat{\theta}) = \frac{N}{\sigma^2} . \quad (2.8)$$

Z (2.7) oraz z klasycznej analizy wiemy, że wariancja $\sigma^2(\hat{\theta}) = \sigma^2(\bar{y}) = \sigma^2/N$ zatem:

$$\mathbf{iF}(\hat{\theta}) = \frac{1}{\sigma^2(\hat{\theta})} . \quad (2.9)$$

Otrzymaliśmy więc ważny wniosek, że większa obserwowana IF parametru θ oznacza mniejszą wariancję jego estymatora $\hat{\theta}$. Zwróćmy uwagę, że powyższe równanie można zapisać w postaci:

$$\sigma^2(\hat{\theta}) \mathbf{iF}(\hat{\theta}) = 1 , \quad (2.10)$$

będące, jak się okaże, dolnym ograniczeniem nierówności Rao-Cramera.

Zdefiniujmy oczekiwaną informację Fishera następująco:

$$I_F(\Theta) \equiv E_{\Theta}(\mathbf{iF}(\Theta)) = \int_{\mathcal{Y}} dy P(y|\Theta) \mathbf{iF}(\Theta) . \quad (2.11)$$

Oznaczenie Θ w indeksie wartości oczekiwanej mówi, że Θ jest prawdziwą wartością parametru, przy której generowane są dane $y \equiv (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N)$, natomiast element różniczkowy dy oznacza $dy \equiv d^N \mathbf{y} = d\mathbf{y}_1 d\mathbf{y}_2 \dots d\mathbf{y}_N$.

Z powyższych rozważań wynika, że istnieją znaczące różnice pomiędzy oczekiwaną, a obserwowaną informacją Fishera [13, 18]. Oczekiwana informacja Fishera I_F ma sens jako funkcja dopuszczalnych (w jego przestrzeni) wartości Θ , natomiast (jak to wynika z MNW) obserwowana informacja Fishera, $\mathbf{iF}(\Theta)$, ma zasadniczo sens tylko w pobliżu $\hat{\Theta}$. Jako związana z obserwowaną wartością wiarygodności, \mathbf{iF} odnosi się do pojedynczego zestawu danych i zmienia się od próbkki do próbkki. Oznacza to, że należy o niej myśleć jako o pojedynczej realizacji statystyki $\mathbf{iF}(\hat{\Theta})$ w próbce a nie jako o funkcji parametru Θ . Natomiast oczekiwana informacja I_F jest średnią wartością dla wszystkich możliwych zestawów danych w całej przestrzeni próby generowanych przez prawdziwą wartość parametru. Zatem $I_F(\Theta)$ jest nie tyle użytecznym wskaźnikiem informacji

dla konkretnego zbioru danych, ile funkcją Θ mówiącą jak trudno jest estymować Θ , co oznacza, że parametr z większą I_F wymaga mniejszej próbki do osiągnięcia wymaganej precyzji.

Kontynuując rozważany przykład rozkładu normalnego z estymacją skalarne parametru θ , otrzymujemy po skorzystaniu z (2.8) oraz unormowania funkcji wiarygodności:

$$\int P(y|\theta)dy = 1, \quad (2.12)$$

wartość oczekiwanej IF dla parametru θ z rozkładu $N(\theta, \sigma^2)$:

$$I_F(\theta) = \int \mathfrak{IF}(\hat{\theta}) P(y|\theta)dy = \int \frac{N}{\sigma^2} P(y|\theta)dy = \frac{N}{\sigma^2}. \quad (2.13)$$

2.1.1 Wartość oczekiwana i wariancja funkcji wynikowej

Pokażmy, że średnia funkcji wynikowej, tzn. gradientu log funkcji wiarygodności jest równa zero,:

$$E_{\Theta}S(\Theta) = 0 \quad (2.14)$$

Istotnie,

$$\begin{aligned} E_{\Theta}S(\Theta) &= \int dy P(y|\Theta) S(\Theta) = \int dy P(y|\Theta) \left(\frac{\partial}{\partial \Theta} \ln P(y|\Theta) \right) \\ &= \int dy P(y|\Theta) \frac{\frac{\partial}{\partial \Theta} P(\Theta)}{P(\Theta)} = \int dy \frac{\partial}{\partial \Theta} P(y|\Theta) = \frac{\partial}{\partial \Theta} \int P(y|\Theta) dy = 0, \end{aligned} \quad (2.15)$$

gdzie skorzystano z interpretacji funkcji wiarygodności jako łącznego rozkładu prawdopodobieństwa i jej unormowania do jedności a zakres całkowania obejmuje całą przestrzeń próby y . W przypadku gdy pierwotna zmienna y jest dyskretna, powyższy dowód przebiega podobnie.

Warunek dla wyciągnięcia różniczkowania po parametrze przed znak całki oznacza spełnienie żądania aby rozkład $P(y|\Theta)$ był wystarczająco gładki jako funkcja Θ [13]. Oznacza to, że Θ nie może być brzegową wartością, tzn. istnieje taka funkcja $g(y)$, dla której $\int g(y) dy$ jest skończona, że $|\partial P(y|\Theta)/\partial \Theta| \leq g(y)$ dla $\partial P/\partial \Theta$ traktowanej jako funkcja y w sąsiedztwie prawdziwej wartości parametru Θ [13]. Zauważmy również, że przy odpowiednim warunku regularności:

$$\int \frac{\partial^2}{\partial \Theta^2} P(y|\Theta) dy = \frac{\partial^2}{\partial \Theta^2} \int P(y|\Theta) dy = 0 \quad (2.16)$$

Zachodzi **Twierdzenie**: Zakładając warunek regularności pozwalający na wyciągnięcie różniczkowania po parametrze przed znak całki, można pokazać, że wariancja funkcji wynikowej jest równa oczekiwanej IF:

$$\sigma_{\Theta}^2 S(\Theta) = I_F(\Theta). \quad (2.17)$$

Zauważmy, że macierz informacji $I_F(\Theta)$ jest $p \times p$ macierzą kowariancji, co oznacza, że jest ona nieujemnie określana².

²Dla dowolnego wektora $\mathbf{a} = (a_1, \dots, a_p)^T \in \mathbf{R}^p$ oraz macierzy kowariancji $C = E \left((Z - E(Z))(Z - E(Z))^T \right)$, gdzie Z jest p -wymiarowym wektorem losowym, zachodzi $\mathbf{a}^T E \left[(Z - E(Z))(Z - E(Z))^T \right] \mathbf{a} = \mathbf{a}^T E(WW^T) \mathbf{a} = E[(\mathbf{a}^T W)^2] = \sigma^2(\mathbf{a}^T W) \geq 0$, gdzie $W = Z - E(Z)$ i $E(W) = 0$, tzn. macierz C jest nieujemnie określona.

Dowód twierdzenia (2.17). Korzystając z (2.14) otrzymujemy:

$$\begin{aligned}
\sigma^2_{\Theta} S(\Theta) &= \int (S(\Theta) - E_{\Theta} S(\Theta))^2 P(y|\Theta) dy \\
&= \int (S(\Theta))^2 P(y|\Theta) dy = \int \left(\frac{\partial}{\partial \Theta} \ln P(y|\Theta) \right)^2 P(y|\Theta) dy \\
&= \int \left[\left(\frac{\partial}{\partial \Theta} P(y|\Theta) \right) / P(y|\Theta) \right]^2 P(y|\Theta) dy = \int \left[\frac{\partial}{\partial \Theta} P(y|\Theta) \right]^2 / P(y|\Theta) dy .
\end{aligned} \tag{2.18}$$

Natomiast korzystając z (2.16) otrzymujemy:

$$\begin{aligned}
I_F(\Theta) &= E_{\Theta} \mathbf{iF}(\Theta) = \int dy P(y|\Theta) \mathbf{iF}(\Theta) = - \int dy P(y|\Theta) \left(\frac{\partial^2}{\partial \Theta^2} \ln P(y|\Theta) \right) \\
&= - \int dy P(y|\Theta) \frac{\partial}{\partial \Theta} \left[\left(\frac{\partial}{\partial \Theta} P(y|\Theta) \right) / P(y|\Theta) \right] \\
&= \int dy P(y|\Theta) \frac{\left[- \left(\frac{\partial^2}{\partial \Theta^2} P(y|\Theta) \right) P(y|\Theta) + \left(\frac{\partial}{\partial \Theta} P(y|\Theta) \right)^2 \right]}{(P(y|\Theta))^2} \\
&= \int dy \left[\frac{\partial}{\partial \Theta} P(y|\Theta) \right]^2 / P(y|\Theta)
\end{aligned} \tag{2.19}$$

co porównując z (2.18) kończy dowód twierdzenia (2.17).

Przy okazji z (2.19) widać, że przy spełnieniu wspomnianych własności regularności, zachodzi:

$$I_F(\Theta) = - \int dy P(y|\Theta) \left(\frac{\partial^2}{\partial \Theta^2} \ln P(y|\Theta) \right) = \int dy P(y|\Theta) \left[\frac{\partial}{\partial \Theta} \ln P(y|\Theta) \right]^2 . \tag{2.20}$$

Powyżej otrzymane twierdzenia, (2.14) oraz (2.15), są słuszne dla ogólnego przypadku wektorowego, więc i z jednym parametrem tzn. skalarne.

W końcu z (2.20) otrzymujemy dla dowolnego N -wymiarowego wektora $v = (v_1, v_2, \dots, v_p)^T$:

$$v^T I_F(\Theta) v = \sum_{i=1}^p \sum_{j=1}^p v_i g_{ij}(\Theta) v_j = \int_{\mathcal{Y}} dy P(y|\Theta) \left\{ \sum_{i=1}^p (v_i \partial_i \ln P(\Theta)) \sum_{j=1}^p (\partial_j \ln P(\Theta) v_j) \right\} \geq 0 , \tag{2.21}$$

co oznacza, że $I_F(\Theta)$ jest dodatnio półokreślona [13], jak to zaznaczyliśmy poniżej (2.17). Jednak w teorii pola interesuje nas zaostrenie warunku (2.21) do własności dodatniej określoności. W Rozdziale 2.2 zwrócimy uwagę na fakt, że macierz informacyjna $(g_{ij}) = I_F(\Theta)$ określa metrykę na przestrzeni statystycznej S . Jak to wynika z (2.17), w teorii pola z ciągłymi rozkładami (co pociąga za sobą ciągłość rozkładu funkcji wynikowej na całej przestrzeni próby), macierz informacyjna $I_F(\Theta)$ jest określona dodatnio.

Zatem: w teorii pola z ciągłymi rozkładami metryka Fishera g_{ij} jest określona dodatnio.

2.2 Wstęp do geometrii różniczkowej na przestrzeni statystycznej i α -koneksja

Niech $\mathcal{S} = \{p_\vartheta \equiv p(\mathbf{y}|\vartheta), \vartheta \equiv (\vartheta_n)_{n=1}^k \in V_\Theta\}$ jest k -wymiarowym modelem statystycznym, tzn. rodziną rozkładów prawdopodobieństwa parametryzowaną przez k (niełosowych) zmiennych o wartościach rzeczywistych $(\vartheta_n)_{n=1}^k$ na \mathcal{Y} (gdzie np. $\mathcal{Y} = \mathbb{R}^4$). Mówimy, że \mathcal{S} jest k -wymiarową przestrzenią statystyczną. Zmienna losowa Y zdefiniowana na \mathcal{Y} przyjmuje wartości \mathbf{y} . Wprowadźmy oznaczenie $\ell_\vartheta \equiv \ln P(\vartheta)$ oraz $\partial_i \equiv \partial/\partial\vartheta^i$. Dla każdego punktu ϑ , $k \times k$ -wymiarowa macierz $(g_{ij}(\vartheta))$ o elementach:

$$g_{ij} := E_\vartheta(\partial_i \ell_\vartheta \partial_j \ell_\vartheta) = \int_{\mathcal{Y}} d\mathbf{y} p(\mathbf{y}|\vartheta) \partial_i \ell(\mathbf{y}|\vartheta) \partial_j \ell(\mathbf{y}|\vartheta), \quad (2.22)$$

jest nazywana macierzą informacyjną Fishera na \mathcal{S} w punkcie ϑ [12]. Wielkość $E_\vartheta(\cdot)$ oznacza tutaj wartość oczekiwaną, a całkowanie przebiega po całej przestrzeni \mathcal{Y} .

Przy założeniu spełnienia warunków regularności (porównaj (2.20)), macierz $(g_{ij}(\vartheta))$ może zostać zapisana następująco:

$$g_{ij} = -E_\vartheta(\partial_i \partial_j \ell_\vartheta) = - \int_{\mathcal{Y}} d\mathbf{y} p(\mathbf{y}|\vartheta) \partial_i \partial_j \ln P(\mathbf{y}|\vartheta). \quad (2.23)$$

Wprowadźmy iloczyn wewnętrzny nazywamy metryką Fishera $\langle \cdot, \cdot \rangle$ na \mathcal{S} , definiując go w układzie współrzędnych $(\vartheta_i)_{i=1}^k$ poprzez związek $\langle \partial_i, \partial_j \rangle = g_{ij}$. Metryka ta jest typu Riemannowskiego. Wprowadźmy też koneksję afiniczną, tzn. k^3 funkcji $\Gamma_{ij,k}^{(\alpha)} : \vartheta \rightarrow (\Gamma_{ij,k}^{(\alpha)})_\vartheta$, określonych na \mathcal{S} , które każdemu punktowi ϑ przyporządkowują k^3 następujących współczynników:

$$(\Gamma_{ij,k}^{(\alpha)})_\vartheta = E_\vartheta \left[\left(\partial_i \partial_j \ell_\vartheta + \frac{1-\alpha}{2} \partial_i \ell_\vartheta \partial_j \ell_\vartheta \right) \partial_k \ell_\vartheta \right]. \quad (2.24)$$

W końcu wprowadźmy α -koneksję $\nabla^{(\alpha)}$ na \mathcal{S} zdefiniowaną następująco [12]:

$$\langle \nabla_{\partial_i}^{(\alpha)} \partial_j, \partial_k \rangle = \Gamma_{ij,k}^{(\alpha)}. \quad (2.25)$$

Koneksja α jest symetryczna. Można pokazać, że: $\nabla^{(\alpha)} = \frac{1+\alpha}{2} \nabla^{(1)} + \frac{1-\alpha}{2} \nabla^{(-1)}$.

Z (2.22) oraz (2.24) można zauważyć, że $\partial_k g_{ij} = \Gamma_{ki,j}^{(0)} + \Gamma_{kj,i}^{(0)}$, co oznacza, że 0-koneksja jest metryczna ze względu na metrykę Fishera [12]. Jednak w ogólności dla $\alpha \neq 0$, α -koneksja ∇^α nie jest metryczna.

Szczególnym przykładem jest tzw. *rodzina eksponentialna* rozkładów, która ma postać:

$$p_\vartheta \equiv p(\mathbf{y}; \vartheta) = \exp \left[C(\mathbf{y}) + \sum_{i=1}^k \vartheta^i F_i(\mathbf{y}) \psi(\vartheta) \right], \quad (2.26)$$

gdzie ϑ^i są tzw. parametrami kanonicznymi. Jest ona $\alpha = 1$ - płaska, co oznacza, że $\Gamma_{ij,k}^{(1)} = 0$, [12]. Wyróżnia ją fakt osiągnięcia dolnego ograniczenia nierówności Rao-Cramera [19]. Przykładem rozkładu z eksponentialnej rodziny jest *rozkład normalny*, dla którego $C(\mathbf{y}) = 0$, $F_1 = \mathbf{y}$, $F_2 = \mathbf{y}^2$, $\vartheta^1 = \frac{\mu}{\sigma^2}$, $\vartheta^2 = -\frac{1}{2\sigma^2}$.

Inną klasą jest tzw. *rodzina mieszanych* rozkładów $p_\vartheta \equiv p(\mathbf{y}; \vartheta) = C(\mathbf{y}) + \sum \vartheta^i F_i(\mathbf{y}) \psi(\vartheta)$, gdzie ϑ^i są tzw. parametrami mieszanymi. Można pokazać, że przestrzeń statystyczna rodziny rozkładów mieszanych jest $\alpha = -1$ - płaska [12].

Modele $\alpha = 1$ bądź -1 płaskie odgrywają szczególną rolę w modelowaniu statystycznym [13].

2.3 Dolne ograniczenie Rao-Cramera

Zgodnie z powyższymi uwagami, estymatory MNW mają asymptotycznie optymalne własności. Poniższy rozdział poświęćmy efektywności estymatora parametru. Estymator nazywamy efektywnym (lub najefektywniejszym) gdy ma najmniejszą z możliwych wariancji, jaką możemy uzyskać w procesie estymacji parametru. Dolnego ograniczenia na wariancję estymatora, czyli wariancję estymatora efektywnego, podaje poniższe twierdzenie Rao-Cramera. Jego sednem jest stwierdzenie, że osiągnięcie przez estymator dolnej granicy wariancji podanej przez twierdzenie Rao-Cramera oznacza, że w klasie estymatorów nieobciążonych, które spełniają warunek regularności (tzn. mają funkcję rozkładu prawdopodobieństwa nie posiadającą punktów nieciągłości zależnych od estymowanego parametru Θ), nie znajdziemy estymatora z mniejszą wariancją.

Twierdzenie Rao-Cramera (TRC): Niech $T(y)$ będzie nieobciążonym estymatorem funkcji skalarnego parametru $g(\theta)$, tzn.:

$$E_{\theta}T(y) = g(\theta) \quad (2.27)$$

oraz niech $I_F(\theta)$ będzie informacją Fishera dla parametru θ wyznaczoną na podstawie próby y . Zakładając warunki regularności, otrzymujemy:

$$\sigma^2_{\theta}T(y) \geq \frac{[g'(\theta)]^2}{I_F(\theta)}. \quad (2.28)$$

W szczególnym przypadku gdy $g(\theta) = \theta$, wtedy z (2.28) otrzymujemy następującą postać nierówności Rao-Cramera:

$$\sigma^2_{\theta}T(y) \geq \frac{1}{I_F(\theta)}. \quad (2.29)$$

Wielkość:

$$\frac{[g'(\theta)]^2}{I_F(\theta)} \quad \text{lub} \quad \frac{1}{I_F(\theta)} \quad \text{dla} \quad g(\theta) = \theta \quad (2.30)$$

nazywana jest dolnym ograniczeniem Rao-Cramera (DORC)³. Przypomnijmy, że ponieważ statystyka $T(y)$ jest estymatorem parametru θ , więc nie zależy ona od tego parametru.

Uwaga: W przypadku gdy funkcja rozkładu zmiennej Y traktowana jako funkcja (estymowanego) parametru θ ma dla pewnych wartości tego parametru punkty nieciągłości, wtedy wariancja estymatora parametru θ występująca po lewej stronie (2.29) może okazać się mniejsza niż wartość po stronie prawej. Tego typu sytuacji nie będziemy jednak rozważali. Przeciwnie, zakładamy, że funkcja rozkładu jest wystarczająco gładka, tzn. posiada rozwinięcie w szereg Taylora wystarczająco wysokiego rzędu (tzn. drugiego lub więcej), w każdym punkcie (pod)przestrzeni statystycznej \mathcal{S} . O ile nie podamy inaczej, to w dalszym ciągu założymy, że rząd rozwinięcia Taylora jest wystarczająco wysoki [10], tzn. albo jest spełniony warunek analityczności funkcji rozkładu we wszystkich estymowanych parametrach (wtedy \mathcal{S} jest rozmaitością), albo funkcja rozkładu określona w otoczeniu punktu $p \in \mathcal{S}$ posiada jet $J_p^r(\mathcal{S}, \mathbb{R})$ wystarczająco wysokiego, choć skończonego rzędu r [20]. Istotną sprawą jest fakt, że zbiór wszystkich r -jetów funkcji w punkcie p tworzy skończenie wymiarową przestrzeń wektorową, natomiast ich suma $J^r(\mathcal{S}, \mathbb{R}) = \bigcup_{p \in \mathcal{S}} J_p^r(\mathcal{S}, \mathbb{R})$ jest wiązką włóknistą nad \mathcal{S} .

³Z angielskiego *Cramér-Rao lower bound* (CRLB).

2.3.1 Dowód TRC (wersja dowodu dla przypadku skalarne)

Współczynnik kowariancji liniowej Pearsona dla dwóch zmiennych losowych $S(y)$ i $T(y)$ zdefiniowany jest następująco:

$$\rho(S, T) = \frac{\text{cov}(S, T)}{\sqrt{\sigma^2(S)}\sqrt{\sigma^2(T)}}. \quad (2.31)$$

Z klasycznej analizy statystycznej wiemy, że $\rho(S, T) \in [-1, 1]$, stąd z (2.31) otrzymujemy:

$$\sigma^2(T) \geq \frac{|\text{cov}(S, T)|^2}{\sigma^2(S)}. \quad (2.32)$$

Równość występuje gdy współczynnik kowariancji liniowej Pearsona jest równy 1, co zachodzi gdy zmienne S i T są idealnie skorelowane.

Niech teraz zmienna losowa S będzie statystykę wynikową $S(\theta) \equiv S(y|\theta)$. Pokażmy, że $\text{cov}(S(\theta), T) = g'(\theta)$. Istotnie, ponieważ:

$$\begin{aligned} \text{cov}(S(\theta), T(y)) &= E(S(\theta)T(y)) = \int S(\theta)T(y)P(y|\theta)dy = \int \frac{\partial}{\partial \theta} \frac{P(y|\theta)}{P(y|\theta)} T(y)P(y|\theta)dy \\ &= \int \frac{\partial}{\partial \theta} P(y|\theta)T(y)dy = \frac{\partial}{\partial \theta} \int P(y|\theta)T(y)dy = \frac{\partial}{\partial \theta} E_{\theta}T(y) = g'(\theta), \end{aligned} \quad (2.33)$$

zatem ponieważ zgodnie z (2.17) zachodzi, $\sigma^2_{\theta}S(\theta) = I_F(\theta)$, więc wstawiając (2.33) do (2.32) otrzymujemy (2.28), co kończy dowód TRC.

Jako przykład rozważmy średnią arytmetyczną $\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$, która, zakładając jedynie identyczne rozkłady zmiennych y_i próby, jest (dla dowolnego rozkładu $f(y)$ zmiennej y) nieobciążonym estymatorem wartości oczekiwanej $E(y) \equiv \int yf(y)dy$, tzn.

$$E(\bar{y}) = \int \bar{y}P(y|\theta)dy = E(y) \quad (2.34)$$

Z bezpośredniego rachunku dla próby prostej otrzymujemy również:

$$\sigma^2(\bar{y}) = \int (\bar{y} - E(\bar{y}))^2 P(y|\theta)dy = \frac{\sigma^2}{N}, \quad (2.35)$$

gdzie σ^2 jest wariancją $\sigma^2(y)$ zmiennej y .

Niech teraz zmienna pierwotna y ma rozkład normalny $N(\theta, \sigma^2)$. Łatwo sprawdzić, że $E(y) = \theta$, zatem $\hat{\theta} = \bar{y}$. Interesuje nas estymacja skalarne parametru θ w próbie prostej y . Jak już wiemy, średnia \bar{y} jest też estymatorem MNW parametru θ wyznaczonym z równania wiarygodności (1.3). W przypadku rozkładu normalnego warunek można sprawdzić bezpośrednio. Istotnie, korzystając z (2.6) oraz (2.13), otrzymujemy:

$$\sigma^2 S(\theta) = \sigma^2 \left(\frac{N}{\sigma^2} (\bar{y} - \theta) \right) = \left(\frac{N}{\sigma^2} \right)^2 \frac{\sigma^2}{N} = \frac{N}{\sigma^2} = I_F(\theta). \quad (2.36)$$

Widzimy więc, że dla zmiennych o rozkładzie normalnym wariancja średniej arytmetycznej $\bar{y} \equiv T(y)$ jest zgodnie z (2.35) równa σ^2/N . Jest to dolne ograniczenie Rao-Cramera (2.29), co oznacza, że średnia arytmetyczna jest efektywnym estymatorem wartości oczekiwanej.

2.3.2 Przypadek wieloparametrowy DORC

Gdy dokonujemy równoczesnej estymacji $p > 1$ parametrów, wtedy funkcja wynikowa $S(\Theta)$ jest wektorem kolumnowym (1.4), natomiast $\mathbf{iF}(\hat{\Theta})$ czyli obserwowana IF w punkcie $\hat{\Theta}$, oraz wartością oczekiwaną z $\mathbf{iF}(\Theta)$ czyli I_F (por.(2.11)) są $p \times p$ wymiarowymi macierzami.

Rozważmy dwuparametrowy ($p = 2$) przykład rozkładu normalnego. Niech $\mathbf{y}_1, \dots, \mathbf{y}_N$ jest realizacją próby prostej dla zmiennej Y o rozkładzie normalnym $N(\theta, \sigma)$. Wektor parametrów przyjmujemy jako $\Theta \equiv (\theta_i)_{i=1}^2 = (\mu, \sigma)$. Funkcja wiarygodności próby ma wtedy postać:

$$\ln P(y|\Theta) = -N \ln(\sqrt{2\pi} \sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - \mu)^2. \quad (2.37)$$

Funkcja wynikowa z nią związana jest równa:

$$S(\Theta) = \begin{pmatrix} \frac{\partial}{\partial \mu} \ln P(y|\Theta) \\ \frac{\partial}{\partial \sigma} \ln P(y|\Theta) \end{pmatrix} = \begin{pmatrix} \frac{N}{\sigma^2} (\bar{y} - \mu) \\ -\frac{N}{\sigma} + \frac{\sum_{i=1}^N (y_i - \mu)^2}{\sigma^3} \end{pmatrix}, \quad (2.38)$$

gdzie $\bar{y} = \frac{1}{N} \sum_{i=1}^N \mathbf{y}_i$ jest średnią arytmetyczną w próbie. Zatem postacie estymatorów MNW, tzn. parametru wartości oczekiwanej μ oraz odchylenia standardowego σ zmiennej Y , otrzymujemy rozwiązując układ równań wiarygodności:

$$S(\Theta)|_{\Theta=\hat{\Theta}} = \begin{pmatrix} \frac{N}{\sigma^2} (\bar{y} - \mu) \\ -\frac{N}{\sigma} + \frac{\sum_{i=1}^N (y_i - \mu)^2}{\sigma^3} \end{pmatrix} |_{\Theta=\hat{\Theta}} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad (2.39)$$

gdzie $\hat{\Theta} = (\hat{\mu}, \hat{\sigma})$ i rozwiązanie to ma postać:

$$\hat{\mu} = \bar{y} \quad \text{oraz} \quad \hat{\sigma} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2}. \quad (2.40)$$

Obserwowana informacja Fishera (2.2) w punkcie $\hat{\Theta}$ wynosi więc:

$$\mathbf{iF}(\hat{\Theta}) = \begin{pmatrix} \frac{N}{\sigma^2} & \frac{2N}{\sigma^3} (\bar{y} - \mu) \\ \frac{2N}{\sigma^3} (\bar{y} - \mu) & -\frac{N}{\sigma^2} + \frac{3}{\sigma^4} \sum_{i=1}^N (y_i - \mu)^2 \end{pmatrix} |_{\Theta=\hat{\Theta}} = \begin{pmatrix} \frac{N}{\hat{\sigma}^2} & 0 \\ 0 & \frac{2N}{\hat{\sigma}^2} \end{pmatrix}, \quad (2.41)$$

W końcu oczekiwana informacja Fishera jest równa:

$$I_F(\Theta) = E_{\Theta} \mathbf{iF}(\Theta) = \begin{pmatrix} \frac{N}{\sigma^2} & 0 \\ 0 & \frac{2N}{\sigma^2} \end{pmatrix}. \quad (2.42)$$

Poprzednio, dowodząc (2.14) okazało się, że przy założeniu warunku regularności, wartość oczekiwana na przestrzeni próby z funkcji wynikowej zeruje się, tzn. $E_{\Theta} S(\Theta) = 0$ i rezultat ten jest słuszny dla ogólnego przypadku wektorowego. Okazało się też (2.17), że wariancja funkcji wynikowej jest równa IF, tzn. $\sigma^2_{\Theta} S(\Theta) = I_F(\Theta)$, co ma znaczenie dla dodatniej określoności metryki Fishera g_{ij} w teorii pola. Przedstawione dowody dla własności (2.14) oraz (2.17) były ogólne. Jako ćwiczenie warto pokazać, że zachodzą one dla (2.38) oraz (2.42) w powyższym przykładzie estymacji parametrów rozkładu normalnego.

Uwaga: Widząc powyższy rozkład jako $N(\theta, \sigma^2)$ i w konsekwencji przyjmując wektor parametrów jako (μ, σ^2) , oraz przeprowadzając analogiczny rachunek jak powyżej, można pokazać, że estymator wartości oczekiwanej ma postać $\hat{\mu} = \bar{y}$, czyli tak jak w (2.40), natomiast estymator

wariancji wynosi $\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2$. W jego mianowniku występuje czynnik N a nie $N - 1$ jak to ma miejsce w przypadku nieobciążonego estymatora wariancji. Estymator $\hat{\sigma}^2$ nie jest też efektywny (tzn. nie posiada najmniejszej z możliwych wariancji). Jednak własności te posiada on asymptotycznie (dla $N \rightarrow \infty$), co jest charakterystyczne dla wszystkich estymatorów MNW [11].

Uwaga: Poniżej pokażemy, że włączenie do analizy dodatkowych parametrów ma (na ogół) wpływ na wartość IF dla interesującego nas, wyróżnionego parametru. Sytuacja ta jest analogiczna do problemu inflacji wariancji estymatora parametru w analizie częstotliwościowej [14]. Poniżej przedstawiona zostanie odnosząca się do tego problemu wieloparametrowa wersja twierdzenia o dolnym ograniczeniu w nierówności Rao-Cramera (DORC).

Twierdzenie RC (wieloparametrowe): Niech $T(y)$ będzie funkcją *skalarną* z wartością oczekiwaną:

$$E_{\Theta} T(y) = g(\Theta) \in \mathbf{R}, \quad (2.43)$$

oraz $I_F(\Theta)$ niech będzie oczekiwaną informacją Fishera (2.11) dla $\Theta \equiv (\theta_i)_{i=1}^p$ wyznaczaną na przestrzeni próby y . Zachodzi wtedy nierówność:

$$\sigma^2_{\Theta}(T) \geq \mathbf{a}^T I_F^{-1}(\Theta) \mathbf{a} \quad (2.44)$$

gdzie I_F^{-1} jest macierzą odwrotną do macierzy informacyjnej Fishera I_F oraz

$$\mathbf{a} = \frac{\partial}{\partial \Theta} g(\Theta), \quad (2.45)$$

jest p -wymiarowym wektorem.

2.3.2.1 Przykład wektorowego DORC

Jako ilustrację powyższego wieloparametrowego Twierdzenia RC przedstawimy przykład, przyjmując szczególną postać skalarnej funkcji $g(\Theta)$, o której zakładamy, że jest liniową funkcją składowych θ_i wektora parametrów [13]:

$$g(\Theta) = \mathbf{a}^T \Theta = \sum_{i=1}^p a_i \theta_i, \quad (2.46)$$

gdzie \mathbf{a} jest pewnym znanym wektorem o stałych składowych a_i , które nie zależą od składowych wektora Θ . Załóżmy chwilowo, że $\mathbf{a}^T = (1, 0, \dots, 0)$, tzn. jedynie $a_1 \neq 0$. Wtedy z (2.46) otrzymujemy $g(\Theta) = \theta_1$, natomiast (2.44) w Twierdzeniu RC, $\sigma^2_{\Theta}(T) \geq \mathbf{a}^T I_F^{-1}(\Theta) \mathbf{a}$, przyjmuje dla rozważanego nieobciążonego estymatora T postać:

$$\sigma^2(T) \geq [I_F^{-1}(\Theta)]_{11} =: I_F^{11}(\Theta), \quad (2.47)$$

gdzie $I_F^{11}(\Theta)$ oznacza element (1,1) macierzy $I_F^{-1}(\Theta)$. Prawa strona nierówności (2.47) podaje dolne ograniczenie wariancji estymatora T , pod warunkiem, że θ_1 jest wyróżnionym parametrem a wartości pozostałych parametrów nie są znane. Oznaczmy wewnętrzną strukturę $p \times p$ -wymiarowych macierzy $I_F(\Theta)$ oraz $I_F(\Theta)^{-1}$ następująco:

$$I_F(\Theta) = \begin{pmatrix} I_{F11} & I_{F12} \\ I_{F21} & I_{F22} \end{pmatrix} \quad (2.48)$$

oraz

$$I_F^{-1}(\Theta) = \begin{pmatrix} I_F^{11} & I_F^{12} \\ I_F^{21} & I_F^{22} \end{pmatrix} \quad (2.49)$$

gdzie I_{F11} oraz $I_F^{11} = [I_F^{-1}(\Theta)]_{11}$ (zgodnie z oznaczeniem wprowadzonym w (2.47)) są liczbami, I_{F22} , I_F^{22} są $(p-1) \times (p-1)$ -wymiarowymi macierzami, natomiast $(I_{F12})_{1 \times (p-1)}$, $(I_{F21})_{(p-1) \times 1}$, $(I_F^{12})_{1 \times (p-1)}$, $(I_F^{21})_{(p-1) \times 1}$ odpowiednimi wierszowymi bądź kolumnowymi wektorami o wymiarze $(p-1)$.

Rozważmy parametr θ_1 . Jego informacja Fishera jest równa $I_{F11} = I_{F11}(\theta_1)$. Nie oznacza to jednak, że $\sigma^2(T)$ oraz I_{F11} są z sobą automatycznie powiązane nierównością $\sigma^2(T) \geq 1/I_{F11}$, która jest słuszna na podstawie Twierdzenia RC (2.29). Udowodniliśmy ją bowiem tylko dla przypadku parametru skalarnego (tzn. gdy tylko jeden parametr jest estymowany a reszta parametrów jest znana).

Określmy relację pomiędzy $(I_F^{11})^{-1}$ oraz I_F^{11} . Oczywiście zachodzi:

$$I_F(\Theta) \cdot I_F^{-1}(\Theta) = \begin{pmatrix} I_{F11} & I_{F12} \\ I_{F21} & I_{F22} \end{pmatrix} \begin{pmatrix} I_F^{11} & I_F^{12} \\ I_F^{21} & I_F^{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (2.50)$$

zatem:

$$\begin{aligned} (I_{F11})_{1 \times 1} (I_F^{11})_{1 \times 1} + (I_{F12})_{1 \times (p-1)} (I_F^{21})_{(p-1) \times 1} &= 1 \\ \Rightarrow (I_F^{11})^{-1} &= I_{F11} + I_{F12} I_F^{21} (I_F^{11})^{-1}, \end{aligned} \quad (2.51)$$

$$\begin{aligned} (I_{F21})_{(p-1) \times 1} (I_F^{11})_{1 \times 1} + (I_{F22})_{(p-1) \times (p-1)} (I_F^{21})_{(p-1) \times 1} &= (0)_{(p-1) \times 1} \\ \Rightarrow I_F^{21} &= -(I_{F22})^{-1} I_{F21} I_F^{11} \end{aligned} \quad (2.52)$$

skąd otrzymujemy:

$$(I_F^{11})^{-1} = I_{F11} - I_{F12} (I_{F22})^{-1} I_{F21}. \quad (2.53)$$

Ponieważ I_{F22} jest macierzą informacyjną (dla parametrów $\theta_2, \theta_3, \dots, \theta_{(p-1)}$), jest więc zgodnie z rozważaniami poniżej (2.21) symetryczna i nieujemnie określona. Symetryczna i nieujemnie określona jest zatem $(I_{F22})^{-1}$. Ponieważ z symetrii macierzy I wynika $I_{F12} = (I_{F21})^T$, zatem ostatecznie forma kwadratowa $I_{F12} (I_{F22})^{-1} I_{F21} \geq 0$, stąd z (2.53) otrzymujemy:

$$(I^{F11})^{-1} \leq I_{F11} \Rightarrow I^{F11} \geq \frac{1}{I_{F11}}, \quad (2.54)$$

co zgodnie z (2.47) oznacza, że:

$$\sigma^2(T) \geq I_F^{11} \geq \frac{1}{I_{F11}}. \quad (2.55)$$

Zatem otrzymujemy, że I_F^{11} daje silniejsze ograniczenie niż $(I_{F11})^{-1}$. Tzn. w przypadku estymacji wieloparametrowej należy zastosować związek $\sigma^2(T) \geq I_F^{11}$, (2.47), gdyż to właśnie on jest właściwy na podstawie wieloparametrowego Twierdzenia RC. Zastosowanie $\sigma^2(T) \geq 1/I_{F11}$, tak jak byśmy mieli do czynienia z przypadkiem skalarnym, błędnie zaniża wartość dolnego ograniczenia na $\sigma^2(T)$.

Uwaga: W “statycznie” ukierunkowanej analizie statystycznej wielkość $(I_F^{11})^{-1}$ jest interpretowana jako informacja Fishera dla θ_1 – **jednak w treści skryptu odstępimy od tej nazwy**. A mianowicie w analizie ukierunkowanej na estymację “dynamiczną”, tzn. generującą równania różniczkowe dla rozkładów, okazuje się za bardziej użyteczne nazwać $(I_F^{11})^{-1}$ po prostu dolnym ograniczeniem RC na wariancję estymatora parametru θ_1 w sytuacji gdy pozostałe parametry są nieznane (tzn. trzeba je estymować z próby równocześnie z θ_1). *Natomiast I_{F11} będziemy nazywali, zgodnie z tym jak to uczyniliśmy, informacją Fishera parametru θ_1 , i to niezależnie od tego czy inne parametry są równocześnie estymowane, czy też nie.*

Uwaga: Należy pamiętać, że estymując parametr θ_1 należy być świadomy faktu występowania równoczesnej estymacji innych parametrów, gdyż wstawienie wartości I_{F11} do nierówności RC może w przypadku estymacji wieloparametrowej doprowadzić do zaniżenia wartości dolnego ograniczenia wariancji tego parametru.

Uwaga: Istnieje jednak pewien wyjątek spowodowany dokładnym zerowaniem się I_{F12} dla dowolnego N . Wtedy z (2.53) wynika, że wzrost wariancji estymatora parametru związany z dodaniem nowych parametrów o nieznanymi wartościami byłby równy zero. Tak też było w rozważanym wcześniej przykładzie rozkładu normalnego $N(\mu, \sigma)$ (porównaj (2.42) z (2.13)). Podobnie, taki szczególny przypadek zachodzi gdy wieloparametrowym rozkładem prawdopodobieństwa jest wiarygodność N -wymiarowej próby $P(\theta_1, \theta_2, \dots, \theta_N) = \prod_{n=1}^N p_{\theta_n}$, gdzie każdy estymowany parametr θ_n określa tylko jeden punktowy rozkład p_{θ_n} . Wtedy macierz informacyjna Fishera I_F jest diagonalna i zachodzi $I_{Fnn} = (I_F^{nn})^{-1}$, a w miejsce (2.55) otrzymujemy dla każdego parametru θ_n :

$$\sigma^2(T_n) \geq I_F^{nn} = \frac{1}{I_{Fnn}}, \quad n = 1, 2, \dots, N \quad \text{gdy } I_F \text{ jest diagonalne,} \quad (2.56)$$

jako szczególny przypadek nierówności RC, gdzie T_n jest estymatorem parametru θ_n .

2.4 Entropia informacyjna

2.4.1 Entropia Shannona i jej własności

W rozdziale tym omówimy pojęcie, które podaje globalną charakterystykę pojedynczego rozkładu prawdopodobieństwa, tzn. entropię Shannona. Dokładniejsze omówienie własności entropii Shannona można znaleźć w [21].

Niech $P(\omega)$ będzie rozkładem prawdopodobieństwa⁴ określonym na przestrzeni próby Ω , gdzie ω jest punktem w Ω . Jeśli przestrzeń próby Ω jest dyskretna to informacyjna entropia Shannona jest zdefiniowana następująco:

$$S(P) = -k \sum_{\omega} P(\omega) \ln P(\omega), \quad (2.57)$$

gdzie k jest liczbą dodatnią⁵, którą przyjmujemy dalej jako równą 1.

Niech Ω jest rozpięta przez skończoną liczbę N elementów będących możliwymi wynikami doświadczenia w rezultacie, którego otrzymujemy wartości x_i , $i = 1, 2, \dots, N$, zmiennej losowej X . Rozkład prawdopodobieństwa P jest wtedy reprezentowany przez wektor $\vec{p} = (p_i)_{i=1}^N$ należący do sympleksu rozkładów prawdopodobieństwa [21], tzn. jego N składowych spełnia warunki:

$$p_i \geq 0 \quad \text{oraz} \quad \sum_i^N p_i = 1. \quad (2.58)$$

Sympleks rozkładów prawdopodobieństwa, czyli zbiór $\Delta^m = \{(p_1, p_2, \dots, p_m) \in \mathbf{R}^m; p_j \geq 0, j = 1, \dots, m, m \leq N, \text{ dla } \sum_{j=1}^m p_j \leq 1\}$ jest zbiorem wypukłym. Entropia Shannona (2.57) przyjmuje wtedy postać:

$$S(P) = - \sum_{i=1}^N p_i \ln p_i. \quad (2.59)$$

Zapis $S(P)$ gdzie w argumentcie pominięto oznaczenie zmiennej X podkreśla, że jedyną rozważaną przez nas cechą zmiennej losowej jest jej rozkład prawdopodobieństwa P .

2.4.1.1 Interpretacja entropii Shannona

Maksymalna możliwa wartość entropii Shannona wynosi $\ln N$ i jest osiągnięta gdy wszystkie wyniki są równo prawdopodobne ($p_i = 1/N$), tzn. gdy stan układu jest maksymalnie zmieszany.

⁴Mówimy, że na przestrzeni próby Ω została określona funkcja $\omega \rightarrow P(\omega)$ spełniająca warunki, $P(\omega) \geq 0$ oraz $\sum_{\omega} P(\omega) = 1$, nazywana wtedy miarą probabilistyczną. Zbiór wszystkich miar prawdopodobieństwa określonych na Ω tworzy przestrzeń stanów modelu.

⁵W przypadku statystycznej entropii fizycznej N może być np. liczbą konfiguracji określonej liczby molekuł przy zadanej energii całkowitej układu. Wtedy k jest utożsamiane ze stałą Boltzmann'a k_B . Dla układu określonego w przestrzeni ciągłej R^3 liczba konfiguracji jest nieskończona. Gdyby ograniczyć się do skończonej podprzestrzeni i podzielić ją na komórki o skończonej wielkości, i podobnie uczynić w przestrzeni pędowej, to liczba możliwych konfiguracji układu byłaby skończona a jego entropia mogłaby być policzona. Jednakże poprawny rachunek entropii wymaga wtedy utożsamienia konfiguracji powiedzmy n cząstek różniących się jedynie permutacją cząstek w ramach jednej klasy równoważności. Na fakt, że właściwa przestrzeń próby ma w tym przypadku nie N lecz $N/n!$ punktów zwrócił uwagę Gibbs, a otrzymaną przestrzeń próby nazywa się przestrzenią próby Gibbsa. Np. dla 1 cm^3 cieczy w zwykłych warunkach $n \approx 10^{23}$. Problemem ten nie będzie rozważany dalej w niniejszym skrypcie.

Gdy jeden z wyników jest pewny, wtedy tylko jedna, odpowiadająca mu współrzędna wektora \vec{p} jest równa jeden, a pozostałe są równe 0. Mówimy wtedy, że układ znajduje się w stanie czystym, a odpowiada mu wartość entropii Shannona jest minimalna i równa zero.

Z powyższych przykładów można wnioskować, że entropię Shannona można interpretować jako miarę niepewności wyniku eksperymentu będącego realizacją rozkładu prawdopodobieństwa P , lub inaczej jako wielkość informacji koniecznej do określenia wyniku, który może się pojawić w rezultacie przeprowadzenia eksperymentu na układzie.

Podstawowe własności entropii Shannona zostały zebrane w Dodatku.

2.4.1.2 Przypadek ciągłego rozkładu prawdopodobieństwa

Przejdźcie z dyskretnego do ciągłego rozkładu prawdopodobieństwa, polega na zastąpieniu sumowania w (2.59) całkowaniem po całym zakresie zmienności zmiennej losowej. W ten sposób otrzymujemy Boltzmanowską postać entropii Shannona:

$$S(P) = - \int_{-\infty}^{+\infty} dx P(x) \ln P(x). \quad (2.60)$$

Jednakże dla pewnych funkcji rozkładu $P(x)$, całka (2.60) może być niewłaściwie określona. Na przykład [21] gdy $P(x)$ przyjmuje w przedziale $[0, t]$ wartość t^{-1} i zero wszędzie poza tym przedziałem, wtedy entropia $S(P)$ jest równa $\ln t$ i dla $t \rightarrow 0$ dąży do minus nieskończoności. Procedura ta odpowiada przejściu do punktowego, czystego stanu klasycznego opisanego dystrybucją delta Diraca, dla której $S(P) = -\infty$. Zatem przyjmując poziom zerowy entropii jako punkt odniesienia, dokładne określenie stanu opisanego deltą Diraca (której fizycznie mógłby odpowiadać nieskończony skok w gęstości rozkładu substancji cząstki), wymaga dostarczenia nieskończonej ilości informacji o układzie. Do sprawy powrócimy w jednym z kolejnych rozdziałów.

Definicja (2.60) ma jednak pewien formalny minus. Ze względu na unormowanie prawdopodobieństwa do jedności, gęstość rozkładu prawdopodobieństwa przekształca się przy transformacji układu współrzędnych tak jak odwrotność objętości. Np. w jednowymiarowym przypadku z unormowania $\int_{-\infty}^{+\infty} dx P(x) = 1$ wynika, że $P(x)$ musi transformować się tak jak $1/x$. Rozważmy transformację układu współrzędnych $x \rightarrow x'$. Zatem $dx = J(\frac{x}{x'})dx'$, oraz $P'(x') = J^{-1}(\frac{x}{x'})P(x)$ gdzie J jest Jacobianem transformacji natomiast $P(x)$ i $P'(x')$ są gęstościami rozkładu prawdopodobieństwa przed i po transformacji. Jak powinno być, transformacja ta nie zmienia unormowania, tzn. $\int_{-\infty}^{+\infty} dx' P'(x') = 1$.

Rozważmy z kolei entropię Shannona układu określoną dla rozkładu ciągłego jak w (2.60), $S(P) = - \int_{-\infty}^{+\infty} dx P(x) \ln P(x)$. Jak to zauważyliśmy powyżej, entropia układu jest miarą nieuporządkowania w układzie (bądź informacji), zatem również i ona powinna być niezmiennicza przy rozważanej transformacji. Niestety ponieważ $\ln P(x) \neq \ln P'(x')$ zatem $S(P) \neq S(P')$. Zatem logarytm z gęstości rozkładu prawdopodobieństwa “zachowuje się niepoprawnie” przy transformacji układu współrzędnych, oczywiście gdybyśmy chcieli utrzymać entropię Shannona (2.60) jako docelową postać miary nieuporządkowania układu (czy informacji w nim zawartej).

Z drugiej strony, ze względu na wyjątkowe pośród innych entropii własności entropii Shannona

dla rozkładu dyskretnego [21], zrezygnowanie z jej ciągłej granicy (2.60) mogłoby się okazać decyzją chybioną. Również jej związek z informacją Fishera omówiony w dalej, przekonuje o istotności pojęcia entropii Shannona w jej formie ciągłej.

Proste rozwiązanie zaistniałego problemu polega na zaobserwowaniu, że ponieważ iloraz dwóch gęstości $P(x)$ oraz $P_{ref}(x)$ transformuje się jak skalar, tzn. $\frac{P'(x')}{P'_{ref}(x')} = \frac{P(x)}{P_{ref}(x)}$, zatem wielkość nazywana entropią względną:

$$S(P|P_{ref}) \equiv \int_{-\infty}^{+\infty} dx P(x) \ln \frac{P(x)}{P_{ref}(x)}, \quad (2.61)$$

gdzie $P_{ref}(x)$ występuje jako pewien rozkład referencyjny, nie posiada już wady braku niezmienniczości przy transformacji układu współrzędnych.

W ten sposób problem dewiancji (czy ogólniej logarytmu ilorazu funkcji wiarygodności) wykorzystywanego w analizie braku dopasowania modelu, powrócił w postaci konieczności wprowadzenia entropii względnej. W obu przypadkach istotą konstrukcji ilorazu rozkładów okazuje się jej użyteczność w porównaniu modeli statystycznych i wybór modelu bardziej “wiarygodnego”. Poza tym, wybór modelu powinien być też niezmienniczy ze względu na transformację układu współrzędnych. Entropią względną zajmiemy się w Rozdziale (2.4.2) i problem ten będzie nam towarzyszył do końca skryptu.

2.4.2 Entropia względna

Z kolei rozważmy więcej niż jednoelementowy zbiór rozkładów prawdopodobieństwa o \aleph możliwych wynikach. Rozważmy eksperyment, którego wyniki są generowane z pewnego określonego, chociaż nieznanego rozkładu prawdopodobieństwa $P_{ref} \equiv (p^i)_{i=1}^{\aleph}$ należące do zbioru E , który jest przestrzenią metryczną zupełną. Oznacza to, że E nie posiada punktów izolowanych oraz można na nim określić odległość. Poniższe twierdzenie Sanova wypowiada się na temat prawdopodobieństwa realizacji w N -elementowej próbie prostej (powstałej poprzez N -krotną realizację pewnego eksperymentu) dowolnego rozkładu $P_{Ob} \equiv (p_{Ob}^i)_{i=1}^{\aleph}$ należącego do E . Indeks Ob sygnalizuje rozkład prawdopodobieństwa otrzymany w obserwacji, tzn. rozkład empiryczny częstości. Podkreślmy jeszcze raz, że wyniki w eksperymencie są generowane w każdym z nich z rozkładu P_{ref} , co byłoby dokładnie rozpoznane gdyby $N \rightarrow \infty$.

Twierdzenie Sanova: Jeśli eksperyment z \aleph możliwymi wynikami generowanymi z rozkładu prawdopodobieństwa P_{ref} jest N -krotnie powtarzany, oraz E jest nie posiadającym izolowanych punktów zbiorem rozkładów prawdopodobieństwa z \aleph wynikami, wtedy dla dużego N prawdopodobieństwo \mathcal{P} uzyskania rozkładu częstości należącego do zbioru E spełnia warunek:

$$\mathcal{P}(E) \sim e^{-NS(P_{Ob}|P_{ref})}, \quad (2.62)$$

gdzie P_{Ob} jest rozkładem należącym do E z najmniejszą wartością entropii względnej $S(P_{Ob}|P_{ref})$, zdefiniowanej dla rozkładów dyskretnych z \aleph wynikami jako:

$$S(P_{Ob}|P_{ref}) = \sum_{i=1}^{\aleph} p_{Ob}^i \ln \frac{p_{Ob}^i}{p^i}. \quad (2.63)$$

Entropię względną nazywana się również entropią Kullbacka-Leiblera (KL) lub dywergencją informacji. Dla rozkładów ciągłych została ona określona w (2.61). W teorii informacji jest ona

miarą określającą jak bardzo rozkłady P_{Ob} oraz P_{ref} różnią się od siebie.

Zauważmy, że zgodnie z prawem wielkich liczb, dla wielkości próby N dążącej do nieskończoności, prawdopodobieństwo zaobserwowania rozkładu częstości należącego do E różnego od prawdziwego rozkładu P_{ref} (tzn. tego który generował wyniki eksperymentu), dąży do zera. Fakt ten wyraża właśnie relacja (2.62), a ponieważ $S(P_{Ob}|P_{ref})$ jest w jej wykładniku, zatem entropia względna określa tempo w jakim prawdopodobieństwo $\mathcal{P}(E)$ dąży do zera wraz ze wzrostem N .

W celu ilustracji twierdzenie Sanova założmy, że przeprowadzamy doświadczenie rzutu niesymetryczną monetą z wynikami orzeł, reszka, zatem $\aleph = 2$. Rozkład teoretyczny P_{ref} jest więc zero-jedynkowy. *Natomiast w wyniku pobrania N -elementowej próbki dokonujemy jego estymacji na podstawie rozkładu empirycznego P_{Ob} częstości pojawienia się wyników orzeł lub reszka.* Zatem:

$$P_{ref} \equiv (p^i)_{i=1}^{\aleph=2} = (p, 1-p) \quad \text{oraz} \quad P_{Ob} \equiv (p_{Ob}^i)_{i=1}^{\aleph=2} = \left(\frac{m}{N}, 1 - \frac{m}{N}\right). \quad (2.64)$$

Twierdzenie Bernoulliego mówi, że prawdopodobieństwo pojawienia się wyniku orzeł z częstością m/N w N -losowaniach wynosi:

$$\mathcal{P}(P_{Ob}) \equiv \mathcal{P}\left(\frac{m}{N}\right) = \binom{N}{m} p^m (1-p)^{N-m}. \quad (2.65)$$

Biorąc logarytm naturalny obu stron (2.65), następnie stosując słuszne dla dużego n przybliżenie Stirlinga, $\ln n! \approx n \ln n - n$, dla każdej silni w wyrażeniu $\binom{N}{m}$, i w końcu biorąc eksponentę obu stron, można otrzymać [21]:

$$\mathcal{P}(P_{Ob}) \approx e^{-NS(P_{Ob}|P_{ref})}, \quad (2.66)$$

gdzie

$$\begin{aligned} S(P_{Ob}|P_{ref}) &= \left[\frac{m}{N} \left(\ln \frac{m}{N} - \ln p \right) + \left(1 - \frac{m}{N} \right) \left(\ln \left(1 - \frac{m}{N} \right) - \ln(1-p) \right) \right] \\ &= \sum_{i=1}^{\aleph=2} p_{Ob}^i \ln \frac{p_{Ob}^i}{p^i}. \end{aligned} \quad (2.67)$$

gdzie w ostatniej równości skorzystano z (2.64) otrzymując entropię względną (2.63) dla przypadku $\aleph = 2$. W powyższym przykładzie entropia względna $S(P_{Ob}|P_{ref})$ pojawiła się jako pojęcie wtórne, wynikające z rachunku prawdopodobieństwa otrzymania w eksperymencie oszacowania P_{Ob} rozkładu P_{ref} . Fakt ten oznacza, że entropia względna *nie jest tworem sztucznym*, wprowadzonym do teorii jedynie dla wygody jako pewna miara odległości pomiędzy rozkładami, lecz, że jest właściwą dla przestrzeni statystycznej rozkładów miarą probabilistyczną tej odległości, tzn. “dywergencją informacji” pomiędzy rozkładami. Okaze się też, że pośród innych miar odległości jest ona wyróżniona poprzez jej związek ze znaną już nam informacją Fishera.

2.5 Geometria przestrzeni rozkładów prawdopodobieństwa i metryka Fishera-Rao

Pojęcie metryki Fishera wprowadziliśmy już w Rozdziale 2.2. Do ujęcia tam przedstawionego dojdziemy jeszcze raz, wychodząc od pojęcia entropii względnej.

Rozważmy dwa rozkłady prawdopodobieństwa, $P = (p^i)_{i=1}^{\aleph}$ oraz $P' = P + dP = (p'^i = p^i + dp^i)_{i=1}^{\aleph}$, różniące się *infinitesimalnie mało*, przy czym $p_i \neq 0$ dla każdego i . Rozkłady P oraz P' spełniają warunek unormowania $\sum_{i=1}^{\aleph} p^i = 1$ oraz $\sum_{i=1}^{\aleph} p^i + dp^i = 1$, skąd:

$$\sum_{i=1}^{\aleph} dp^i = 0. \quad (2.68)$$

Ponieważ dp^i/p^i jest wielkością infinitesimalnie małą, zatem z rozwinięcia:

$$\ln\left(1 + \frac{dp^i}{p^i}\right) = \frac{dp^i}{p^i} - 1/2\left(\frac{dp^i}{p^i}\right)^2 + \dots \quad (2.69)$$

otrzymujemy, że względna entropia KL rozkładów P oraz P' wynosi:

$$S(P|P + dP) = \sum_{i=1}^{\aleph} p^i \ln \frac{p^i}{p^i + dp^i} = - \sum_{i=1}^{\aleph} p^i \ln \frac{p^i + dp^i}{p^i} \approx \frac{1}{2} \sum_{i=1}^{\aleph} \frac{dp^i dp^i}{p^i}. \quad (2.70)$$

Ostatnia postać $S(P|P + dP)$ sugeruje, że entropia względna KL określa w naturalny sposób na przestrzeni statystycznej \mathcal{S} infinitesimalny kwadrat odległości pomiędzy rozważanymi rozkładami, co oznaczałoby również *lokalne* określenie na \mathcal{S} pewnej metryki. Do jej związku z metryką Rao-Fisher'a wprowadzoną w Rozdziale 2.2 powrócimy później.

Niech $d\vec{p} = (dp^1, \dots, dp^{\aleph})$ jest infinitesimalnym wektorem w przestrzeni rozkładów prawdopodobieństwa, spełniającym warunek (2.68). Wprowadzając na przestrzeni statystycznej \mathcal{S} aparat matematyczny geometrii różniczkowej [12], zapiszmy kwadrat różniczkowego interwału w tej przestrzeni następująco:

$$ds^2 = \sum_{i,j=1}^{\aleph} g_{ij} dp^i dp^j. \quad (2.71)$$

Aby uzgodnić (2.70) z (2.71) wprowadźmy na \mathcal{S} metrykę:

$$g_{ij} = \frac{\delta_{ij}}{p^i}. \quad (2.72)$$

Związek (2.71) oznacza, że liczba możliwych wyników \aleph określa wymiar przestrzeni \mathcal{S} oraz, że entropia względna KL definiuje dla infinitesimalnie bliskich rozkładów symetryczną metrykę na przestrzeni statystycznej \mathcal{S} , pozwalającą mierzyć odległość pomiędzy tymi rozkładami. Okazuje się, że metryka ta jest *metryką Rao-Fisher'a* wprowadzoną w Rozdziale 2.2.

Powyżej stan układu określony był w reprezentacji rozkładu prawdopodobieństwa zmiennej losowej, tzn. określony był przez podanie rozkładu prawdopodobieństwa. Dwóm infinitesimalnie leżącym stanom opowiadały rozkłady P oraz P' . Gdy interesują nas czysto geometryczne własności metryki Fishera, wtedy wygodne jest użycie innej reprezentacji do opisu stanu układu,

określonej jak następuje. Niech $Q \equiv (q^i)_{i=1}^{\aleph}$ oraz $Q' \equiv (q'^i)_{i=1}^{\aleph} = (q^i + dq^i)_{i=1}^{\aleph}$ opisują te same dwa nieskończenie blisko siebie leżące stany układu, tyle, że zarządzajmy, aby nieskończenie mały kwadrat interwału pomiędzy nimi równy (2.71) był równy:

$$ds^2 = 4 \sum_{i=1}^{\aleph} dq^i dq^i . \quad (2.73)$$

Porównując formułę (2.73) z (2.71), zauważamy, że zgadzają się one z sobą o ile $Q = (q^i)_1^{\aleph}$ oraz $P = (p^i)_1^{\aleph}$ są powiązane związkiem:

$$dq^i = \frac{dp^i}{2\sqrt{p^i}} \quad i = 1, \dots, \aleph , \quad (2.74)$$

co zachodzi wtedy gdy

$$q^i = \sqrt{p^i} \quad i = 1, \dots, \aleph . \quad (2.75)$$

Wielkości q^i nazywamy amplitudami układu⁶. Definiują one na \mathcal{S} nowe współrzędne dla których $q^i \geq 0$ dla każdego i , dla których otrzymana geometria jest geometrią jednostkowej sfery, tzn. ze względu na unormowania rozkładu prawdopodobieństwa do jedności, amplitudy spełniają następujący związek unormowania na promieniu jednostkowym:

$$\sum_{i=1}^{\aleph} p^i = \sum_{i=1}^{\aleph} q^i q^i = 1 , \quad (2.76)$$

Na sferze tej możemy określić odległość geodezyjną D_{Bhatt} , tzw. odległość Bhattacharyya', dla dwóch rozkładów prawdopodobieństwa P oraz P' , jako długość kątową liczoną wzdłuż koła wielkiego pomiędzy dwoma wektorami Q oraz Q' o składowych będących amplitudami $q^i = \sqrt{p^i}$ oraz $q'^i = \sqrt{p'^i}$. Kwadrat nieskończenie małego interwału (2.73) jest przykładem odległości Bhattacharyya, którą ogólnie określamy następująco: Jeśli $P = (p^i)$ oraz $P' = (p'^i)$ są rozkładami prawdopodobieństwa, wtedy odległość Bhattacharyya pomiędzy Q oraz Q' jest iloczyn wewnętrzny określony na przestrzeni statystycznej \mathcal{S} następująco:

$$\cos D_{Bhatt} = \sum_{i=1}^{\aleph} q^i q'^i = \sum_{i=1}^{\aleph} \sqrt{p^i p'^i} \equiv B(p, p') . \quad (2.77)$$

Statystyczna odległością Bhattacharyya' wygląda więc jak iloczyn wewnętrzny mechaniki kwantowej. W kolejnym rozdziale przekonamy się, że nie jest to luźne skojarzenie.

Zauważmy, że metryka Rao-Fishera jest Hessianem (tzn. macierzą drugich pochodnych) entropii Shannona:

$$g_{ij} = -\partial_i \partial_j S(p) = \frac{\partial}{\partial p^i} \frac{\partial}{\partial p^j} \sum_{k=1}^{\aleph} p^k \ln p^k = \frac{\delta^{ij}}{p^j} \geq 0 , \quad (2.78)$$

zgodnie z (2.72). Powyższy związek oznacza, że fakt wklęsłości entropii Shannona [21] daje dodatnią określoność metryki Fishera-Rao.

⁶Fisher korzystał z amplitud prawdopodobieństwa niezależnie od ich pojawienia się w mechanice kwantowej

Uzasadnijmy w końcu fakt nazwania metryki (2.72) metryką Rao-Fishera. Załóżmy, że interesuje nas pewna podprzestrzeń przestrzeni rozkładów prawdopodobieństwa (która tworzy sympleks). Wprowadźmy na niej układem współrzędnych θ^a . Korzystając z (2.72) widać, że metryka $g_{ij} = \frac{\delta^{ij}}{p^j}$ indukuje w tej podprzestrzeni metrykę:

$$g_{ab} = \sum_{i,j=1}^{\aleph} \frac{\partial p^i}{\partial \theta^a} \frac{\partial p^j}{\partial \theta^b} g_{ij} = \sum_{i=1}^{\aleph} \frac{\partial_a p^i \partial_b p^i}{p^i} \quad \text{gdzie} \quad \partial_a p^i \equiv \partial p^i / \partial \theta^a . \quad (2.79)$$

Korzystając ze związku $q^i = \sqrt{p^i}$, (2.75), pomiędzy prawdopodobieństwami a amplitudami, metrykę Rao-Fishera (2.79) można zapisać następująco:

$$g_{ab} = 4 \sum_{i=1}^{\aleph} \frac{\partial q^i}{\partial \theta^a} \frac{\partial q^i}{\partial \theta^b} . \quad (2.80)$$

Gdy liczba parametrów θ^a (więc i wektorów bazowych $\vec{\theta}^a$) rozpinających osie układu współrzędnych rozważanej podprzestrzeni jest skończona, wtedy można dokonać następującego uogólnienia metryki na przypadek ciągłego rozkładu prawdopodobieństwa $P(y)$ zmiennej losowej Y :

$$g_{ab} = \int_{\mathcal{Y}} dy \frac{\partial_a P}{P} \frac{\partial_b P}{P} = \int_{\mathcal{Y}} dy P \partial_a l \partial_b l \equiv E(\partial_a l \partial_b l) = E_{\Theta} \mathbf{iF}(\Theta) = I_F(\Theta) , \quad (2.81)$$

gdzie ponieważ przestrzeń zdarzeń ciągłego rozkładu prawdopodobieństwa jest nieskończenie wymiarowa, zatem w miejsce sumowania po $i = 1, \dots, \aleph$ pojawiło się całkowanie po wartościach $y \in \mathcal{Y}$ zmiennej losowej Y , oraz skorzystano z zapisu $l(y|\Theta) = \ln P(y|\Theta)$ z Rozdziału 2.2. Otrzymana postać metryki jest jawnie niezmiennicza ze względu na reparametryzację przestrzeni \mathcal{Y} , tzn. ze względu na przejście do nowego układu współrzędnych $y'(y)$.

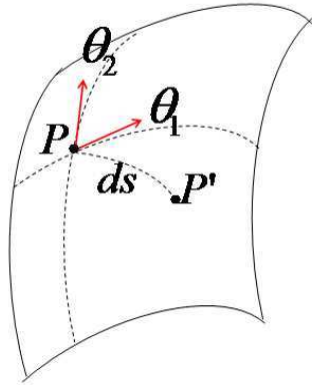
Gdy rozkład prawdopodobieństwa $P(y)$ jest funkcją wiarygodności próby, wtedy zbiór \mathcal{Y} jest N -wymiarową przestrzenią próby.

W Rozdziale 2.2 wzór (2.22) zdefiniowaliśmy metrykę Rao-Fishera jako $g_{ab} = E_{\vartheta}(\partial_a l \partial_b l)$. Tak więc ostatecznie otrzymaliśmy zgodność nazwania metryki (2.72) metryką Rao-Fishera. Na koniec, dokonując reparametryzacji i przechodząc do amplitud $q(y|\Theta) = \sqrt{P}(y|\Theta)$, można metrykę (2.81) dla przypadku rozkładu ciągłego zapisać następująco:

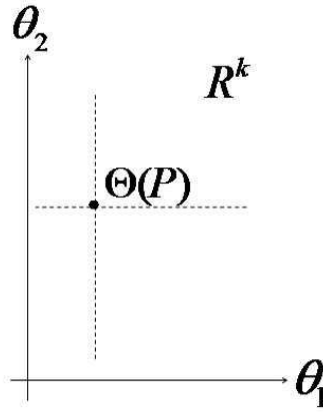
$$g_{ab} = 4 \int_{\mathcal{Y}} dy \partial_a q(y|\Theta) \partial_b q(y|\Theta) . \quad (2.82)$$

Omówiną sytuację odległości na przestrzeni statystycznej przedstawia graficznie poniższy rysunek.

przestrzeń statystyczna \mathcal{S}



układ współrzędnych (θ_k)



Dla dwóch nieskończenie bliskich rozkładów:

P oraz $P' = P + dP$

$$ds^2 = g_{ij} d\theta^i d\theta^j \quad \text{gdzie } g_{ij} \text{ jest metryką Rao-Fishera}$$

Przykład: Wyznaczyć kwadrat nieskończonego interwału dla dwóch stanów posiadających rozkład normalny $N(\mu, \sigma)$. Rozkład normalny ma postać:

$$P(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}, \quad (2.83)$$

gdzie wektor parametrów $\Theta \equiv (\theta^a) = (\mu, \sigma)$. Macierz informacyjna Fishera dla rozkładu normalnego $N(\mu, \sigma)$ ma wyznaczoną poprzednio postać (2.42). Zatem metryka Rao-Fishera na 2-wymiarowej przestrzeni normalnych rozkładów prawdopodobieństwa z układem współrzędnych μ, σ ma (dla próby $N = 1$), postać:

$$(g_{ab}) = I_F(\Theta) = E_{\Theta} \mathbf{iF}(\Theta) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix}, \quad (2.84)$$

tzn. składowe metryki (2.81) są równe

$$g_{\mu\mu} = \frac{1}{\sigma^2}, \quad g_{\mu\sigma} = g_{\sigma\mu} = 0, \quad g_{\sigma\sigma} = \frac{2}{\sigma^2}. \quad (2.85)$$

Zatem otrzymany kwadrat nieskończonego interwału na 2-wymiarowej (pod)przestrzeni statystycznej \mathcal{S} wynosi [21]:

$$ds^2 = g_{\mu\mu} d\mu^2 + g_{\mu\sigma} d\mu d\sigma + g_{\sigma\sigma} d\sigma^2 = \frac{1}{\sigma^2} (d\mu^2 + 2d\sigma^2). \quad (2.86)$$

Odpowiada on metryce Poincarégo ze stałą ujemną krzywizną. W końcu, z postaci rozkładu (2.83) otrzymujemy również bazę w przestrzeni stycznej do \mathcal{S} :

$$\partial_{\mu} \ln P(y|\mu, \sigma) = \frac{y - \mu}{\sigma^2}, \quad \partial_{\sigma} \ln P(y|\mu, \sigma) = \frac{(y - \mu)^2}{\sigma^3} - \frac{1}{\sigma}. \quad (2.87)$$

Na osi μ na której $\sigma = 0$ leżą punkty odpowiadające klasycznym stanom czystym i są one zgodnie z (2.86) nieskończenie daleko odległe od dowolnego punktu we wnętrzu górnej półpłaszczyzny $\sigma > 0$. W konsekwencji oznacza to, że wynik pewny, któremu odpowiada rozkład klasyczny z $\sigma = 0$, jest łatwy do odróżnienia od każdego innego.

2.6 Informacja Fishera

2.6.1 Informacja Fishera jako entropia

W powyższych rozważaniach informacja Fishera I_F pojawiła się poprzez nierówność Rao-Cramera (2.29) jako wielkość określająca graniczną dobroć procedury estymacyjnej parametru rozkładu, tzn. o ile estymator efektywny istnieje to informacja Fishera określa minimalną wartość jego wariancji. Ponieważ im informacja Fishera mniejsza tym to graniczne oszacowanie parametru gorsze, zatem jest ona również miarą stopnia nieuporządkowania układu określonego rozkładem prawdopodobieństwa. Takie zrozumienie informacji Fishera odnosi się do typu analizowanego rozkładu prawdopodobieństwa (spróbuj pomyśleć o sensie oszacowania wartości oczekiwanej rozkładu jednorodnego) i jako miara nieuporządkowania rozkładu oznacza brak przewidywalności procedury estymacyjnej. Poniżej uzasadnimy stwierdzenie, że informacja Fishera okazuje się być proporcjonalna do entropii względnej Kullbacka-Leiblera rozkładów różniących się infinytezymalnie mało w parametrze rozkładu [7].

Rozważmy N -wymiarową próbę Y_1, Y_2, \dots, Y_N , gdzie każda ze zmiennych losowych Y_n jest określona na przestrzeni \mathcal{Y} i posiada rozkład prawdopodobieństwa $p(y_n|\theta)$. Funkcja wiarygodności $P(y|\theta)$ ma postać:

$$P(y|\theta) = \prod_{n=1}^N p_n(y_n|\theta) \quad (2.88)$$

gdzie $y \equiv (\mathbf{y}_n)_{n=1}^N = y_1, y_2, \dots, y_N$ jest realizacją próby. Dla uproszczenia rozważań θ jest parametrem skalarnym.

Zauważmy wpraw, że informacja Fishera parametru θ określona w (2.11) i (2.2):

$$I_F = - \int dy P(y|\theta) \frac{\partial^2 \ln P(y|\theta)}{\partial \theta^2}, \quad (2.89)$$

gdzie $dy \equiv d^N \mathbf{y} = dy_1 dy_2 \dots dy_N$, można zapisać następująco:

$$I_F = \int dy P \left(\frac{\partial \ln P}{\partial \theta} \right)^2. \quad (2.90)$$

Postać (2.90) otrzymujemy z (2.89) po skorzystaniu z:

$$\frac{\partial}{\partial \theta} \left(\frac{\partial \ln P}{\partial \theta} \right) = \frac{\partial}{\partial \theta} \left(\frac{1}{P} \frac{\partial P}{\partial \theta} \right) = - \frac{1}{P^2} \left(\frac{\partial P}{\partial \theta} \right)^2 + \frac{1}{P} \left(\frac{\partial^2 P}{\partial \theta^2} \right), \quad (2.91)$$

oraz warunku regularności pozwalającego na wyłączenie różniczkowania po parametrze przed całkę (Rozdział):

$$\int dy \left(\frac{\partial^2 P}{\partial \theta^2} \right) = \frac{\partial^2}{\partial \theta^2} \int dy P = \frac{\partial^2}{\partial \theta^2} 1 = 0. \quad (2.92)$$

Po skorzystaniu z (2.88) zauważamy, że:

$$\ln P(y; \theta) = \sum_{n=1}^N \ln p_n \quad \text{zatem} \quad \frac{\partial \ln P}{\partial \theta} = \sum_{n=1}^N \frac{1}{p_n} \frac{\partial p_n}{\partial \theta} \quad (2.93)$$

i podnosząc ostatnie wyrażenie do kwadratu otrzymujemy:

$$\left(\frac{\partial \ln P}{\partial \theta}\right)^2 = \sum_{\substack{n,m=1 \\ m \neq n}}^N \frac{1}{p_m} \frac{1}{p_n} \frac{\partial p_m}{\partial \theta} \frac{\partial p_n}{\partial \theta} + \sum_{n=1}^N \frac{1}{p_n^2} \left(\frac{\partial p_n}{\partial \theta}\right)^2. \quad (2.94)$$

Wstawiając (2.88) oraz (2.94) do (2.90) otrzymujemy:

$$I_F = \int dy \prod_{k=1}^N p_k \sum_{\substack{n,m=1 \\ m \neq n}}^N \frac{1}{p_m} \frac{1}{p_n} \frac{\partial p_m}{\partial \theta} \frac{\partial p_n}{\partial \theta} + \int dy \prod_{k=1}^N p_k \sum_{n=1}^N \frac{1}{p_n^2} \left(\frac{\partial p_n}{\partial \theta}\right)^2. \quad (2.95)$$

Ze względu na warunek normalizacji rozkładów brzegowych:

$$\int dy_n p_n(y_n|\theta) = 1 \quad (2.96)$$

w (2.95) z iloczynu $\prod_{k=1}^N p_k$ pozostaje w pierwszym składniku jedynie $p_m p_n$ (dla $k \neq m$ oraz $k \neq n$) natomiast w drugim składniku pozostaje jedynie p_n (dla $k \neq n$), tzn.:

$$I_F = \sum_{\substack{n,m=1 \\ m \neq n}}^N \iint dy_m dy_n \frac{\partial p_m}{\partial \theta} \frac{\partial p_n}{\partial \theta} + \sum_{n=1}^N \int dy_n \frac{1}{p_n} \left(\frac{\partial p_n}{\partial \theta}\right)^2. \quad (2.97)$$

W końcu, ze względu na:

$$\int dy_n \frac{\partial p_n}{\partial \theta} = \frac{\partial}{\partial \theta} \int dy_n p_n = \frac{\partial}{\partial \theta} 1 = 0, \quad (2.98)$$

pierwszy składnik w (2.97) zeruje się i pozostaje jedynie drugi, zatem:

$$I_F = \sum_{n=1}^N \int dy_n \frac{1}{p_n} \left(\frac{\partial p_n}{\partial \theta}\right)^2 = \sum_{n=1}^N \int dy_n p_n \left(\frac{\partial \ln p_n}{\partial \theta}\right)^2. \quad (2.99)$$

Ostatecznie otrzymujemy więc następującą postać informacji Fishera:

$$I_F = \sum_{n=1}^N I_{F_n} \quad \text{gdzie} \quad I_{F_n} = \int dy_n p_n(y_n|\theta) \left(\frac{\partial \ln p_n(y_n|\theta)}{\partial \theta}\right)^2. \quad (2.100)$$

Aby pokazać związek informacji Fishera z entropią Kullbacka-Leiblera zastąpmy całkę w (2.99) sumą Riemanna:

$$\begin{aligned} I_F &= \sum_{n=1}^N \sum_k \Delta y_{nk} \frac{1}{p_n(y_{nk}|\theta)} \left[\frac{p_n(y_{nk}|\theta + \Delta\theta) - p_n(y_{nk}|\theta)}{\Delta\theta} \right]^2 \\ &= (\Delta\theta)^{-2} \sum_{n=1}^N \Delta y_n \sum_k p_n(y_{nk}|\theta) \left[\frac{p_n(y_{nk}|\theta + \Delta\theta)}{p_n(y_{nk}|\theta)} - 1 \right]^2. \end{aligned} \quad (2.101)$$

Powyższe przejście od (2.100) do (2.101) jest słuszne dla różniczkowania w granicy $\Delta\theta \rightarrow 0$. Natomiast zamiana całkowania na sumę została wprowadzona dla wygody i jest ścisłym przejściem w granicy $\Delta y_{nk} \xrightarrow{k \rightarrow \infty} 0$. W drugiej równości w (2.101) przyjęto równe przyrosty $\Delta y_{nk} = \Delta y_n$ dla każdego k , co w tej granicy nie zmienia wyniku.

W granicy $\Delta\theta \rightarrow 0$ każde z powyższych wyrażeń $p_n(y_{nk}|\theta + \Delta\theta)/p_n(y_{nk}|\theta)$ dąży do 1. Wtedy każda z wielkości:

$$\delta_{\Delta\theta}^n \equiv \frac{p_n(y_{nk}|\theta + \Delta\theta)}{p_n(y_{nk}|\theta)} - 1 \quad (2.102)$$

staje się mała i rozwijając funkcję logarytm do wyrazu drugiego rzędu:

$$\ln(1 + \delta_{\Delta\theta}^n) \approx \delta_{\Delta\theta}^n - (\delta_{\Delta\theta}^n)^2/2, \quad (2.103)$$

otrzymujemy:

$$(\delta_{\Delta\theta}^n)^2 \approx 2[\delta_{\Delta\theta}^n - \ln(1 + \delta_{\Delta\theta}^n)]. \quad (2.104)$$

Korzystając z (2.102) oraz (2.104) wyrażenie (2.101) można zapisać następująco:

$$\begin{aligned} I_F &= -2(\Delta\theta)^{-2} \sum_{n=1}^N \Delta y_n \left[\sum_k p_n(y_{nk}|\theta) \ln \left(\frac{p_n(y_{nk}|\theta + \Delta\theta)}{p(y_k|\theta)} \right) \right. \\ &\quad \left. - \sum_k p_n(y_{nk}|\theta + \Delta\theta) + \sum_k p_n(y_{nk}|\theta) \right], \quad \text{dla } \Delta\theta \rightarrow 0. \end{aligned} \quad (2.105)$$

Ze względu na warunek normalizacji zachodzi $\sum_k p_n(y_{nk}|\theta + \Delta\theta) = \sum_k p_n(y_{nk}|\theta) = 1$, zatem dwie ostatnie sumy po k w nawiasie kwadratowym znoszą się wzajemnie i (2.105) redukuje się do postaci:

$$\begin{aligned} I_F &= -2(\Delta\theta)^{-2} \sum_{n=1}^N \sum_k \Delta y_n p_n(y_{nk}|\theta) \ln \left(\frac{p_n(y_{nk}|\theta + \Delta\theta)}{p_n(y_{nk}|\theta)} \right) \\ &= -2(\Delta\theta)^{-2} \sum_{n=1}^N \int dy_n p_n(y_n|\theta) \ln \left(\frac{p_n(y_n|\theta + \Delta\theta)}{p_n(y_n|\theta)} \right) \\ &= -2(\Delta\theta)^{-2} \sum_{n=1}^N S[p_n(y_n|\theta) | p_n(y_n|\theta + \Delta\theta)], \quad \text{dla } \Delta\theta \rightarrow 0, \end{aligned} \quad (2.106)$$

gdzie w ostatnim przejściu skorzystaliśmy z (2.61). Drugą z równości w (2.106) można zapisać również następująco:

$$\begin{aligned} I_F &= -2(\Delta\theta)^{-2} \prod_{\substack{m=1 \\ m \neq n}}^N \int dy_m p_m(y_m|\theta) \sum_{n=1}^N \int dy_n p_n(y_n|\theta) \ln \left(\frac{p_n(y_n|\theta + \Delta\theta)}{p_n(y_n|\theta)} \right) \\ &= -2(\Delta\theta)^{-2} \int dy \prod_{m=1}^N p_m(y_m|\theta) \ln \left(\frac{\prod_{n=1}^N p_n(y_n|\theta + \Delta\theta)}{\prod_{n=1}^N p_n(y_n|\theta)} \right) \\ &= -2(\Delta\theta)^{-2} S[P(y|\theta) | P(y|\theta + \Delta\theta)], \quad \text{dla } \Delta\theta \rightarrow 0. \end{aligned} \quad (2.107)$$

Tak więc otrzymaliśmy (jakby) wsparcie dla intuicji wspomnianej na początku obecnego Rozdziału, która mówi, że skoro entropia jest miarą nieuporządkowania układu to informacja Fishera jest również miarą jego nieuporządkowania (2.107). Jednakże o własnościach globalnych rozkładu wypowiada się entropia Shannona, natomiast wniosek wynikający z (2.107) ma sens tylko dla entropii względnej i to lokalnie. Zatem związek ten może być co najwyżej sygnałem, że niektóre własności układu związane z entropią Shannona mogą być ujęte w języku informacji Fishera. Na

niektóre sytuacje, w których ma to miejsce zwrócimy uwagę w przyszłości.

Z porównania (2.107) z (2.106) otrzymujemy dodatkowo, że entropia względna rozkładów $P(y|\theta)$ oraz $P(y|\theta + \Delta\theta)$ jest sumą entropii względnych odpowiadających im rozkładów punktowych $p_n(y_n|\theta)$ oraz $p_n(y_n|\theta + \Delta\theta)$:

$$S[P(y|\theta) | P(y|\theta + \Delta\theta)] = \sum_{n=1}^N S[p_n(y_n|\theta) | p_n(y_n|\theta + \Delta\theta)] \quad \text{dla dowolnego } \Delta\theta. \quad (2.108)$$

Ponieważ $\Delta\theta \rightarrow 0$ nie był wykorzystywany w przejściu od (2.107) z (2.106), zatem związek ten jest słuszny dla dowolnego $\Delta\theta$. Natomiast należy pamiętać, że rozkłady punktowe w definicji funkcji wiarygodności są *nieskorelowane* dla różnych n , dlatego też związek (2.108) jest słuszny tylko w tym przypadku.

2.7 Pojęcie kanału informacyjnego

Niech pierwotna zmienna losowa Y przyjmuje wartości wektorowe $\mathbf{y} \in \mathcal{Y}$. Wektor \mathbf{y} może być np. wektorem położenia. Zatem, aby wprowadzony opis był wystarczająco ogólny, wartości $\mathbf{y} \equiv (\mathbf{y}^\nu)$, mogą posiadać np. indeks wektorowy ν . Wartości te są realizowane zgodnie z (łącznym) rozkładem $p(\mathbf{y}|\Theta)$ właściwym dla badanego układu.

Rozważmy N -wymiarową próbę. Oznaczmy przez $y \equiv (\mathbf{y}_n)_{n=1}^N = (\mathbf{y}_1, \dots, \mathbf{y}_N)$ dane będące realizacjami próby $\tilde{Y} = (Y_1, Y_2, \dots, Y_N)$ dla pierwotnej zmiennej Y , gdzie $\mathbf{y}_n \equiv (\mathbf{y}_n^\nu)$ oznacza n -tą wektorową obserwację w próbce ($n = 1, 2, \dots, N$). Rozkład łączny próby jest określony przez $P(y|\Theta)$.

Podobnie postąpimy dla parametrów rozkładu. Niech indeks α określa pewną dodatkową współrzędną wektorową parametru θ_i , gdzie jak w poprzednich rozdziałach $i = 1, 2, \dots, p$. Zatem wektor parametrów ma teraz postać:

$$\Theta = (\theta_1, \theta_2, \dots, \theta_p) \quad \text{gdzie} \quad \theta_i = (\theta_i^\alpha) . \quad (2.109)$$

Wariancja estymatora $\hat{\theta}_i^\alpha$ parametru θ_i^α otrzymana z próby ma postać:

$$\sigma^2(\hat{\theta}_i^\alpha) = \int dy P(y|\Theta) \left(\hat{\theta}_i^\alpha(y) - \theta_i^\alpha \right)^2 , \quad (2.110)$$

gdzie $\hat{\theta}_i^\alpha(y)$ jest estymatorem MNW parametru θ_i^α , a całkowanie przebiega po całej przestrzeni próby, tzn. po wszystkich możliwych realizacjach y .

Według Rozdziału 2.3.2.1, dla każdego wyróżnionego parametru θ_i^α wariancja $\sigma^2(\hat{\theta}_i^\alpha)$ jego estymatora (2.110) jest związana z informacją Fishera $I_{Fi\alpha}$ parametru θ_i^α , poprzez nierówność (2.55):

$$\sigma^2(\hat{\theta}_i^\alpha) \geq I_F^{i\alpha} \geq \frac{1}{I_{Fi\alpha}} , \quad (2.111)$$

gdzie $I^{i\alpha} \equiv I_F^{i\alpha, i\alpha}$ jest dolnym ograniczeniem RC dla parametru θ_i^α w przypadku wieloparametrowym, natomiast:

$$I_{i\alpha} \equiv I_{Fi\alpha, i\alpha} , \quad (2.112)$$

jest pojemnością informacyjną w *pojedynczym kanale informacyjnym* (i, α) czyli *informacją Fishera dla parametru θ_i^α* . Zgodnie z (2.90) jest ona równa⁷:

$$I_{i\alpha} \equiv I_F(\theta_{i\alpha}) = \int dy P(y|\Theta) \left(\frac{\partial \ln P(y|\Theta)}{\partial \theta_i^\alpha} \right)^2 . \quad (2.113)$$

Wielkość $\frac{1}{\sigma^2(\hat{\theta}_i^\alpha)}$ nazwał Fisher *wewnętrzną dokładnością* pojedynczego kanału (i, α). Sumując ją po indeksach α oraz i , otrzymujemy tzw. informację Stama I_S [22]:

$$I_S \equiv \sum_i \sum_\alpha \frac{1}{\sigma^2(\hat{\theta}_i^\alpha)} , \quad (2.114)$$

która jest skalarną miarą jakości jednoczesnej estymacji we wszystkich kanałach informacyjnych. W końcu, sumując lewą i prawą stroną (2.121) po indeksach α oraz i widać, że I_S spełnia nierówność:

$$I_S \equiv \sum_i \sum_\alpha \frac{1}{\sigma^2(\hat{\theta}_i^\alpha)} \leq \sum_i \sum_\alpha I_{i\alpha} =: I \equiv C , \quad (2.115)$$

⁷Porównaj przejście od (2.89) do (2.90).

gdzie C , oznaczana dalej jako I , nazywana jest *pojemnością informacyjną układu*. Zgodnie z (2.115) i (2.113) jest ona równa:

$$I = \sum_{n=1}^N \sum_{\alpha} I_{i\alpha} = \sum_i \int dy P(y|\Theta) \sum_{\alpha} \left(\frac{\partial \ln P(y|\Theta)}{\partial \theta_i^{\alpha}} \right)^2. \quad (2.116)$$

Jak się okaże, pojemność informacyjna I jest najważniejszym pojęciem statystyki leżącym u podstaw np. członów kinetycznych różnych modeli teorii pola. Jest ona uogólnieniem pojęcia informacja Fishera dla przypadku pojedynczego, skalarnego parametru, na przypadek wieloparametrowy.

2.7.1 Pojemność informacyjna dla zmiennej losowej położenia

Zawężmy obszar analizy do szczególnego przypadku gdy interesującym nas (naturalnym) parametrem jest wartość oczekiwana zmiennej położenia układu Y :

$$\theta \equiv E(Y) = (\theta^{\nu}) \quad \text{gdzie} \quad \theta^{\nu} = \int dy p(y) y^{\nu}. \quad (2.117)$$

Wtedy N -wymiarowa próbka $y \equiv (\mathbf{y}_n)_{n=1}^N = (\mathbf{y}_1, \dots, \mathbf{y}_N)$ jest realizacją próby \tilde{Y} dla *położenia układu*, a wartość oczekiwana θ_n położenia układu w n -tym punkcie (tzn. pomiarze) próby wynosi:

$$\theta_n \equiv E(Y_n) = (\theta_n^{\nu}) \quad \text{gdzie} \quad \theta_n^{\nu} = \int dy P(y|\Theta) \mathbf{y}_n^{\nu}. \quad (2.118)$$

Gdy, jak to ma miejsce w rozważanym przypadku, jedynym parametrem rozkładu, który nas interesuje jest wartość oczekiwana położenia $\theta_n \equiv (\theta_n^{\nu})$, gdzie $n = 1, 2, \dots, N$ jest indeksem próby, wtedy parametr wektorowy $\Theta = (\theta_n)_{n=1}^N$. Oznacza to, że wymiar parametru Θ jest taki sam jak wymiar próby \tilde{Y} . Zatem liczba parametrów p pokryła się z wymiarem próby N , a indeksy parametru θ_i^{α} są następujące: $i \equiv n$, $n = 1, 2, \dots, N$ oraz $\alpha \equiv \nu$, gdzie ν jest indeksem wektorowym współrzędnej \mathbf{y}_n^{ν} .

Założenie: W rozważaniach niniejszego skryptu zmienne Y_n próby \tilde{Y} są niezależne (tzn. zakładamy, że pomiary dla $m \neq n$ są w próbie niezależne). Oznacza to również, że wartość oczekiwana położenia $\theta_m = \int dy P(y|\Theta) \mathbf{y}_m$, nie ma wpływu na rozkład $p_n(\mathbf{y}_n|\theta_n)$ dla indeksu próby $m \neq n$. Wtedy dane są generowane zgodnie z punktowymi rozkładami spełniającymi warunek:

$$p_n(\mathbf{y}_n|\Theta) = p_n(\mathbf{y}_n|\theta_n), \quad \text{gdzie} \quad n = 1, \dots, N, \quad (2.119)$$

a wiarygodność próby jest iloczynem:

$$P(y|\Theta) = \prod_{n=1}^N p_n(\mathbf{y}_n|\theta_n). \quad (2.120)$$

W konsekwencji, zgodnie z Uwagą zawartą na końcu Rozdziału 2.3.2.1, dla każdego wyróżnionego parametru θ_n^{ν} wariancja $\sigma^2(\hat{\theta}_n^{\nu})$ jego estymatora (2.110) jest związana z informacją Fishera $I_{n\nu}$ parametru θ_n^{ν} , poprzez nierówność (2.56):

$$\sigma^2(\hat{\theta}_n^{\nu}) \geq \frac{1}{I_{n\nu}}, \quad (2.121)$$

będącą szczególnym uogólnieniem nierówności informacyjną Rao-Cramera (2.29), gdzie $I_{n\nu}$ jest informacją Fishera w pojedynczym (z definicji) kanale informacyjnym.

W takiej sytuacji pojemność informacyjna I zdefiniowana ogólnie w (2.116) przyjmuje postać⁸:

$$I = \sum_{n=1}^N \sum_{\nu} I_{n\nu} = \sum_{n=1}^N \int dy P(y|\Theta) \sum_{\nu} \left(\frac{\partial \ln P(y|\Theta)}{\partial \theta_{n\nu}} \frac{\partial \ln P(y|\Theta)}{\partial \theta_n^{\nu}} \right), \quad (2.123)$$

gdzie (z powodów wyjaśnionych dalej) wprowadzono rozróżnienie pomiędzy współrzędnymi tzw. *kowariantnymi* $\theta_{n\nu}$ i *kontrawariantnymi* θ_n^{ν} parametru. Związek pomiędzy współrzędnymi kowariantnymi i kontrawariantnymi, tak dla wartości losowego wektora położenia jak i dla odpowiednich wartości oczekiwanych, jest następujący:

$$\mathbf{y}_{n\nu} = \sum_{\mu=0}^3 \eta_{\nu\mu} \mathbf{y}_n^{\mu}, \quad \theta_{n\nu} = \sum_{\mu=0}^3 \eta_{\nu\mu} \theta_n^{\mu}, \quad (2.124)$$

gdzie $(\eta_{\nu\mu})$ jest tensorem metrycznym przestrzeni bazowej \mathcal{Y} . W przypadku wektorowego indeksu Minkowskiego $\nu = 0, 1, 2, 3, \dots$ przyjmujemy następującą postać tensora metrycznego:

$$\eta_{\nu\mu} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & \ddots \end{pmatrix} \quad (2.125)$$

lub w skrócie $(\eta_{\nu\mu}) = \text{diag}(1, -1, -1, -1, \dots)$. Symbol “diag” oznacza macierz diagonalną z niezerowymi elementami na przekątnej głównej oraz zerami poza nią. Natomiast dla Euklidesowego indeksu wektorowego $\nu = 1, 2, 3, \dots$, tensor metryczny ma postać:

$$(\eta_{\nu\mu}) = \text{diag}(1, 1, 1, \dots) . \quad (2.126)$$

Uwaga 1: W pomiarze wybranej ν -tej współrzędnej położenia nie można wykluczyć odchyżeń (fluktuacji) wartości współrzędnych do niej ortogonalnych. Oznacza to, że wartość oczekiwana ν -tej współrzędnej położenia nie jest w (2.117) liczona z jakiegoś rozkładu typu $p(\mathbf{y}^{\nu})$ lecz musi być liczona z łącznego rozkładu $p(\mathbf{y})$ dla wszystkich współrzędnych \mathbf{y}^{ν} . W konsekwencji, w przypadku zmiennych położenia przestrzennego i ich parametrów naturalnych określonych w (2.117) całkowanie w (2.123) nie może zostać sfaktoryzowane ze względu na współrzędną wektorową ν .

Uwaga 2: Faktu zależności statystycznej zmiennych położenia przestrzennego dla różnych indeksów ν nie należy mylić z ich niezależnością analityczną, którą posiadają, tzn. zmienne Y są tzw. zmiennymi Fisher’owskimi, dla których:

$$\frac{\partial \mathbf{y}^{\nu}}{\partial \mathbf{y}^{\mu}} = \delta_{\mu}^{\nu} . \quad (2.127)$$

⁸Lokalne własności funkcji wiarygodności $P(y|\Theta)$ opisuje obserwowana macierz informacji Fishera:

$$\mathbf{iF} = \left(-\frac{\partial^2 \ln P(\Theta)}{\partial \theta_n^{\nu} \partial \theta_n^{\mu}} \right), \quad (2.122)$$

która pozostaje symetryczna i dodatnio określona. Jak wiemy, zadaje ona geometryczną strukturę na $S_{N \times 4}$ nazywaną metryką Fishera [12].

Nierówność Rao-Cramera okazuje się niezmiennicza ze względu na podstawowe transformacje [7, 23]. Istotnie, zgodnie z Uwagą 1 właściwym pomiarem niezależnym od przyjętego układu współrzędnych jest pomiar kwadratu długości $\sum_{\nu=0}^3 \mathbf{y}_\nu \mathbf{y}^\nu$ a nie pojedynczej współrzędnej \mathbf{y}^ν . Zatem minimalnym uogólnieniem “jednokanałowej” nierówności Rao-Cramera (2.121), potrzebnym z punktu widzenia przeprowadzanego pomiaru jest nierówność:

$$I_{S_n} := \sum_{\nu=0}^3 \frac{\eta^{\nu\nu}}{\sigma^2(\hat{\theta}_i^\nu)} \leq \int dy P(y|\Theta) \sum_{\nu=0}^3 \frac{\partial \ln P(y|\Theta)}{\partial \theta_{n\nu}} \frac{\partial \ln P(y|\Theta)}{\partial \theta_n^\nu} =: I_n, \quad n = 1, \dots, N \quad (2.128)$$

gdzie I_{S_n} jest informacją Stama czaso-przestrzennych kanałów n -tego pomiaru w próbce, natomiast I_n jest pojemnością informacyjną tych kanałów. Tensor $(\eta^{\nu\mu}) = \text{diag}(1, -1, -1, -1, \dots)$ jest dualny do $(\eta_{\nu\mu})$, tzn. $\sum_{\mu=0}^3 \eta_{\nu\mu} \eta^{\gamma\mu} = \delta_\nu^\gamma$, przy czym δ_ν^γ jest deltą Kroneckera.

Ostatecznie w przestrzeni bazowej \mathcal{Y} Minkowskiego, otrzymujemy w sytuacji estymowania parametru położenia, że informacja Stama I_S dla N -wymiarowej próby powinna być zdefiniowana następująco:

$$I_S \equiv \sum_{n=1}^N \sum_{\nu=0}^3 \frac{\eta^{\nu\nu}}{\sigma^2(\hat{\theta}_i^\nu)}. \quad (2.129)$$

Informacja (2.129) spełnia uogólnioną relację (2.115):

$$I_S = \sum_{n=1}^N I_{S_n} \leq I := \sum_{n=1}^N I_n, \quad (2.130)$$

gdzie I :

$$I = \sum_{n=1}^N I_n, \quad (2.131)$$

jest pojemnością informacyjną układu.

Uwaga o indeksie próby: Indeks próby n jest najmniejszym indeksem kanału informacyjnego, w którym dokonywany jest pomiar. Tzn. gdyby indeks próby można było dodatkowo “zaindeksować”, np. indeksem czaso-przestrzennym, wyznaczając podkanały, to i tak nie możnaby dokonać pomiaru tylko w jednym z tak wyznaczonych podkanałów (nie dokonując go równocześnie w pozostałych podkanałach posiadających indeks próby n).

Uwaga: Dla metryki Minkowskiego jak w (2.125) nie dokonujemy estymacji w przypadku gdy którakolwiek z I_{S_n} , dla $n = 1, 2, \dots, N$ jest ujemna. W przyszłości (Rozdział 4.3) okaże się, że I w (2.131) jest nieujemnie określona dla teorii pola bez tachionów. W każdym konkretnym przypadku estymacji w przestrzeni Minkowskiego należałoby ten fakt sprawdzać. Natomiast estymacja w przestrzeni Euklidesowej zawsze prowadzi do nieujemnej określoności zarówno dla I_{S_n} jak i I_n .

Uwaga o symetrii: Wydaje się oczywiste z punktu widzenia pomiaru, że błąd estymacji, a co za tym idzie, wewnętrzna dokładność estymacji kanałów składających się na oszacowanie długości cztero-wektora powinna być niezależna od układu współrzędnych. Dlatego I_{S_n} określona

w (2.128) jest niezmiennicza, ze względu na transformację Lorentz'a (pchnięcia i obroty). Jeśli chodzi o pojemność informacyjną I_n , to w przypadku niezależności pomiarów w próbie, jest ona również niezmiennicza ze względu na transformację Lorentz'a w przestrzeni z metryką Minkowskiego (czy transformację Galileusza w przestrzeni Euklidesowej), o ile niezmiennicze jest każde I_n .⁹ Ten warunek niezmienniczości I okazuje się mocniejszy niż warunek łącznej niezmienniczości wewnętrznej dokładności estymacji dla czaso-przestrzennych kanałów n -tego pomiaru [7]. Jednak oba warunki schodzą się gdy w nierówności Rao-Cramera (2.128) osiągnięta jest równość. Więcej na temat niezmienniczości DORC w nierówności Rao-Cramera ze względu na przesunięcie, odbicie przestrzenne, obroty i transformację afiniczną oraz transformacje unitarne można znaleźć w [23].

Na koniec zauważmy, że sumowanie w (2.123) przebiega od $n = 1$ do $n = N$. Każdy n -ty wyraz w sumie wnosi analityczny wkład jako stopień swobody dla I . O ile dodane stopnie swobody nie wpływają na już istniejące, to ponieważ każdy (cały) wyraz w sumie po n jest nieujemny to informacja I ma tendencję do wzrostu wraz ze wzrostem N . Kryterium minimalizacji I ze względu na N posłużyło Friedenowi i Sofferowi jako dodatkowy warunek przy konstrukcji np. równań ruchu. Nie znaczy to, że modele z większym N zostały automatycznie wykluczone, tylko, że im większe jest N tym więcej stopni swobody wchodzi do opisu obserwowanego zjawiska, i opisywane zjawisko jest bardziej złożone. Zagadnienie to omówimy w przykładach w dalszej części skryptu.

2.7.2 Podstawowe założenie fizyczne podejścia Friedena-Soffera

Jak w Rozdziale 2.7.1, rozważmy zmienną losową Y przyjmującą wartość \mathbf{y} , która jako położenie układu jest punktem zbioru \mathcal{Y} . Może to być punkt czaso-przestrzenny przestrzeni Minkowskiego, co ma miejsce w rozważaniach związanych z opisem układu np. w mechanice falowej. Wartości $\mathbf{y} \equiv (\mathbf{y}^\nu)_{\nu=0}^3 \in \mathcal{Y} \equiv \mathbb{R}^4$, są realizowane zgodnie z rozkładem $p(\mathbf{y})$ właściwym dla układu¹⁰.

Niech dane $y \equiv (\mathbf{y}_n)_{n=1}^N = (\mathbf{y}_1, \dots, \mathbf{y}_N)$ są realizacjami próby dla położenia układu, gdzie $\mathbf{y}_n \equiv (\mathbf{y}_n^\nu)_{\nu=0}^3$. Zgodnie z założeniem zaproponowanym przez Frieden'a i Soffer'a [7], ich zebranie następuje przez sam układ w zgodzie z rozkładami gęstości prawdopodobieństwa, $p_n(\mathbf{y}_n|\theta_n)$, gdzie $n = 1, \dots, N$. Każdy z tych rozkładów jest punktem zbioru $S = \{p_n(\mathbf{y}_n|\theta_n)\}$ parametryzowanego przez naturalny parametr, tzn. przez wartość oczekiwaną $\theta_n \equiv (\theta_n^\nu)_{\nu=0}^3 = E(Y_n)$ jak w (2.118). Dlatego też wymiary próby y i wektora parametrów $\Theta \equiv (\theta_n^\nu)_{n=1}^N$ są takie same¹¹. Dlatego, ponieważ y jest $N \times 4$ -wymiarową zmienną losową, zatem zbiór $S_{N \times 4} = \{p_n(y|\Theta)\}$ jest rozmaitością w przestrzeni statystycznej \mathcal{S} , gdzie parametry $(\theta_n^\nu)_{n=1}^N$ tworzą $N \times 4$ -wymiarowy lokalny

⁹Pojemność informacyjna:

$$I_n = \int dy P(y|\Theta) \sum_{\nu=0}^3 \mathbf{iF}_{nn\nu}^\nu, \quad (2.132)$$

jest niezmiennicza ze względu na gładkie odwracalne odwzorowania, $Y \rightarrow X$, gdzie X jest nową zmienną [19]. Jest ona również niezmiennicza ze względu na odbicia przestrzenne i czasowe.

¹⁰Ale przy wyprowadzaniu równań generujących (rozkład) w fizyce statystycznej \mathbf{y} może być np. wartością energii układu [7] i wtedy $\mathbf{y} \equiv \mathbf{E} \in \mathcal{Y} \equiv \mathbb{R}$.

¹¹Jednakże przypomnijmy, że w ogólnym przypadku estymacji, wymiar wektora parametrów $\Theta \equiv (\theta_i)_1^k$ oraz próby $y \equiv (\mathbf{y}_n)_1^N$ może być inny.

układ współrzędnych na $S_{N \times 4}$. Podsumujmy treść powyższego założenia fizycznego następująco: *Układ próbkuje dostępną mu czasoprzestrzeń, "zbierając dane i dokonując analizy statystycznej" zgodnie z zasadami informacyjnymi wprowadzonymi w poprzednim rozdziale.*

Uwaga: Jednak pamiętajmy, że estymacji dokonuje tylko człowiek, zatem na metodę EFI należy patrzeć tylko jak na *pewien model analizy statystycznej*.

Interesująca nas statystyczna procedura estymacyjna dotyczy wnioskowania o $p_n(\mathbf{y}_n|\theta_n)$ na podstawie danych y z wykorzystaniem funkcji wiarygodności $P(y|\Theta)$. Gdyby traktować ją jako funkcję próby y , to jest ona proporcjonalna do łącznej gęstości prawdopodobieństwa realizacji $N \times 4$ wartości \mathbf{y}_n (otrzymanego) zbioru danych. Gdy dane te są zbierane niezależnie, to wiarygodność próby y jest zadana w postaci formuły faktoryzacyjnej przyjmującej (zazwyczaj) postać [13]:

$$P(\Theta) \equiv P(y|\Theta) = \prod_{n=1}^N p_n(\mathbf{y}_n|\theta_n), \quad (2.133)$$

gdzie θ_m dla $m \neq n$ nie ma wpływu na \mathbf{y}_n .

Centralna część pracy Frieden'a i Soffer'a związana jest z transformacją postaci pojemności informacyjnej I zadanej równaniem (2.123) oraz (2.131):

$$I = \sum_{n=1}^N I_n = \sum_{n=1}^N \int dy P(y|\Theta) \sum_{\nu=0}^3 \left(\frac{\partial \ln P(y|\Theta)}{\partial \theta_{n\nu}} \frac{\partial \ln P(y|\Theta)}{\partial \theta_n^\nu} \right) \quad (2.134)$$

do tzw. postaci kinematycznej wykorzystywanej w teorii pola oraz fizyce statystycznej.

Założmy, że dane zbierane przez układ $y = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N)$ są uzyskane niezależnie tak, że łączny rozkład prawdopodobieństwa dla próby faktoryzuje się na rozkłady brzegowe:

$$P(y|\Theta) = \prod_{n=1}^N p_n(\mathbf{y}_n|\Theta) = \prod_{n=1}^N p_n(\mathbf{y}_n|\theta_n), \quad (2.135)$$

gdzie w ostatniej równości skorzystano z założenia, że parametr θ_k dla $k \neq n$ nie ma wpływu na rozkład zmiennej Y_n . Wprowadźmy oznaczenie $dy := d^4\mathbf{y}_1 \dots d^4\mathbf{y}_N$, gdzie $d^4\mathbf{y}_n = dy_n^0 dy_n^1 dy_n^2 dy_n^3$. Rachunek analogiczny jaki doprowadził z (2.90) do (2.100) wygląda teraz następująco. Przekształćmy pochodną $\ln P$ w (2.134) do postaci:

$$\frac{\partial \ln P(y|\Theta)}{\partial \theta_{n\nu}} = \frac{\partial}{\partial \theta_{n\nu}} \sum_{n=1}^N \ln p_n(\mathbf{y}_n|\theta_n) = \frac{1}{p_n(\mathbf{y}_n|\theta_n)} \frac{\partial p_n(\mathbf{y}_n|\theta_n)}{\partial \theta_{n\nu}}. \quad (2.136)$$

Pamiętając o unormowaniu rozkładów brzegowych, $\int_{\mathbf{y}} d^4\mathbf{y} p_n(\mathbf{y}_n|\theta_n) = 1$, otrzymujemy postać pojemności informacyjnej:

$$I = \sum_{n=1}^N \int d^4\mathbf{y}_n \frac{1}{p_n(\mathbf{y}_n|\theta_n)} \sum_{\nu=0}^3 \left(\frac{\partial p_n(\mathbf{y}_n|\theta_n)}{\partial \theta_{n\nu}} \frac{\partial p_n(\mathbf{y}_n|\theta_n)}{\partial \theta_n^\nu} \right), \quad (2.137)$$

będącą uogólnieniem (2.100).

W końcu przejdźmy do amplitud $q_n(\mathbf{y}_n|\theta_n)$ określonych jak w (2.75):

$$p_n(\mathbf{y}_n|\theta_n) = q_n^2(\mathbf{y}_n|\theta_n). \quad (2.138)$$

Proste rachunki dają wtedy:

$$I = 4 \sum_{n=1}^N \int d^4 \mathbf{y}_n \sum_{\nu=0}^3 \left(\frac{\partial q_n(\mathbf{y}_n | \theta_n)}{\partial \theta_{n\nu}} \frac{\partial q_n(\mathbf{y}_n | \theta_n)}{\partial \theta_n^\nu} \right), \quad (2.139)$$

czyli prawie kluczową postać pojemności informacyjnej dla rachunku metody EFI Friedena-Soffera. “Jedyne” co trzeba zrobić, to przejść od podprzestrzeni statystycznej \mathcal{S} dla amplitud danych pomiarowych \mathbf{y}_n z bazą parametryczną θ_n^ν do przestrzeni amplitud dla przesunięć (fluktuacji) od wartości oczekiwanych θ_n tych danych określonych na przestrzeni bazowej \mathcal{Y} . Poświęćmy temu Rozdział 3.3.

Rozdział 3

Zasady informacyjne

3.1 Estymacja w statystyce klasycznej a estymacja fizyczna. Postawienie problemu

W dotychczasowej analizie przedstawiona została MNW w statystyce. Polega ona na estymacji parametrów pewnego zadanego rozkładu. Na przykład w analizie regresji na podstawie pewnej wcześniejszej wiedzy na temat zachowania się zmiennej objaśnianej, zakresu wartości jakie może przyjmować oraz jej charakteru (ciągła czy dyskretna) postulujemy warunkowy rozkład i model regresji, a następnie konstruujemy funkcję wiarygodności, którą maksymalizując otrzymujemy estymatory parametrów strukturalnych modelu. Opracowanie skutecznego algorytmu znajdowania estymatorów MNW oraz ich odchyłeń standardowych jest centralnym problemem np. w rutynowych aplikacjach służących do analizy uogólnionych regresyjnych modeli liniowych. W analizie tej najważniejszym wykorzystywanym algorytmem jest ogólny algorytm metody iteracyjnie ważonych najmniejszych kwadratów a jedną z jego głównych analitycznych procedur jest procedura Newton-Raphson'a [13, 15].

Niech parametr wektorowy $\Theta \equiv (\theta_n)_1^N$ jest zbiorem wartości oczekiwanych zmiennej losowej położenia układu w N pomiarach, jak to przyjęliśmy w Rozdziale 2.7. Przypomnijmy więc, że MNW jest wtedy skoncentrowana na układzie N równań wiarygodności [1]:

$$S(\Theta) |_{\Theta=\hat{\Theta}} \equiv \frac{\partial}{\partial \Theta} \ln P(\Theta) |_{\Theta=\hat{\Theta}} = 0, \quad (3.1)$$

których rozwiązanie daje N elementowy zbiór $\hat{\Theta} \equiv (\hat{\theta}_n)_1^N$ estymatorów parametrów. Tzn. układ równań wiarygodności tworzy N warunków na estymatory parametrów, które maksymalizują wiarygodność próbki.

Estymacja w fizyce musi się rozpocząć na wcześniejszym etapie. Wychodząc od zasady informacyjnych, którym poświęcony będzie kolejny rozdział, estymujemy odpowiednie dla opisywanego zagadnienia fizycznego równania ruchu, których rozwiązanie daje odpowiedni rozkład wraz z parametrami. Tak więc zastosowanie zasad informacyjnych nałożonych na funkcję wiarygodności zamiast MNW stanowi o podstawowej różnicy pomiędzy analizą statystyczną wykorzystywaną w konstrukcji modeli fizycznych a statystyką klasyczną. Oczywiście oznacza to, że informacja Fishera zdefiniowana poprzednio na przestrzeni statystycznej \mathcal{S} musi zostać związana z bazową przestrzenią \mathcal{Y} przestrzeni próby, tak aby można ją wykorzystać do konstrukcji równań ruchu.

3.1.1 Strukturalna zasada informacyjna. Metoda EFI

Poniższe rozważania prezentują analizę, leżącą u podstaw strukturalnej zasady informacyjnej [10]. Ta z kolei leży u podstaw metody estymacji statystycznej EFI zaproponowanej przez Friedena i Soffera [7].

Niech V_Θ jest przestrzenią parametru Θ , tzn. $\Theta \in V_\Theta$. Wtedy logarytm funkcji wiarygodności $\ln P : V_\Theta \rightarrow \mathbb{R}$ jest funkcją określoną na przestrzeni V_Θ o wartościach w zbiorze liczb rzeczywistych \mathbb{R} . Rozpocznijmy od rozwinięcia $\ln P(\hat{\Theta})$ w szereg Taylora wokół prawdziwej wartości Θ a następnie scałkujemy po przestrzeni próby z miarą $dy P(\Theta)$, gdzie jak zwykle stosujemy oznaczenie $dy \equiv d^N \mathbf{y}$. W wyniku otrzymujemy pierwsze równanie estymacji modeli metody EFI:

$$\begin{aligned} & \int dy P(\Theta) \left(\ln \frac{P(\hat{\Theta})}{P(\Theta)} - R_3 - \sum_{n=1}^N \frac{\partial \ln P(\Theta)}{\partial \theta_n} (\hat{\theta}_n - \theta_n) \right) \\ &= \frac{1}{2} \int dy P(\Theta) \sum_{n,n'=1}^N \frac{\partial^2 \ln P(\Theta)}{\partial \theta_{n'} \partial \theta_n} (\hat{\theta}_n - \theta_n) (\hat{\theta}_{n'} - \theta_{n'}), \end{aligned} \quad (3.2)$$

gdzie $P(\Theta) \equiv P(\hat{\Theta})|_\Theta$, a R_3 jest resztą rozwinięcia trzeciego rzędu. Wyrażenie po lewej stronie (3.2) ma postać zmodyfikowanej entropii względnej. Zdefiniujmy (obserwowaną) strukturę układu \mathfrak{tF} w następujący sposób:

$$\mathfrak{tF} \equiv \ln \frac{P(\hat{\Theta})}{P(\Theta)} - R_3. \quad (3.3)$$

Następnie, definiując $\tilde{\mathcal{Q}}$ jako:

$$\tilde{\mathcal{Q}} = \int dy P(\Theta) \left(\mathfrak{tF} - \sum_{n=1}^N \frac{\partial \ln P}{\partial \theta_n} (\hat{\theta}_n - \theta_n) \right) \quad (3.4)$$

otrzymujemy równanie będące pierwotną formą strukturalnej zasady informacyjnej:

$$-\tilde{\mathcal{Q}} = \tilde{I} \equiv \frac{1}{2} \int dy P(\Theta) \sum_{n,n'=1}^N \left(-\frac{\partial^2 \ln P}{\partial \theta_{n'} \partial \theta_n} \right) (\hat{\theta}_n - \theta_n) (\hat{\theta}_{n'} - \theta_{n'}). \quad (3.5)$$

Postulując ważność równania (3.5) na poziomie mikroskopowym, tzn. pod całką, otrzymujemy:

$$\Delta_{LHS} \equiv \sum_{n=1}^N 2 \frac{\partial \ln P}{\partial \theta_n} (\hat{\theta}_n - \theta_n) - \sum_{n=1}^N 2 \frac{\mathfrak{tF}}{N} = \sum_{n,n'=1}^N \mathfrak{iF}_{nn'} (\hat{\theta}_n - \theta_n) (\hat{\theta}_{n'} - \theta_{n'}) \equiv \Delta_{RHS}, \quad (3.6)$$

gdzie \mathfrak{iF} jest znaną już z Rozdziału 2.1 wzór I obserwowana (obserwowaną) macierzą informacyjną Fishera:

$$\mathfrak{iF} = \left(-\frac{\partial^2 \ln P(\Theta)}{\partial \theta_{n'} \partial \theta_n} \right) \quad (3.7)$$

która jako macierz odwrotna do macierzy kowariancji, jest symetryczna i dodatnio określona (Rozdział 2.1.1), tzn. zakładamy, że $\ln P$ jest wypukła w otoczeniu prawdziwej wartości parametru Θ . Oznacza to, że istnieje ortogonalna macierz U taka, że Δ_{RHS} w (3.6), a zatem również Δ_{LHS} , może być zapisane w tzw. postaci normalnej [24]:

$$\sum_{n,n'=1}^N \mathfrak{iF}_{nn'} (\hat{\theta}_n - \theta_n) (\hat{\theta}_{n'} - \theta_{n'}) \equiv \Delta_{RHS} = \Delta_{LHS} = \sum_{n=1}^N m_n \hat{v}_n^2, \quad (3.8)$$

gdzie \hat{v}_n są pewnymi funkcjami estymatorów $\hat{\theta}_n$, a m_n są elementami dodatnio określonej macierzy \mathbf{mF} (otrzymanymi dla Δ_{LHS}), która z powodu równości (3.8) musi być równa macierzy diagonalnej otrzymanej dla Δ_{RHS} , tzn.:

$$\mathbf{mF} = D^T U^T \mathbf{iF} U D . \quad (3.9)$$

Macierz D jest diagonalną macierzą skalującą o elementach $d_n \equiv \sqrt{\frac{m_n}{\lambda_n}}$ gdzie λ_n są wartościami własnymi macierzy \mathbf{iF} .

Równość (3.9) pozwala w końcu na zapisanie mikroskopowej strukturalnej zasady informacyjnej Frieden'a:

$$\mathbf{qF} + \mathbf{iF} = 0 , \quad (3.10)$$

gdzie

$$\mathbf{qF} = -U (D^T)^{-1} \mathbf{mF} D^{-1} U^T , \quad (3.11)$$

nazwijmy (*obserwowaną*) *macierzą struktury*.

Istnienie postaci normalnej (3.8) jest bardzo mocnym warunkiem, który czyni całą analizę Friedena-Soffera w ogóle możliwą. Istnieją dwa szczególne przypadki, które prowadzą do prostych realizacji fizycznych. Pierwszy z nich jest związany z założeniem, że $\mathbf{tF} = 0$ i wtedy z (3.6) otrzymuje się postać "równania master" (porównaj dalej (3.54)). W przypadku tym:

$$\mathbf{mF} = \text{diag} \left(2 \frac{\partial \ln P}{\partial \theta_n} \right) , \quad \hat{v}_n = \sqrt{\hat{\theta}_n - \theta_n} , \quad \mathbf{tF} = 0 , \quad \text{oraz} \quad d_n = \sqrt{2 \frac{\partial \ln P}{\partial \theta_n} / \lambda_n} , \quad (3.12)$$

co oznacza, że nie istnieje złożona struktura układu.

Natomiast drugi przypadek związany jest z założeniem, że rozkład jest *regularny* [13]. Wtedy, zakładając dodatkowo, że dla wszystkich $n = 1, \dots, N$ zachodzi $\frac{\partial \ln P}{\partial \theta_n} = 0$, z równania (3.6) widzimy, że:

$$\mathbf{mF} = (2 \delta_{nn'}) , \quad \hat{v}_n = \sqrt{\frac{\mathbf{tF}}{N}} , \quad \text{oraz} \quad d_n = \sqrt{2/\lambda_n} . \quad (3.13)$$

Sumując po wszystkich elementach zarówno obserwowanej macierzy informacyjnej Fishera \mathbf{iF} jak i (obserwowanej) macierzy struktury \mathbf{qF} , oraz całkując obie strony równania (3.10) po przestrzeni próby z miarą $d^N \mathbf{y} P(\Theta)$, otrzymujemy całkową postać *informacyjnej zasady strukturalnej*:

$$Q + I = 0 , \quad (3.14)$$

gdzie I jest uogólnieniem pojemności informacyjnej Fishera (2.116) (por. (2.20)):

$$I = \int dy P(\Theta) \sum_{n,n'=1}^N (\mathbf{iF})_{nn'} , \quad (3.15)$$

natomiast Q jest informacją strukturalną (SI):

$$Q = \int dy P(\Theta) \sum_{n,n'=1}^N (\mathbf{qF})_{nn'} . \quad (3.16)$$

Pierwotnie, w innej, informatycznej formie i interpretacji, zasada ta została zapostulowana w [7]. Powyższa, fizyczna postać zasady strukturalnej (3.14) została zapostulowana w [9] a następnie wyprowadzona jak to przedstawiono powyżej w [10]. Zasada ta jest równaniem strukturalnym wielu współczesnych modeli fizycznych wyprowadzanych metodą EFI.

3.1.1.1 I oraz Q dla parami niezależnych zmiennych położeniowych próby

Rozważmy jeszcze postać IS wyrażoną w amplitudach w szczególnym przypadku zmiennych Y_n parami niezależnych. W takim przypadku zarówno amplituda q_n nie zależy od \mathbf{y}_n dla $n' \neq n$, czyli ma postać $q_n(\mathbf{y}_n)$, jak również (\mathfrak{IF}) jest diagonalna, tzn. ma postać:

$$(\mathfrak{IF})_{nn'} = \delta_{nn'} \mathfrak{IF}_n , \quad (3.17)$$

gdzie $\delta_{nn'}$ jest deltą Kroneckera. W takim razie, zgodnie z (3.9) oraz (3.11) obserwowana macierz strukturalna jest diagonalna i jej ogólna postać jest następująca:

$$(\mathfrak{IF})_{nn'} = \delta_{nn'} \mathfrak{F}_n \left(q_n(\mathbf{y}_n), q_n^{(r)}(\mathbf{y}_n) \right) \equiv \mathfrak{F}_n (q_n(\mathbf{y}_n)) , \quad (3.18)$$

tzn. nie zależy od amplitud $q_{n'}(\mathbf{y}_{n'})$ i jej pochodnych dla $n' \neq n$. Powyżej $q_n^{(r)}(\mathbf{y}_n)$ oznaczają pochodne rzędu $r = 1, 2, \dots$. Zobaczymy, że dla teorii pola w \mathfrak{F}_n pojawią się pochodne co najwyżej pierwszego rzędu. Fakt ten wynika stąd, że swobodne pola rangi N , z którymi będziemy mieli do czynienia, będą spełniały równanie Kleina-Gordona.

Uwaga: Oznaczenie \mathfrak{F}_n jak również jawne zaznaczenie w argumencie obserwowanej IS tylko amplitudy $q_n(\mathbf{y}_n)$, będą stosowane w dalszej części tego skryptu.

Podsumowując, pojemność informacyjna (3.15) przyjmuje w rozważanym przypadku formę:

$$I = \int dy P(\Theta) \sum_{n=1}^N \mathfrak{IF}_n , \quad (3.19)$$

która ma postać (2.123), natomiast informacja strukturalna (3.16):

$$Q = \int dy P(\Theta) \sum_{n=1}^N \mathfrak{F}_n (q_n(\mathbf{y}_n)) . \quad (3.20)$$

Ze względu na unormowanie rozkładów brzegowych $\int d^4\mathbf{y}_n p_n(\mathbf{y}_n|\theta_n) = 1$, powyższą postać Q można zapisać następująco:

$$Q = \sum_{n=1}^N \int d^4\mathbf{y}_n p_n(\mathbf{y}_n|\theta_n) \mathfrak{F}_n (q_n(\mathbf{y}_n)) . \quad (3.21)$$

Ważna kinematyczna postać I zostanie wprowadzona w Rozdziale 3.3, natomiast postacie Q będą pojawiały się w toku rozwiązywania konkretnych fizycznych problemów.

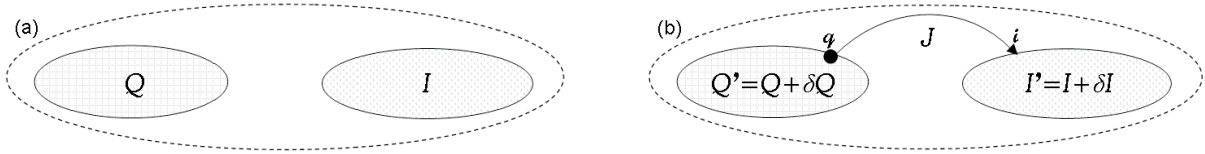
3.2 Przepływ informacji

Pojemność informacyjna I jest infinitezymalnym typem entropii Kulback-Leibler'a (Rozdział 2.6.1) wzór (2.107). W statystycznej estymacji KL służy jako narzędzie analizy wyboru modelu [25, 26], o czym możemy się przekonać, zauważając, że jest ona związana z wartością oczekiwaną statystyki ilorazu wiarygodności (1.44) wprowadzonej w Rozdziale 1.2.2, właśnie w celu porównywania wiarygodności modeli.

Chociażby z tego powodu, pojawia się więc przypuszczenie, że I mogłaby (po nałożeniu jak się okazuje strukturalnej i wariacyjnej zasady informacyjnej [7, 9]), stać się podstawą równań ruchu (lub równań generujących rozkład) układu fizycznego. Równania te miałyby być najlepsze z punktu widzenia zasad informacyjnych, co jest sednem metody EFI Friedena-Soffera.

Sformułujmy wstępnie jej główną statystyczną myśl: Próbkowanie czaso-przestrzeni następuje przez sam układ, nawet wtedy gdy on sam nie jest poddany rzeczywistemu pomiarowi. Tzn. układ dokonuje próbkowania czaso-przestrzeni używając charakterystycznego, swojego własnego pola (i związanej z nim amplitudy) rangi N , która jest wymiarem próby, próbkując swoimi kinematycznymi "Fisherowskich" stopni swobody przestrzeni położeń mu dostępną. Przejście od postaci statystycznej informacji Fishera (2.123) do jej reprezentacji kinematycznej zostanie omówione poniżej w Rozdziale 3.3.

Rozważmy następujący, informacyjny schemat układu. Zanim nastąpi pomiar (który dokonuje sam układ), układ ma swoją pojemność informacyjną I zawartą w kinematycznych stopniach swobody, oraz informację strukturalną Q układu zawartą w swoich strukturalnych stopniach swobody, jak to przedstawiono symbolicznie na Rysunku 3.2. "Włączmy" teraz pomiar, pod-



Rysunek 3.1: Panel: (a) The system before the measurement: Q is the SI of the system contained in the structural degrees of freedom and I is the IF of the system contained in the kinematical degrees of freedom. (b) The system after the measurement: Q' is the SI and I' is the IF of the system; $\delta Q = Q' - Q \leq 0$ and $\delta I = I' - I \geq 0$ as the transfer of the information (TI) in the measurement takes place with $J \geq 0$. In the ideal measurement $\delta I = -\delta Q$.

czas którego przepływ (transfer) informacji (TI) przebiega zgodnie z następującymi zasadami (Rysunku 3.2):

$$J \geq 0 \quad \text{zatem} \quad \delta I = I' - I \geq 0, \quad \delta Q = Q' - Q \leq 0, \quad (3.22)$$

gdzie I' , Q' są odpowiednio IF oraz SI układu po pomiarze, natomiast J jest dokonanym transferem informacji (TI). Postulujemy, że w pomiarze TI "w punkcie q " jest idealny, co oznacza, że:

$$Q = Q' + J = Q + \delta Q + J \quad \text{zatem} \quad \delta Q = -J. \quad (3.23)$$

Oznacza to, że "w punkcie q " cała zmiana SI jest przekazana.

Z drugiej strony “w punkcie i ” zasada związana z TI jest następująca:

$$I' \leq I + J \quad \text{zatem} \quad 0 \leq \delta I = I' - I \leq J . \quad (3.24)$$

Dlatego:

$$\text{ponieważ} \quad J \geq 0 , \quad \text{zatem} \quad |\delta I| \leq |\delta Q| , \quad (3.25)$$

co jest rozsądnym rezultatem, gdyż w pomiarze może nastąpić utrata informacji. Gdyby i “w punkcie i ” TI był idealny, wtedy cały pomiar byłby idealny, tzn.:

$$\delta Q = -\delta I \Leftrightarrow \text{pomiar idealny} . \quad (3.26)$$

W [9, 16] zostało zapostulowane istnienie nieujemnej addytywnej całkowitej (totalnej) fizycznej informacji (TFI):

$$K = I + Q \geq 0 . \quad (3.27)$$

Wybór intuicyjnego warunku $K \geq 0$ [16], prowadzi do następującej postaci strukturalnej zasady informacyjnej:

$$I + \kappa Q = 0 \quad (3.28)$$

lub

$$I + Q = 0 \quad \text{dla} \quad \kappa = 1 . \quad (3.29)$$

Powyższy warunek został formalnie wyprowadzony w Rozdziale 3.1.1 jako zasada strukturalna (3.14) [10]. Współczynnik κ został nazwany w [7] współczynnikiem efektywności. W praktyce przyjmuje on dwie możliwe wartości:

$$\kappa = 1 \quad \vee \quad \frac{1}{2} . \quad (3.30)$$

Jego znaczenie zostanie omówione w Rozdziale 4. W przypadku określonym w (3.29) otrzymujemy, że K jest równe:

$$K = Q + I = 0 \quad \text{for} \quad \kappa = 1 . \quad (3.31)$$

Drugą zasadą informacyjną jest *zasada wariacyjna* (skalarna). Ma ona postać [9]:

$$\delta K = \delta(Q + I) = 0 \Rightarrow K = Q + I \text{ jest ekstremalne} . \quad (3.32)$$

W końcu zauważmy, że w zgodzie z zapostulowanym zachowaniem się układu w pomiarze, otrzymaliśmy z warunków (3.23) i (3.24), że $\delta I \leq J = -\delta Q$, z czego wynika, że:

$$K' = I' + Q' \leq (I + J) + (Q - J) = I + Q = K \Rightarrow K' \leq K . \quad (3.33)$$

W przypadku (3.26) pomiaru idealnego, $\delta I = -\delta Q$ skąd $K' = K$, co oznacza, że informacja fizyczna TFI pozostaje w tym przypadku niezmienną. Jeśli pomiar idealny byłby wykonany na poziomie próbkowania czaso-przestrzeni przez sam układ, wtedy warunek ten mógłby prowadzić do wariacyjnej zasady informacyjnej (3.32), tzn.:

$$\delta I = -\delta Q \Rightarrow \delta(Q + I) = 0 . \quad (3.34)$$

Chociaż rozumowanie powyższe wydaje się być rozsądne, jednak ściśle mówiąc słuszność przyjęcia zasad informacyjnych, strukturalnej oraz wariacyjnej, powinno wynikać z dwóch rzeczy. Po pierwsze z ich wyprowadzenia a po drugie z ich użyteczności. Wyprowadzenie zasady strukturalnej (dla $\kappa = 1$) zostało pokazane w Rozdziale 3.1.1, natomiast wnioskowanie które doprowadziło do warunku (3.34) oraz sama implikacja wewnątrz niego, może służyć jedynie jako przesłanka słuszności zasady wariacyjnej, dla której do tej pory brak formalnego wyprowadzenia. Natomiast sama użyteczność zasady wariacyjnej w metodzie EFI jest oczywista, bowiem prowadzi ona do owocnego w zastosowaniach równania Eulera-Lagrange'a. A dokładniej, obie zasady (3.28) oraz (3.32) są podstawą metody estymacyjnej EFI. W ich mikroskopowej postaci tworzą układ dwóch równań różniczkowych na amplitudy układu (2.75) wprowadzone w Rozdziale 2.5. Układ ten może być zgodny dając samo-spójne rozwiązanie dla amplitud [7], i prowadząc dla $\kappa = 1$ lub $1/2$ do dobrze znanych modeli teorii pola (Rozdział 4) oraz fizyki statystycznej (Rozdział 5). Ponadto, strukturalna (wewnętrzna) zasada informacyjna (3.28) [9] jest operacyjnie równoważna zapostulowanej przez Frieden'a [7], więc jako wyprowadzona powinna mieć przynajmniej taką samą moc przewidywania jak tamta [7]. Wiele z podstawowych modeli zostało już wyliczonych [7], jednak ich ponowne przeliczenie [16] przy powyżej podanej interpretacji informacji fizycznej K może dać lepsze zrozumienie metody EFI, jej związku z istniejącym już modelowaniem zjawisk w fizyce, ale również jej ograniczeń.

W końcu, w strukturalnej zasadzie informacyjnej ciekawe jest również to, że stanowi ona warunek zerowego podziału dla TFI , który jest dawno poszukiwanym warunkiem zasady ekwipartycji entropii (w tym przypadku infinitezymalnej entropii względnej).

Wspomnieliśmy o tym, że pomysł metody EFI pochodzi od Friedena. Jednak mówiąc w skrócie, Frieden i Soffer [7] podeszli inaczej do informacji strukturalnej. W [7] wprowadzono tzw. informację związaną J , która ma interpretację informacji zawartej w układzie przed pomiarem. Chociaż, aksjomaty Frieden'a są równoważne powyższymi warunkom (3.28) oraz (3.32), o ile $J = -Q$, to jednakże różnica pomiędzy podejściami jest widoczna. A mianowicie, o ile w podejściu Friedena-Soffera układ doświadcza transferu informacji $J \rightarrow I$, mając w każdej chwili czasu tylko jeden z tych typów informacji, o tyle w naszym podejściu system jest charakteryzowany jednocześnie przez I oraz Q w każdej chwili czasu.

3.3 Kinematyczna postać informacji Fishera

Centralna część pracy Frieden'a i Soffer'a związana jest z transformacją postaci pojemności informacyjnej I zadanej równaniem (2.123) oraz (2.131) do tzw. postaci kinematycznej wykorzystywanej w teorii pola oraz fizyce statystycznej. W obecnym rozdziale zaprezentujemy podstawowe założenia, które doprowadziły do konstrukcji kinematycznego członu (całki) działania dla cztero-wymiarowych modeli teorii pola. Przejście to ma następującą postać [7].

Zgodnie z podstawowym założeniem Friedena-Soffera, N -wymiarowa próbka $\mathbf{y}_n \equiv (\mathbf{y}_n^\nu)$ jest pobierana przez układ posiadający rozkład $p_n(\mathbf{y}_n)$, gdzie obok indeksu próby $n = 1, 2, \dots, N$ wprowadzono indeks (czaso)przestrzenny $\nu = (0), 1, 2, 3$. Zgodnie z Rozdziałem 2.5, wzór (2.75), metryka Fishera na (pod)rozmaitości \mathcal{S} prowadzi w naturalny sposób do pojęcia rzeczywistej amplitudy $q(\mathbf{y}_n|\theta_n) \equiv \sqrt{p(\mathbf{y}_n|\theta_n)}$ pola układu. Od razu skorzystano też z zapisu, który sugeruje niezależność rozkładu dla Y_n od θ_m , gdy $m \neq n$. Jak przedstawiliśmy w Rozdziale 2.7.2 pojemność informacyjna (2.134) może zostać zapisana jako (2.139):

$$I \equiv I(\Theta) = 4 \sum_{n=1}^N \int d^4 \mathbf{y}_n \sum_{\nu=0}^3 \left(\frac{\partial q_n(\mathbf{y}_n|\theta_n)}{\partial \theta_{n\nu}} \frac{\partial q_n(\mathbf{y}_n|\theta_n)}{\partial \theta_n^\nu} \right). \quad (3.35)$$

Niech $\mathbf{x}_n \equiv (\mathbf{x}_n^\nu)$ są przesunięciami (np. addytywnymi fluktuacjami) danych $\mathbf{y}_n \equiv (\mathbf{y}_n^\nu)$ od ich wartości oczekiwanych θ_n^ν , tzn.:

$$\mathbf{y}_n^\nu = \theta_n^\nu + \mathbf{x}_n^\nu. \quad (3.36)$$

Przesunięcia \mathbf{x}_n^ν są zmiennymi Fisher'owskimi, spełniając warunek $\frac{\partial \mathbf{x}_n^\nu}{\partial \mathbf{x}^\mu} = \delta_\mu^\nu$, (2.127).

Odwołując się do "reguły łańcuchowej" dla pochodnej:

$$\frac{\partial}{\partial \theta_n^\nu} = \frac{\partial(\mathbf{y}_n^\nu - \theta_n^\nu)}{\partial \theta_n^\nu} \frac{\partial}{\partial(\mathbf{y}_n^\nu - \theta_n^\nu)} = - \frac{\partial}{\partial(\mathbf{y}_n^\nu - \theta_n^\nu)} = - \frac{\partial}{\partial \mathbf{x}_n^\nu}, \quad (3.37)$$

oraz uwzględniając $d^4 \mathbf{x}_n = d^4 \mathbf{y}_n$, co wynika z tego, że parametry θ_n są stałymi, możemy przejść od postaci statystycznej (3.35) do postaci kinematycznej IF :

$$I = 4 \sum_{n=1}^N \int d^4 \mathbf{x}_n \sum_{\nu} \frac{\partial q_n(\mathbf{x}_n)}{\partial \mathbf{x}_{n\nu}} \frac{\partial q_n(\mathbf{x}_n)}{\partial \mathbf{x}_n^\nu}, \quad (3.38)$$

gdzie $d^4 \mathbf{x}_n = d\mathbf{x}_n^0 d\mathbf{x}_n^1 d\mathbf{x}_n^2 d\mathbf{x}_n^3$. W (3.38) wprowadzono oznaczenie:

$$q_n(\mathbf{x}_n) \equiv q_n(\mathbf{x}_n + \theta_n|\theta_n) = q_n(\mathbf{y}_n|\theta_n), \quad (3.39)$$

pozostawiając całą informację o θ_n w indeksie n amplitudy $q_n(\mathbf{x}_n)$.

Zakładając, że zakres zmienności wszystkich \mathbf{x}_n^ν jest dla każdego n taki sam, możemy pominąć indeks n przy tej zmiennej (ale nie przy amplitudzie q_n), otrzymując:

$$I = 4 \sum_{n=1}^N \int d^4 \mathbf{x} \sum_{\nu} \frac{\partial q_n(\mathbf{x})}{\partial \mathbf{x}_\nu} \frac{\partial q_n(\mathbf{x})}{\partial \mathbf{x}^\nu}, \quad (3.40)$$

postać, którą wykorzystamy przy wyprowadzeniu równań generujących fizyki statystycznej [7], ale która została również wykorzystana do wyprowadzenia elektrodynamiki Maxwella metodą EFI [7].

Uwaga: Wymiar próby N jest rangą pola układu zdefiniowanego jako zbiór amplitud $(q_n(\mathbf{x}_n))_{n=1}^N$.

W Rozdziale 3.1.1 pokazaliśmy, że strukturalna zasada informacyjna $I + Q = 0$ jest artefaktem istnienia rozwinięcia $\ln P(\hat{\Theta})$ w szereg Taylora¹ wokół prawdziwej wartości parametru Θ . Obecnie znamy już ogólną postać kinematyczną I części pomiarowej zasady strukturalnej. W metodzie EFI, jej część strukturalna Q ma postać zależną od np. fizycznych więzów nałożonych na układ. Zagadnieniem tym zajmiemy się w kolejnych Rozdziałach 4 oraz 5.

Kolejnym założeniem jest konstrukcja *składowych funkcji falowej* składanych z amplitud² w następujący sposób [7]:

$$\psi_n(\mathbf{x}_{2n-1}, \mathbf{x}_{2n}) \equiv \frac{1}{\sqrt{N}} (q_{2n-1}(\mathbf{x}_{2n-1}) + i q_{2n}(\mathbf{x}_{2n})) , \quad n = 1, \dots, N/2 . \quad (3.41)$$

Powyższa postać jest uogólnieniem konstrukcji Friedena, który tworząc funkcję falową układu złożył n -tą *składową funkcję falową* z amplitud w następujący sposób [7]:

$$\psi_n(\mathbf{x}) \equiv \frac{1}{\sqrt{N}} (q_{2n-1}(\mathbf{x}) + i q_{2n}(\mathbf{x})) , \quad n = 1, \dots, N/2 . \quad (3.42)$$

Dokładniej mówiąc, aby posłużenie się funkcją falową (3.42) miało sens, musi przynajmniej pod całą zachodzić równoważność zmiennych:

$$\mathbf{x}_m \equiv \mathbf{x}_n \equiv \mathbf{x} \quad \text{dla wszystkich } n, m = 1, 2, \dots, N . \quad (3.43)$$

Założenie to całkiem wystarcza przy liczeniu wartości oczekiwanych oraz prawdopodobieństw. Przy założeniu postaci (3.42) dla n -tej składowej³, postać *funkcji falowej* Friedena jest następująca:

$$\psi(\mathbf{x}) \equiv (\psi_n(\mathbf{x}))_{n=1}^{N/2} . \quad (3.44)$$

Zbiór N składowych funkcji falowych ψ_n nazwijmy *funkcją falową układu rangi N* .

Zauważmy, że zachodzą następujące równości:

$$\sum_{n=1}^N q_n^2 = (q_1^2 + q_3^2 + \dots + q_{N-1}^2) + (q_2^2 + q_4^2 + \dots + q_N^2) = \sum_{n=1}^{N/2} (q_{2n-1})^2 + (q_{2n})^2 \quad (3.45)$$

i analogicznie:

$$\sum_{n=1}^N \frac{\partial q_n}{\partial \mathbf{x}_{n\nu}} \frac{\partial q_n}{\partial \mathbf{x}'_n} = \sum_{n=1}^{N/2} \left(\frac{\partial q_{2n-1}}{\partial \mathbf{x}_{n\nu}} \frac{\partial q_{2n-1}}{\partial \mathbf{x}'_n} + \frac{\partial q_{2n}}{\partial \mathbf{x}_{n\nu}} \frac{\partial q_{2n}}{\partial \mathbf{x}'_n} \right) . \quad (3.46)$$

Zakładając dla wszystkich poniższych rozważań słuszność (3.43) przynajmniej pod całą, oraz postać *funkcji falowych* (3.42), dokonajmy następującego ciągu przekształceń dla (3.40):

$$\begin{aligned} I &= 4 \sum_{n=1}^N \int d^4 \mathbf{x} \sum_{\nu} \frac{\partial q_n}{\partial \mathbf{x}_{\nu}} \frac{\partial q_n}{\partial \mathbf{x}'_{\nu}} = 4 \sum_{n=1}^{N/2} \int d^4 \mathbf{x} \sum_{\nu} \left[\frac{\partial q_{2n-1}}{\partial \mathbf{x}_{\nu}} \frac{\partial q_{2n-1}}{\partial \mathbf{x}'_{\nu}} + \frac{\partial q_{2n}}{\partial \mathbf{x}_{\nu}} \frac{\partial q_{2n}}{\partial \mathbf{x}'_{\nu}} \right] \\ &= 4N \sum_{n=1}^{N/2} \int d^4 \mathbf{x} \sum_{\nu} \frac{1}{\sqrt{N}} \frac{\partial (q_{2n-1} - i q_{2n})}{\partial \mathbf{x}_{\nu}} \frac{1}{\sqrt{N}} \frac{\partial (q_{2n-1} + i q_{2n})}{\partial \mathbf{x}'_{\nu}} , \end{aligned} \quad (3.47)$$

¹W [10] była użyta miara $d^N \mathbf{x} P(\Theta)$ zamiast $d^N \mathbf{y} P(\Theta)$. Nie zmienia to jednak dowodu strukturalnej zasady informacyjnej lecz poszerza jego zastosowanie na sytuacje, które nie posiadają niezmienniczości przesunięcia, założenia, które nie było wykorzystane w dowodzie.

²Amplitudy q_n są w przypadku rozkładów ciągłych związane z p_n , które są *gęstościami* prawdopodobieństw.

³O ile nie będzie to prowadziło do nieporozumień będziemy pominieli słowo "składowa".

gdzie w ostatnim przejściu skorzystano z przekształcenia typu:

$$q_{2n-1}^k q_{2n-1}^k + q_{2n}^k q_{2n}^k = (q_{2n-1}^k - i q_{2n}^k)(q_{2n-1}^k + i q_{2n}^k) . \quad (3.48)$$

gdzie indeks k oznacza pochodną rzędu $k = 0, 1, \dots$. Odwołując się do definicji (3.42) funkcji falowej, otrzymujemy:

$$I = 4N \sum_{n=1}^{N/2} \int d^4 \mathbf{x} \sum_{\nu} \frac{\partial \psi_n^*(\mathbf{x})}{\partial \mathbf{x}_{\nu}} \frac{\partial \psi_n(\mathbf{x})}{\partial \mathbf{x}^{\nu}} . \quad (3.49)$$

Pojemność informacyjna (3.49) ma typową postać np. dla relatywistycznej mechaniki falowej, odpowiadającą części kinetycznej całki działania. Dlatego właśnie oczekiwaną informację Fishera nazwał Frieden *informacją kinetyczną*. W [7] użyto jej do wyprowadzenia równań Kleina-Gordona oraz Diraca metodą EFI [7].

Uwaga: Korzystając z twierdzenia o prawdopodobieństwie całkowitym, *gęstość* rozkładu prawdopodobieństwa przesunięcia (lub fluktuacji) w układzie może być zapisana następująco [7]:

$$\begin{aligned} p(\mathbf{x}) &= \sum_{n=1}^N p(\mathbf{x}|\theta_n) r(\theta_n) = \sum_{n=1}^N p_n(\mathbf{x}_n|\theta_n) r(\theta_n) = \frac{1}{N} \sum_{n=1}^N q_n^2(\mathbf{x}_n|\theta_n) \\ &= \frac{1}{N} \sum_{n=1}^N q_n^2(\mathbf{x}) , \end{aligned} \quad (3.50)$$

gdzie skorzystano z założenia, że n -ta wartość oczekiwana θ_n nie ma dla $m \neq n$ wpływu na rozkład przesunięcia \mathbf{x}_m , oraz jak zwykle z postaci amplitudy $q_n^2 = p_n$. Prawdopodobieństwo p_n jest prawdopodobieństwem pojawienia się wartości \mathbf{x}_n zmiennej losowej przesunięcia (lub fluktuacji) z rozkładu generowanego z parametrem θ_n , tzn. ma ono interpretację prawdopodobieństwa warunkowego $p_n(\mathbf{x}_n|\theta_n)$. Funkcję $r(\theta_n) = \frac{1}{N}$ można nazwać funkcją “niewiedzy”, gdyż jej postać jest odzwierciedleniem całkowitego braku wiedzy odnośnie tego, która z N możliwych wartości θ_n pojawi się w konkretnym n -tym z N eksperymentów próby.

W końcu, korzystając z (3.42), (3.46), (3.48) oraz (3.50) widać, że:

$$p(\mathbf{x}) = \sum_{n=1}^{N/2} \psi_n^*(\mathbf{x}) \psi_n(\mathbf{x}) , \quad (3.51)$$

jest gęstością rozkładu prawdopodobieństwa przesunięcia (lub fluktuacji) \mathbf{x} w układzie opisanym funkcją falową (3.44).

Uwaga: W całym powyższym wyprowadzeniu nie użyliśmy podstawowego założenia Friedena-Soffera o *niezmienniczości rozkładu ze względu na przesunięcie*, tzn.:

$$p_n(\mathbf{x}_n) = p_{x_n}(\mathbf{x}_n|\theta_n) = p_n(\mathbf{y}_n|\theta_n) \quad \text{gdzie} \quad \mathbf{x}_n^{\nu} \equiv \mathbf{y}_n^{\nu} - \theta_n^{\nu} . \quad (3.52)$$

gdzie $\mathbf{y}_n \equiv (\mathbf{y}_n^{\nu})$, $\mathbf{x}_n \equiv (\mathbf{x}_n^{\nu})$ oraz $\theta_n \equiv (\theta_n^{\nu})$. Założenie to nie jest potrzebne przy wyprowadzeniu postaci (3.38) pojemności informacyjnej.

Uwaga: Co więcej, *informacja o θ_n musi pozostać w rozkładzie p_n oraz jego amplitudzie q_n* . Wcześniej umówiliśmy się, że indeks n zawiera tę informację. Po umiejscowieniu informacji o θ_n w indeksie n , można w razie potrzeby, wynikającej np. z fizyki zjawiska, zażądać dodatkowo niezmienniczości ze względu na przesunięcie.

3.3.0.2 Postać kinematyczna pojemności zapisana w prawdopodobieństwie

Poniżej podamy postać kinematyczną pojemności zapisaną w (punktowych) prawdopodobieństwach próby. Postać ta jest bardziej pierwotna niż (3.35), chociaż w treści skryptu wykorzystywana jedynie w Dodatku.

Puntem wyjścia jest pojemność (2.137):

$$I \equiv I(\Theta) = \sum_{n=1}^N \int d^4 \mathbf{y}_n \frac{1}{p_n(\mathbf{y}_n | \theta_n)} \sum_{\nu=0}^3 \left(\frac{\partial p_n(\mathbf{y}_n | \theta_n)}{\partial \theta_{n\nu}} \frac{\partial p_n(\mathbf{y}_n | \theta_n)}{\partial \theta_n^\nu} \right).$$

Korzystając z przejścia do addytywnych przesunięć $\mathbf{x}_n \equiv (\mathbf{x}_n^\nu)$, (3.36), oraz z “reguły łańcuchowej” (3.37) dla pochodnej, otrzymujemy (podobnie do (3.38)), następującą *kinematyczną postać pojemności informacyjnej* wyrażoną w prawdopodobieństwach:

$$I = \sum_{n=1}^N \int d^4 \mathbf{x}_n \frac{1}{p_n(\mathbf{x}_n)} \sum_{\nu=0}^3 \left(\frac{\partial p_n(\mathbf{x}_n)}{\partial \mathbf{x}_{n\nu}} \frac{\partial p_n(\mathbf{x}_n)}{\partial \mathbf{x}_n^\nu} \right).$$

gdzie, podobnie jak poprzednio dla amplitud, pozostawiliśmy całą informację o θ_n w *indeksie n rozkładu $p_n(\mathbf{x}_n)$* . W końcu, zakładając, że zakres zmienności wszystkich \mathbf{x}_n^ν jest dla każdego n taki sam, pomijamy indeks n przy tej zmiennej (ale nie przy rozkładzie p_n), otrzymując:

$$I = \sum_{n=1}^N \int d^4 \mathbf{x} \frac{1}{p_n(\mathbf{x})} \sum_{\nu=0}^3 \left(\frac{\partial p_n(\mathbf{x})}{\partial \mathbf{x}_\nu} \frac{\partial p_n(\mathbf{x})}{\partial \mathbf{x}^\nu} \right).$$

Postać tę wykorzystamy w Dodatku jako pierwotną przy wyprowadzeniu elektrodynamiki Maxwella, granicy słabego pola w teorii grawitacji, oraz twierdzenia I fizyki statystycznej.

3.4 Równania master

Podjdźmy nieco inaczej niż w Rozdziale 3.1.1 do problemu estymacji. Rozwińmy $P(\hat{\Theta})$ w szereg Taylora wokół prawdziwej wartości parametru Θ i wycałkujemy po przestrzeni próby, otrzymując:

$$\int dy \left(P(\hat{\Theta}) - P(\Theta) \right) = \int dy \left(\sum_{n=1}^N \frac{\partial P(\Theta)}{\partial \theta_n} (\hat{\theta}_n - \theta_n) + \frac{1}{2} \sum_{n,n'=1}^N \frac{\partial^2 P(\Theta)}{\partial \theta_{n'} \partial \theta_n} (\hat{\theta}_n - \theta_n) (\hat{\theta}_{n'} - \theta_{n'}) + \dots \right), \quad (3.53)$$

gdzie jak zwykle użyto oznaczenia $P(\Theta) \equiv P(\hat{\Theta})|_{\Theta}$. Gdy całkowanie zostaje wykonane po całej przestrzeni próby, wtedy pomijając człony wyższego rzędu⁴ i biorąc pod uwagę warunek normalizacji $\int dy P(\Theta) = \int dy P(\hat{\Theta}) = 1$ widać, że lewa strona równania (3.53) jest równa zero. Dlatego otrzymujemy wynik, który dla lokalnie nieobciążonych estymatorów [12] oraz po zapostulowaniu zerowania się wyrażenia podcałkowego po prawej stronie, przyjmuje dla konkretnych n oraz n' następującą mikroskopową postać *równania master*:

$$\frac{\partial^2 P(\Theta)}{\partial \theta_{n'} \partial \theta_n} (\hat{\theta}_n - \theta_n) (\hat{\theta}_{n'} - \theta_{n'}) = 0. \quad (3.54)$$

Gdy parametr θ_n^ν ma index Minkowskiego ν , wtedy można pokazać, że $P = \prod_{n=1}^N p_n(\mathbf{y}_n)$ wraz z (3.54) prowadzi, po przejściu do zmiennych Fisherowskich (porównaj (2.127)), do równania mającego postać równania ciągłości strumienia:

$$\frac{\partial p_n(\mathbf{y}_n)}{\partial t_n} + \sum_{i=1}^3 \frac{\partial p_n(\mathbf{y}_n)}{\partial \mathbf{y}_n^i} \hat{v}_n^i = 0, \quad \hat{v}_n^i \equiv \frac{\hat{\theta}_n^i - \theta_n^i}{\hat{\theta}_n^0 - \theta_n^0}, \quad (3.55)$$

gdzie $t_n \equiv \mathbf{y}_n^0$. W (3.55) θ_n^i oraz θ_n^0 są wartością oczekiwaną położenia oraz czasu układu. Ponieważ (3.55) jest słuszne dla każdego n więc index n mógłby zostać pominięty.

3.5 Podsumowanie rozważań

Podstawowym przesłaniem wyniesionym z metody estymacyjnej Friedena-Soffera jest to, że *TFI* jest poprzednikiem Lagrangianu układu [7]. Temat ten rozwinieemy w kolejnym rozdziale. Pewnym minusem teorii Friedena-Soffera mogła wydawać się konieczność zapostulowania nowych zasad informacyjnych. Co prawda z punktu widzenia fenomenologii wydaje się być całkiem satysfakcjonująca skuteczność tych zasad w wyprowadzeniu dużej liczby modeli użytecznych do opisu zjawisk, jednak wyprowadzenie tych zasad przesunęłoby teorię do obszaru bardziej podstawowego. Pozwoliłoby to zarówno na podanie jej przyszłych ograniczeń fenomenologicznych jak i jej możliwych teoretycznych uogólnień. W tym kontekście wyprowadzenie strukturalnej zasady informacyjnej jako konsekwencji analityczności logarytmu funkcji wiarygodności w otoczeniu prawdziwej wartości parametru Θ spełnia tą właśnie rolę. Pozostaje jeszcze do wyprowadzenia wariacyjna zasada informacyjna, która leży u podstaw zasady ekstremizacji działania fizycznego.

⁴Co jest słuszne nawet na poziomie gęstości o ile tylko $P(\hat{\Theta}) \in \mathcal{S}$ nie posiada w Θ wyższych jetów niż 2-iego rzędu.

W obecnym rozdziale zwrócono uwagę, że u podstaw takiego wyprowadzenia może leżeć idea idealnego pomiaru [10], przy której wariacja IF jest równa co do znaku wariacji IS .

W powyższym rozdziale wyprowadzono też równanie master (3.54) dla funkcji wiarygodności, które prowadzi do równania ciągłości strumienia dla punktowego rozkładu w próbie (3.55). Ciekawe jest to, że równanie master pojawia się z rozwinięcia funkcji wiarygodności w szereg Taylora wokół prawdziwej wartości parametru θ . Siłą rzeczy (por.3.53) nie pojawia się więc w nich część nieliniowa struktury układu. Ta gałąź uogólnienia MNW w klasycznej statystycznej estymacji leży bliżej teorii procesów stochastycznych niż EFI.

Wyprowadzając w Rozdziale 3.1.1 strukturalną zasadę informacyjną [10], wykazano, że metoda Friedena-Soffera jest pewną modyfikacją MNW, pozwalającą jak się okaże na nieparametryczną estymację równań ruchu teorii pola lub równań generujących (rozkład) fizyki statystycznej [7]. Wiele z tych równań otrzymano już w [7], zgodnie z informatycznym rozumieniem Friedena-Soffera wspomnianym na końcu Rozdziału 3.2. W [16] wyprowadzenia te zostały sprawdzone dla przyjętej w obecnym skrypcie fizycznej postaci zasad informacyjnych [9]. Jednakże dopiero wyprowadzenie strukturalnej zasady informacyjnej pozwala na faktoryzację z obserwowanej IS części będącej miarą całkową i w związku z tym na prawidłowe umieszczenie rozkładów spełniających równania różniczkowe metody EFI w odpowiednich podprzestrzeniach przestrzeni statystycznej. Dlatego omówieniu bądź przeliczeniu niektórych rozwiązań EFI z uwzględnieniem tego faktu poświęcimy dwa następne rozdziały. Należy jednak podkreślić, że Frieden, Soffer i ich współpracownicy Plastino i Plastino, podali bardzo skuteczną metodę rozwiązania układu zasad informacyjnych dla problemu EFI, gdyż poza warunkami brzegowymi i ewentualnymi równaniami ciągłości nie jest ona ograniczona przez żadną konkretną postać rozkładu. Metodę tą wykorzystamy w dalszym ciągu analizy.

Rozdział 4

Kryteria informacyjne w teorii pola

Główne estymacyjne przesłanie metody EFI. Jak stwierdziliśmy poprzednio, ponieważ podstawowa myśl stojąca za metodą EFI jest następująca: Skoro IF jest infinitezymalnym typem entropii względnej Kulback-Leibler'a, która służy do statystycznego wyboru pomiędzy zaproponowanymi *ręcznie* modelami, zatem po dodatkowym *ręcznym*, aczkolwiek uzasadnionym nałożeniu różniczkowych zasad strukturalnych na układ, staje się ona metodą estymującą równania ruchu i ich wyboru drogą wymogu spełnienia zasad¹ informacyjnych. Bądź, jeśli ktoś woli, metoda EFI jest metodą estymującą rokłady, które są rozwiązaniami tych równań. Jest więc to metoda estymacji nieparametrycznej. Wspomniane równania to np. równania ruchu teorii pola bądź równania generujące rozkłady fizyki statystycznej.

4.1 Informacja Fishera i klasyfikacja modeli

Obecny rozdział poświęcony jest głównie przedstawieniu wstępnej klasyfikacji modeli fizycznych ze względu na skończoność (bądź nieskończoność) pojemności informacyjnej I . Ponadto, poniższe rozważania dla modeli ze skończonym I dotyczą wyłącznie modeli metody EFI. Kolejna, bardziej szczegółowa klasyfikacja pozwala sklasyfikować modele ze względu na wielkość próby N . Jak pokażemy mechanika klasyczna posiada nieskończoną pojemność informacyjną I . Ściśle mówiąc, mechanika klasyczna jest teorią z symplektyczną strukturą (rozmaitości) i nie posiada struktury statystycznej. Czasami jednak słyszy się stwierdzenie, że jest ona stochastyczną granicą mechaniki kwantowej. Ale i na odwrót, według von Neumann'a [27] teoria kwantowa jest niespójna z istnieniem zespołów nie posiadających rozmycia (rozproszenia). W związku z tym, dość powszechnie uważa się, że występowanie odchyłek od klasycznego zachowania się układów można uchwycić jedynie na poziomie statystycznym [28].

Poniżej udowodnimy twierdzenie klasycznej statystyki mówiące o niemożliwości wyprowadzenia mechaniki falowej² metody EFI z mechaniki klasycznej. W tym celu wykorzystamy statystyczne pojęcie pojemności informacyjnej, które jest narzędziem dla dwóch sprzężonych z sobą zagadnień, a mianowicie powyżej wspomnianego statystycznego dowodu o niewyprowadzalności mechaniki

¹W tym punkcie w stosunku do estymowanego równania ruchu działa ona jak MNW w stosunku do estymowanego parametru rozkładu znanego typu.

²Wspólnotę mechaniki falowej z kwantową ograniczymy do typów równań różniczkowych zagadnienia Sturm'a-Liouville'a oraz zasady nieoznaczoności Heisenberga. Dowód przeprowadzony w ramach mechaniki falowej w obszarze wspólnym dla obu teorii falowych, uznajemy co najwyżej jako przesłankę jego słuszności w mechanice kwantowej.

kwantowej z klasycznej, i związanego z nim problemu konsystencji samo-spójnego formalizmu. Ostatni fakt wykorzystywany jest w takich gałęziach badań fizycznych jak nadprzewodnictwo [29], fizyka atomowa i cząstek elementarnych [30] oraz astrofizyka [31].

4.1.1 Podział modeli ze względu na N oraz kategorie I

Jak dotąd nie odnieśliśmy się do wartości N wymiaru próby. Pierwsza klasyfikacja związana z N jest ogólna. Tzn. pokażemy, że modele należą do dwóch różnych ogólnych kategorii z różną wartością N . Pierwsza z nich posiada skończoną wartość N i jest związana ze skończoną wartością I . Obejmuje ona modele mechaniki falowej i klasycznych teorii pola, gdyż jedno skończenie wymiarowe, polowe rozwiązanie równań ruchu określa ewolucję układu wraz z pełnym określeniem jego struktury w przestrzeni i czasie. Natomiast mechanika klasyczna należy do drugiej kategorii z nieskończonym N , gdyż rozwiązanie równania ruchu nie określa struktury cząstki, która musi być niezależnie od tego równania określona poprzez zdefiniowanie w każdym punkcie toru cząstki jej punktowej struktury (np. poprzez dystrybucję δ -Diraca). Mechanika klasyczna okazuje się posiadać nieskończoną wartość I .

Jeśli chodzi o modele ze skończoną wartością I , to poniższe rozważania dotyczą jedynie modeli posiadających rozwinięcie logarytmu funkcji wiarygodności w szereg Taylora i w konsekwencji, zgodnie z założeniami Rozdziału 3.1.1 są wyprowadzalne z pomocą strukturalnej zasady informacyjnej $I + Q = 0$.

4.1.1.1 Dowód podziału na dwie kategorie I

Ubieżmy powyższe słowa następującą analizą. Dla uproszczenia rozważmy układ jednowymiarowy w położeniu³. Załóżmy wpierw, że układ jest opisany przez nieosobliwą dystrybucję. Wtedy dla $N \rightarrow \infty$ pojemność informacyjna I , (2.100), rozbiega się do nieskończoności. Taka sama sytuacja zachodzi jednak dla każdej *osobliwej* dystrybucji jak np. dystrybucja δ -Diraca. Sprawdźmy, że tak jest istotnie. Rozważmy punktową cząstkę swobodną, dla uproszczenia w spoczynku w położeniu θ , oraz δ -Diracowski ciąg funkcji, np. ciąg funkcji Gaussa:

$$\left\{ \delta_k(y_n|\theta) = \frac{k}{\sqrt{\pi}} \exp(-k^2(y_n - \theta)^2) \right\}, \quad \text{gdzie } k = 1, 2, 3, \dots \quad (4.1)$$

Wtedy, ponieważ dla określonego indeksu k ciągu (4.1), pojemność informacyjna (2.100):

$$I_k = \sum_{n=1}^N \int dy_n \delta_k(y_n|\theta) \left(\frac{\partial \ln \delta_k(y_n|\theta)}{\partial \theta} \right)^2, \quad \text{gdzie } k = 1, 2, 3, \dots \quad (4.2)$$

jest równa:

$$I_k = \frac{N}{\sigma_k^2}, \quad \text{gdzie } k = 1, 2, 3, \dots \quad (4.3)$$

gdzie $\sigma_k^2 = \frac{1}{2k^2}$ opisuje wariancję położenia cząstki dla k -tego elementu ciągu (4.1), więc widzimy, że I_k rozbiega się do nieskończoności dla $N \rightarrow \infty$, i nawet jeszcze mocniej gdy dodatkowo $k \rightarrow \infty$. Posumowując, dla $N \rightarrow \infty$ pojemność informacyjna I nie istnieje, obojętnie z jaką dystrybucją mielibyśmy do czynienia.

³Wielowymiarowość czasoprzestrzenna może, w kontekście obecnych rozważań, zmienić co najwyżej znak pojemności informacyjnej.

Tak więc *istnieją dwie, powyżej wymienione, rozdzielne kategorie modeli odnoszące się do wymiaru N próby*. Tzn. dla jednych, takich jak mechanika falowa i teorie pola, N oraz I są skończone, podczas gdy mechanika klasyczna tworzy osobną klasę z nieskończonym N oraz I . To kończy dowód [9] o niewyprowadzalności modeli falowych i teorii polowych z mechaniki klasycznej.

Powyższy dowód nie obejmuje możliwości wyprowadzenia mechaniki falowej (czy też teorii kwantowych) z klasycznej teorii pola bądź samo-spójnej teorii pola [30, 32, 33].

Uwaga: Oznacza to, że mechanika klasyczna nie ma skończonego statystycznego pochodzenia⁴, chyba, że tak jak w (4.1) wprowadzi się nieskończoną liczbę statystycznych parametrów, co jednak pociąga za sobą nieskończoność pojemności informacyjnej I .

4.1.2 Podział modeli ze skończonym I na podklasy z różnym N

Jak już wspomnieliśmy, Frieden i Soffer [7] wyprowadzili modele falowe posługując się pojęciem pojemności informacyjnej I oraz zasadami informacyjnymi estymacyjnej metody EFI. Rozwińmy ten temat w dalszej części obecnego rozdziału. Na razie zauważmy, że stosując (zasadniczo) jednocześnie obie zasady informacyjne, strukturalną $I + \kappa Q = 0$, (3.28), oraz wariacyjną $\delta(I + Q) = 0$, (3.32), oraz uwzględniając odpowiednie fizyczne więzy (wyrażone narzuceniem na układ np. równania ciągłości, symetrii oraz warunków brzegowych), otrzymujemy zróżnicowanie ze względu na N modeli posiadających skończone wartości N oraz I . I tak równanie Kleina-Gordona (oraz równanie Schrödingera jako jego nierelatywistyczna granica, por. Dodatek) posiadają rangę pola $N = 2$, równanie Diraca posiada $N = 8$, równania Maxwell'a posiadają $N = 4$, a teoria grawitacji, zasadniczo bardziej w ujęciu Logunova [34] niż ogólnej teorii względności, posiada $N = 10$ (Dodatek).

4.1.3 Konkluzje i konsekwencje podziału modeli na kategorie I

Powyżej otrzymaliśmy rezultat mówiący, że wszystkie modele opisane strukturalną zasadą informacyjną należą do kategorii skończonej wartości pojemności informacyjnej I oraz, że mechanika klasyczna należy do kategorii nieskończonego I . Zatem w ramach zagadnień rozważanych w skrypcie, granica nie leży pomiędzy tym co micro a makro, ale przebiega pomiędzy teoriami, które mają pochodzenie statystyczne oraz tymi, które mają pochodzenie klasyczno-mechaniczne. Albo lepiej, pomiędzy tym co ma pochodzenie falowe, lub szerzej, teorio-polowe, oraz tym co ma pochodzenie ściśle punktowe.

Ponieważ w konstrukcji modeli klasycznej teorii pola oraz mechaniki falowej, użyty jest ten sam statystyczny formalizm informacji Fishera, dlatego jest ona również właściwym narzędziem w konstrukcji samo-spójnych teorii pola [30], która łączy modele mechaniki falowej i klasycznej teorii pola w jeden, logicznie spójny aparat matematyczny.

Jak wiemy aby otrzymać jakąkolwiek teorię pola, metoda EFI używa dwóch nowych zasad, wariacyjnej, która minimalizuje całkowitą fizyczną informację układu, oraz strukturalną, która (co pokazaliśmy) tą informację zeruje. Frieden i Soffer [7] zwrócili uwagę, że pojęcie informacji poprzedza pojęcie fizycznego działania, i wiele w tej kwestii już zrobiono, a wprowadzony forma-

⁴Frieden co prawda wyprowadził również mechanikę klasyczną z mechaniki falowej, ale jedynie jako graniczny przypadek $\hbar \rightarrow 0$, a wartość N w tym wyprowadzeniu jest nieistotna [7].

lizm można słusznie nazwać podejściem Friedena do równań ruchu. Jednakże liczne zagadnienia, ze względu na różne niż w [7] zrozumienie zasady strukturalnej (patrz [9, 16, 10] oraz obecny skrypt), wymagają ponownego zinterpretowania i zrozumienia. Ciągłe na ogólne opracowanie czeka wprowadzenie do formalizmu informacyjnych poprzedników źródeł oraz lepsze zrozumienie fizyki leżącej u podstaw znaczenia wymiaru próby N . Poniższe rozważania służą usystematyzowaniu istniejącego już statystycznego aparatu pojęciowego informacji kinetycznej i strukturalnej metody EFI oraz lepszemu opisowi związku informacji fizycznej z całą działaniem.

4.2 Równania zasad informacyjnych dla równań modeli fizycznych

Kolejna część obecnego rozdziału poświęcona jest omówieniu rozwiązań zasad informacyjnych metodą EFI dla modeli mechaniki falowej i teorii pola [35]. Punktem wyjścia jest pojemność informacyjna I w jej kinematycznych postaciach (3.40):

$$I = 4 \sum_{n=1}^N \int d^4 \mathbf{x} \sum_{\nu=0}^3 \frac{\partial q_n(\mathbf{x})}{\partial \mathbf{x}_\nu} \frac{\partial q_n(\mathbf{x})}{\partial \mathbf{x}^\nu},$$

bądź (3.49):

$$I = 4N \sum_{n=1}^{N/2} \int d^4 \mathbf{x} \sum_{\nu=0}^3 \frac{\partial \psi_n^*(\mathbf{x})}{\partial \mathbf{x}_\nu} \frac{\partial \psi_n(\mathbf{x})}{\partial \mathbf{x}^\nu},$$

wyprowadzonych w Rozdziale 3.3, gdzie \mathbf{x}'_n są zgodnie z (3.36) przesunięciami wartości pomiarowych położenia zebranych przez układ od ich wartości oczekiwanych. Wyprowadzenie (3.40) oraz (3.49) zostało zaprezentowane z Rozdziale 3.3 i nie zakłada ono (w przeciwieństwie do oryginalnego wyprowadzenia Friedena-Soffera) konieczności istnienia niezmienniczości przesunięcia rozkładów prawdopodobieństwa. Uogólnienia powyższych kinematycznych postaci na przypadek występowania w układzie pól cechowania omówimy w dalszej części rozdziału. Pewne informacje na temat niezmienniczości Lorentzowskiej pojemności informacyjnej I zostały podane w Rozdziale 2.7.1.

W końcu, wyprowadzenie strukturalnej zasady informacyjnej zostało przedstawione w Rozdziale 3.1.1. Odwoływało się ono do pełnych danych pomiarowych $(\mathbf{y}_n)_{n=1}^N$, ale jego postać dla przesunięć $(\mathbf{x}_n)_{n=1}^N$ jest dokładnie taka sama [10]. Tak więc, poniżej stosowane zasady informacyjne, strukturalna oraz wariacyjna, będą odwoływały się do miary probabilistycznej $d\mathbf{x} p_n(\mathbf{x})$ określonej na przestrzeni przesunięć $\mathbf{x} \in \mathcal{X}$ jako przestrzeni bazowej, gdzie \mathcal{X} jest czaso-przestrzenią Minkowskiego R^4 .

Przystąpmy do przedstawienia konstrukcji mechaniki falowej i teorii pola zgodnie z metodą EFI. Według równania (3.27) TPI została określona jako $K = Q + I$. Ponieważ przesunięcie \mathbf{x}_n nie zależy od parametru θ_m dla $m \neq n$, oraz zakres całkowania dla wszystkich \mathbf{x}_n jest taki sam, dlatego I redukuje się do diagonalnych postaci (3.40) bądź (3.49), a Q do postaci:

$$Q = \sum_{n=1}^N \int d^4 \mathbf{x} q_n^2(\mathbf{x}) \mathfrak{F}_n(q_n(\mathbf{x})), \quad (4.4)$$

zgodnie z oznaczeniem w (3.21), bądź w przypadku pola $\psi(\mathbf{x})$, (3.44), do ogólnej (jak zwykle rzeczywistej) postaci:

$$Q \equiv Q_\psi = \int d^4\mathbf{x} \sum_{n,n'=1}^{N/2} \psi_n^*(\mathbf{x})\psi_{n'}(\mathbf{x}) \mathfrak{F}_{nn'}^\psi(\psi(\mathbf{x}), \psi^*(\mathbf{x}), \psi^{(l)}(\mathbf{x}), \psi^{*(l)}(\mathbf{x})), \quad (4.5)$$

przy czym (cała) funkcja podcałkowa jest wielomianem pól $\psi(\mathbf{x})$ oraz $\psi^*(\mathbf{x})$, stopnia nie mniejszego niż 2, oraz ich pochodnych rzędu $l = 1, 2, \dots$ (por.(3.18)), natomiast $\mathfrak{F}_{nn'}^\psi$ jest pewną obserwowaną (w ogólności zespoloną) informacją strukturalną układu. Konkretną, jak się okazuje prostą postać Q dla przypadku pól skalarnych Kleina-Gordona oraz pola Diraca omówimy poniżej.

Korzystając z (3.40), (4.4) oraz (3.27), możemy zapisać TFI w postaci:

$$\mathbb{S} \equiv K = \int d^4\mathbf{x} k, \quad (4.6)$$

gdzie dla pola opisanego amplitudami q_n :

$$k = 4 \sum_{n=1}^N \left[\sum_{\nu=0}^3 \frac{\partial q_n(\mathbf{x})}{\partial \mathbf{x}_\nu} \frac{\partial q_n(\mathbf{x})}{\partial \mathbf{x}^\nu} + \frac{1}{4} q_n^2(\mathbf{x}) \mathfrak{F}_n(q_n(\mathbf{x})) \right]. \quad (4.7)$$

natomiast dla pola opisanego amplitudami ψ_n :

$$k = 4N \sum_{n,n'=1}^{N/2} \left[\sum_{\nu=0}^3 \delta_{nn'} \frac{\partial \psi_n^*(\mathbf{x})}{\partial \mathbf{x}_\nu} \frac{\partial \psi_n(\mathbf{x})}{\partial \mathbf{x}^\nu} + \frac{1}{4} \psi_n^*(\mathbf{x})\psi_{n'}(\mathbf{x}) \mathfrak{F}_{nn'}^\psi(\psi(\mathbf{x}), \psi^*(\mathbf{x}), \psi^{(l)}(\mathbf{x}), \psi^{*(l)}(\mathbf{x})) \right]. \quad (4.8)$$

Równoważność w (4.6) sugeruje, że K pełni funkcję (statystycznego) poprzednika (całki) działania \mathbb{S} , natomiast k będące funkcją gęstości TFI, jest poprzednikiem gęstości Lagrangianu \mathcal{L} . Sprawie tej poświęcimy jeden z poniższych rozdziałów.

Uwaga: W dalszej części skryptu założymy, że obserwowana informacja strukturalna nie zawiera pochodnych pól rzędu wyższego niż $l = 1$. Założenie to ma charakter fizyczny. Oznacza ono, że jeśli współrzędne uogólnione (u nas amplitudy) oraz prędkości uogólnione (u nas pochodne amplitud) układu są zadane w pewnej chwili czasu, to ewolucja układu jest całkowicie określona, o ile równania ruchu są 2-giego rzędu. Odpowiada to sformułowaniu Lagrange'a wykorzystywanemu w badaniu dynamicznych i termodynamicznych własności układów.

Fakt ten z punktu widzenia statystycznego oznacza, że interesują nas tylko takie (pod)przestrzenie statystyczne, dla których wszystkie możliwe logarytmy funkcji wiarygodności posiadają r -jety $J_p^r(\mathcal{S}, \mathbb{R})$ w otoczeniu U_p punktu $p \equiv P(\Theta) \in \mathcal{S}$ rzędu $r \leq 2$ (por. Rozdział 2.3). Fakt ten jest istotny z punktu widzenia obserwowanej informacji Fishera (3.7) zdefiniowanej (pierwotnie) poprzez drugie pochodne logarytmu funkcji wiarygodności po parametrach, co z kolei umożliwia konstrukcję strukturalnej zasady informacyjnej (Rozdział 3.1.1), która jest równaniem metody EFI. Jednocześnie (oczekiwana) informacja Fishera (3.15) wchodzi w nierówność Rao-Cramera, której pewną postacią jest, po dokonaniu w informacji Fishera transformacji Fouriera do przestrzeni pędowej, zasada nieoznaczoności Heisenberga (Dodatek). Zatem fakt występowania w

funkcji Lagrange'a (kwadratu) pierwszych pochodnych byłby (z tego punktu widzenia) artefaktem konieczności wykonania przez układ estymacji jego czaso-przestrzennych położeń, która to estymacja posiada dolne ogranicze Rao-Cramera na dokładność jednoczesnej estymacji położenia oraz prędkości.

Odpowiedzmy jeszcze na pytanie co z punktu widzenia statystycznego oznacza niewystępowanie w rozwinięciu $\ln P(\hat{\Theta})$ w szereg Taylora (3.2) wyrazów wyższego rzędu niż drugi (tzn. brak jetów o rzędzie $r > 2$). Sytuacja ta ma miejsce gdy obserwowana informacja Fishera \mathbf{iF} , (3.7), nie zależy od parametru $\Theta \in V_{\Theta}$, gdzie V_{Θ} jest przestrzenią parametru Θ . Wtedy bowiem jej pochodne po parametrze $\Theta = (\theta_n)$ są w równe zero dla każdego punktu $p' = P(\Theta')$ w otoczeniu U_p . Zatem:

$$\frac{\partial}{\partial \Theta'} \mathbf{iF} \Big|_{p' \in U_p} = \frac{\partial}{\partial \Theta'} \left(-\frac{\partial^2 \ln P(\Theta')}{\partial \Theta'^2} \right) \Big|_{p' \in U_p} = 0 \Rightarrow j_{p'}^r(\mathcal{S}, \mathcal{R}) - j_{p'}^2(\mathcal{S}, \mathcal{R}) = 0 \text{ dla } r > 2, \quad (4.9)$$

gdzie $j_p^r(\mathcal{S}, \mathcal{R})$ jest dowolnym elementem przestrzeni jetów $J_p^r(\mathcal{S}, \mathcal{R})$. Lewa strona powyższej implikacji oznacza, że:

$$(\mathbf{iF})_{nn'} \Big|_{p' \in U_p} = \text{const. na } U_p, \quad (4.10)$$

tzn. obserwowana informacja Fishera \mathbf{iF} nie zależy w U_p od parametru Θ , z czego wynika niezależność (obserwowanej) IF , czyli metryki Rao-Fishera od parametru Θ w U_p .

Takie zachowanie się metryki Rao-Fishera ma następującą ciekawą konsekwencję. Otóż w Rozdziale 5.2.3 okaże się, że fakt stałości metryki Rao-Fishera (5.206) jest odpowiedzialny za otrzymanie w ramach metody EFI znanych formuł mechaniki kwantowej (5.219) opisujących splątanie w problemie EPR-Bohm'a.

Nałożmy strukturalną $I + \kappa Q = 0$, (3.28), i wariacyjną $\delta K = \delta(I + Q) = 0$, (3.32), zasadę informacyjną na K [7, 9]. Z zasady strukturalnej na poziomie obserwowanym (mikroskopowym), tzn. pod całką, wynika warunek zerowy:

$$\sum_{\nu=0}^3 \frac{\partial q_n(\mathbf{x})}{\partial \mathbf{x}^{\nu}} \frac{\partial q_n(\mathbf{x})}{\partial \mathbf{x}^{\nu}} + \frac{\kappa}{4} q_n^2(\mathbf{x}) \mathbf{F}_n(q_n(\mathbf{x})) = 0, \quad n = 1, 2, \dots, N. \quad (4.11)$$

Natomiast z zasady wariacyjnej wynika układ równań Eulera-Lagrange'a:

$$\sum_{\nu=0}^3 \frac{\partial}{\partial \mathbf{x}^{\nu}} \left(\frac{\partial k}{\partial \left(\frac{\partial q_n(\mathbf{x})}{\partial \mathbf{x}^{\nu}} \right)} \right) = \frac{\partial k}{\partial q_n(\mathbf{x})}, \quad n = 1, 2, \dots, N. \quad (4.12)$$

Odpowiednia dla pola ψ obserwowana postać zasady strukturalnej jest następująca:

$$\begin{aligned} & \sum_{\nu=0}^3 \frac{\partial \psi_n^*(\mathbf{x})}{\partial \mathbf{x}^{\nu}} \frac{\partial \psi_n(\mathbf{x})}{\partial \mathbf{x}^{\nu}} \\ & + \frac{\kappa}{4} \sum_{n'=1}^{N/2} \psi_n^*(\mathbf{x}) \psi_{n'}(\mathbf{x}) \mathbf{F}_{nn'}^{\psi}(\psi(\mathbf{x}), \psi(\mathbf{x}), \frac{\partial \psi(\mathbf{x})}{\partial \mathbf{x}}, \frac{\partial \psi^*(\mathbf{x})}{\partial \mathbf{x}}) = 0, \quad n = 1, 2, \dots, N/2. \end{aligned} \quad (4.13)$$

Natomiast układ równań Eulera-Lagrang'a wynikający z zasady wariacyjnej, jest następujący:

$$\sum_{\nu=0}^3 \frac{\partial}{\partial \mathbf{x}^{\nu}} \left(\frac{\partial k}{\partial \left(\frac{\partial \psi_n^*(\mathbf{x})}{\partial \mathbf{x}^{\nu}} \right)} \right) = \frac{\partial k}{\partial \psi_n^*(\mathbf{x})}, \quad n = 1, 2, \dots, N. \quad (4.14)$$

Zauważmy jednak, że powyższa postać k , (4.8), jest na tyle ogólna, że aby zobaczyć działanie metody EFI wynikające z równań (4.13) i (4.14), należy podać konkretną postać k , dla każdego zagadnienia z polem typu ψ z osobna.

Natomiast jak się przekonamy, równania (4.11) oraz (4.12) są już zapisane w postaci bliskiej ich bezpośredniego użycia i otrzymania jawnej postaci \mathbf{q}_n oraz rozwiązań metody EFI, czyli odpowiednich fizycznych równań ruchu (bądź równań generujących, por. Rozdział 5) dla amplitud $q_n(\mathbf{x})$. Zobaczymy, że wszystkie modelowe różnice leżą po stronie \mathbf{q}_n , której postać zależy od konkretnego fizycznego scenariusza, włączając w to symetrie oraz warunki brzegowe. Po raz pierwszy (4.11) oraz (4.12) otrzymali Frieden i Soffer [7]. Jednak powyższa ich forma uwzględnia inną interpretację IS (jako obecnej stałe podczas ewolucji układu) oraz faktoryzację probabilistycznego czynnika $q_n^2(\mathbf{x})$ zawartego obligatoryjnie w mierze całkowitej⁵.

4.3 TFI oraz specyficzne formy Q w teorii pola

Poniżej są zebrane i rozwinięte wyniki EFI otrzymane poprzednio w [7]. Jednakże zapisano je, szczególnie dla Q , w otwartej formie z punktu widzenia analizy porównawczej modeli [35]. Metoda EFI prowadzi do takiego sformułowania metody teorii pola, która jest zgodna z dzisiejszym opisem mechaniki falowej dla szerokiej klasy (kwantowych) struktur.

Jednakże, czasami metoda EFI wraz z całym towarzyszącym jej statystycznym aparatem pojęciowym może doprowadzić do korekty istniejącej analizy. Z sytuacją taką możemy mieć do czynienia np. w przypadku sformułowania zasady nieoznaczoności Heisenberga dla pola *światłowego*. Otóż w świetle nowych eksperymentów, w wyniku których otrzymano *za wąski impuls świetlny w częstotliwości* [36], standardowa Fourierowska podstawa zasady nieoznaczoności jest ostatnio kwestionowana. Wytlumaczenie istoty nierówności Heisenberga w oparciu o nierówność Rao-Cramera wraz z rozróżnieniem pomiędzy estymacją parametru w przypadku skalarnym [7] (por. Dodatek) i wektorowym, dla którego zachodzi ciąg nierówności (2.55), może okazać się kluczem do zrozumienia pytań, narastających na skutek nowych eksperymentów ze światłem.

W metodzie EFI, informacja strukturalna Q musi zostać wyprowadzona z użyciem zasady strukturalnej, wariacyjnej i czasami pewnych dodatkowych warunków symetrii, biorących pod uwagę specyficzny fizyczny scenariusz teorii. Szeroki opis metod stosowanych przy rozwiązywaniu równań (4.12) oraz (4.11) można znaleźć w [7].

Jednak poniższe rozważania powinny okazać się pomocne w zrozumieniu metody, szczególnie dla układu zasad informacyjnych (4.13) oraz (4.14) pola ψ . Poniżej zostanie pokazane jak możliwe rozwiązania zasad informacyjnych, strukturalnej i wariacyjnej, przewiduje pojawienia się trzech typów pól: N -skalarów [35], fermionów oraz bozonów [7].

Rozważmy cząstkę jako układ opisany polem rangi N , poprzez zbiór zespolonych funkcji falowych $\psi_n(\mathbf{x})$, $n = 1, 2, \dots, N/2$, określonych w czaso-przestrzeni położenia $\mathbf{x} \equiv (\mathbf{x}^\mu)_{\mu=0}^3$, zgodnie z konstrukcją (3.42) oraz (3.43) z Rozdziału 3.3. Ich transformaty Fouriera $\phi_n(\mathbf{p})$ w sprzężonej (do czaso-przestrzeni) przestrzeni energetyczno-pędowej $\mathbf{p} \equiv (\wp^\mu)_{\mu=0}^3 = (\frac{E}{c}, \wp^1, \wp^2, \wp^3)$, mają

⁵Również wyprowadzenie zasady strukturalnej (4.11) czyni ją mniej fundamentalną niż w sformułowaniu Friedena-Soffera [7], a tym co staje się fundamentalne, jest funkcja wiarygodności $P(\Theta)$ próbki, czyli łączna gęstość prawdopodobieństwa jej (niewidocznej dla badacza) realizacji.

postać:

$$\psi_n(\mathbf{x}) = \frac{1}{2\pi\hbar} \int d^4\mathbf{p} \phi_n(\mathbf{p}) e^{-i(\sum_{\nu=0}^3 \mathbf{x}^\nu \wp_\nu)/\hbar}, \quad (4.15)$$

gdzie \hbar jest stałą Plancka.

Uwaga: Transformacja Fouriera jest *unitarną transformacją zachowującą miarę* na przestrzeni L^2 funkcji całkowalnych z kwadratem, tzn.:

$$\int d^4\mathbf{x} \psi_n^*(\mathbf{x}) \psi_m(\mathbf{x}) = \int d^4\mathbf{p} \phi_n^*(\mathbf{p}) \phi_m(\mathbf{p}), \quad (4.16)$$

zatem wykorzystując warunek normalizacji prawdopodobieństwa (3.51), otrzymujemy⁶:

$$\sum_{n=1}^{N/2} \int d^4\mathbf{x} |\psi_n(\mathbf{x})|^2 = \sum_{n=1}^{N/2} \int d^4\mathbf{p} |\phi_n(\mathbf{p})|^2 = 1, \quad (4.18)$$

gdzie $|\psi_n(\mathbf{x})|^2 \equiv \psi_n^*(\mathbf{x}) \psi_n(\mathbf{x})$ oraz $|\phi_n(\mathbf{p})|^2 \equiv \phi_n^*(\mathbf{p}) \phi_n(\mathbf{p})$. Korzystając z (4.15) możemy zapisać I podane wzorem (3.49) w następujący sposób:

$$I[\psi(\mathbf{x})] = I[\phi(\mathbf{p})] = \frac{4N}{\hbar^2} \int d^4\mathbf{p} \sum_{n=1}^{N/2} |\phi_n(\mathbf{p})|^2 \left(\frac{E^2}{c^2} - \vec{\wp}^2 \right), \quad (4.19)$$

gdzie $\vec{\wp}^2 = \sum_{k=1}^3 \wp_k \wp^k$.

Ponieważ I jest z definicji wartością oczekiwaną, dlatego kwadrat *masy* cząstki zdefiniowany jako [7]:

$$m^2 := \frac{1}{c^2} \int d^4\mathbf{p} \sum_{n=1}^{N/2} |\phi_n(\mathbf{p})|^2 \left(\frac{E^2}{c^2} - \vec{\wp}^2 \right) \quad (4.20)$$

jest stałą niezależnie od statystycznych fluktuacji energii E oraz pędu $\vec{\wp}$, tzn. (przynajmniej) gdy całkowanie jest wykonane, czyli jako średnia. Tak więc możemy (4.19) dla cząstki swobodnej zapisać następująco:

$$I[\psi(\mathbf{x})] = I[\phi(\mathbf{p})] = 4N \left(\frac{m c}{\hbar} \right)^2 = const. \quad (4.21)$$

Powyższy warunek oznacza, że:

$$K_F \equiv I[\psi(\mathbf{x}^\mu)] - I[\phi(p^\mu)] = 0, \quad (4.22)$$

co korzystając ze stałości $4N \left(\frac{m c}{\hbar} \right)^2$ oraz (4.18) można zapisać jako warunek spełniony przez pole swobodne rangi N :

$$K_F = 4N \int d^4\mathbf{x} \sum_{n=1}^{N/2} \left[\sum_{\nu=0}^3 \left(\frac{\partial \psi_n}{\partial \mathbf{x}_\nu} \right)^* \frac{\partial \psi_n}{\partial \mathbf{x}^\nu} - \left(\frac{m c}{\hbar} \right)^2 \psi_n^* \psi_n \right] = 0. \quad (4.23)$$

⁶Równość:

$$\int d^4\mathbf{x} \psi_n^*(\mathbf{x}) \psi_n(\mathbf{x}) = \int d^4\mathbf{p} \phi_n^*(\mathbf{p}) \phi_n(\mathbf{p}), \quad (4.17)$$

jest treścią twierdzenia Parseval'a.

Wielkość K_F definiuje tzw. *informację Fouriera* (F), a k_F jej gęstość:

$$k_F = 4N \sum_{n=1}^{N/2} \left[\sum_{\nu=0}^3 \left(\frac{\partial \psi_n}{\partial \mathbf{x}_\nu} \right)^* \frac{\partial \psi_n}{\partial \mathbf{x}^\nu} - \left(\frac{m c}{\hbar} \right)^2 \psi_n^* \psi_n \right]. \quad (4.24)$$

“Nie biorąc pod uwagę” powyższych rachunków, z których m^2 wyłania się jako średnia, równanie (4.22) (oraz (4.23)) jest odbiciem twierdzenia Parseval’s (4.17) i jako takie jest ono zdaniem tautologicznym. Fakt ten oznacza, że (unitarna) transformacja Fouriera odzwierciedla jedynie zmianę bazy w przestrzeni statystycznej \mathcal{S} . Dlatego też, sam z siebie, warunek (4.23) nie nakłada żadnego dodatkowego więzu na układ, *chyba, że $4N(\frac{m c}{\hbar})^2$ jest zadana jako informacja strukturalna Q układu*. Tylko wtedy (4.23) staje się informacyjną zasadą strukturalną dla układu definiującego szczególny typ pola omawianego w Rozdziale 4.3.1.

Powtórka powyższych rachunków wykonana dla pola rangi N , określonego poprzez rzeczywiste amplitudy $q_n(\mathbf{x})$ i posiadającego pojemność informacyjną (3.40), prowadzi przy transformacji Fouriera:

$$q_n(\mathbf{x}) = \frac{1}{2\pi\hbar} \int d^4\mathbf{p} \tilde{q}_n(\mathbf{p}) e^{-i(\sum_{\nu=0}^3 \mathbf{x}^\nu p_\nu)/\hbar} \quad (4.25)$$

i warunku unormowania (por. (3.50)):

$$\frac{1}{N} \sum_{n=1}^N \int d^4\mathbf{x} q_n^2(\mathbf{x}) = 1, \quad (4.26)$$

do następującej postaci informacji Fouriera:

$$K_F = 4 \int d^4\mathbf{x} \sum_{n=1}^N \left[\sum_{\nu=0}^3 \frac{\partial q_n}{\partial \mathbf{x}_\nu} \frac{\partial q_n}{\partial \mathbf{x}^\nu} - N \left(\frac{m c}{\hbar} \right)^2 q_n^2 \right] = 0. \quad (4.27)$$

4.3.1 Skalary Kleina-Gordona

Równanie ruchu Kleina-Gordona dla swobodnego pola rangi N wynika z wariacyjnej zasady informacyjnej $\delta_\psi K = 0$, (4.12), po narzuceniu na Q następującego warunku:

$$Q[\phi(\mathbf{p})] = Q_S \equiv -4N \left(\frac{m c}{\hbar} \right)^2, \quad (4.28)$$

co oznacza, że TFI swobodnego pola Kleina-Gordona jest równe jego informacji Fouriera (4.23), tzn. $K = K_F$. Z wariacyjnej zasady informacyjnej $\delta_{(\psi^*)} K = 0$ wynika N równań Eulera-Lagrange’a (4.14), które dla gęstości TFI równej $k = k_F$, (4.24), prowadzą do N równań Kleina-Gordona⁷. Ich wyprowadzenie można znaleźć w [7, 16].

Ponieważ warunek (4.23) stanowi w tym przypadku strukturalną zasadę informacyjną dla układu $I + Q = 0$, zatem dla swobodnego pola skalarnego transformacja Fouriera jest transformacją unitarną, ze względu na którą warunek ten pozostaje niezmienniczy. Fakt ten zauważony w [7] oznacza, że transformacja Fouriera tworzy rodzaj samo-splątania pomiędzy realizowanymi wartościami (bo takie “siedzą” w I) zmiennych układu [10]. Zauważmy, że wyprowadzenie zasady $I + Q = 0$ jako konsekwencji rozwinięcia w szereg Taylora funkcji wiarygodności należącej do

⁷Patrz równanie (4.49) w przypisie.

przestrzeni statystycznej \mathcal{S} , [10], wyjaśnia charakter tego splątania pomiędzy różnymi stopniami swobody funkcji wiarygodności jako splątania, spowodowanego masą układu pomiędzy różnymi modami pędowymi w (4.20).

Podajmy wynikające z powyższej analizy metody EFI dwa typy pól skalarnych:

Zwykłe, (naładowane) pole skalarne: Pole skalarne mające rangę $N = 2$ ma tylko jedną składową zespoloną, tzn. $\psi \equiv (\psi_n) = \psi_1$ [7]. W przypadku tym informacja strukturalna (4.28) jest równa $Q_S = -8(\frac{mc}{\hbar})^2$. Naładowane pole skalarne Higgsa H^+ [37] mogłoby teoretycznie być jego przykładem.

N-skalary: W przypadku $n > 1$ składowe ψ_n podlegają ewolucji opisanej przez n nie sprzężonych równań Kleina-Gordona z dwoma dodatkowymi więzami. Pierwszy z nich oznacza, że wszystkie pola ψ_n mają taką samą masę m , a drugim jest warunek normalizacji (4.18). Informacja strukturalna Q takiego układu jest określona przez ogólną postać (4.28) dla pola skalarnego rangi N . Owe skalarne pola Kleina-Gordona rangi N nazwijmy *N-skalarami* [35]. Są one teoretycznie realizowane w ramach tzw. σ -modeli teorii pola [38, 35].

W kolejnym rozdziale omówimy postać TFI oraz Q dla równania Diraca. Podstawowe fakty metody EFI dla pól cechowania w elektrodynamice Maxwella [7] omówione są w Dodatku. Również w Dodatku zamieszczona jest postać Q w teorii grawitacji [7]. W opracowaniu jest postać Q dla pól nieabelowych [35].

4.3.2 TFI równania Kleina-Gordona dla pól rangi N

Rozdział ten poświęcony jest konstrukcji równania Diraca metodą EFI, z uwzględnieniem pól cechowania⁸. Szczególną uwagę zwrócono na problem kwadratury TFI pola Kleina-Gordona [7, 35].

4.3.2.1 Wstępna foliacja \mathcal{S} oraz pochodna kowariantna

Ogólny zarys problemu: Model EFI, z którego wyłoni się model teorii pola jest budowany na przestrzeni bazowej B , będącej w rozważanych przez nas przypadkach czaso-przestrzenią Minkowskiego $\mathcal{X} \equiv R^4$. W celu uczynienia (kinematycznej postaci) pojemności informacyjnej I niezmienniczą ze względu na lokalne transformacje cechowania, musimy, poprzez zdefiniowanie pochodnej kowariantnej na przestrzeni bazowej \mathcal{X} , zapisać I w postaci współzmienniczej. Z kolei, zdefiniowanie tej pochodnej kowariantnej oznacza konieczność podania układu współrzędnych na przestrzeni statystycznej \mathcal{S} , co wynika z tego, że pola cechowania są (jak się okaże) amplitudami typu q_n .

Zatem wprowadzenie układu współrzędnych na \mathcal{S} nie jest zadaniem trywialnym [12]. Sytuacja ta wynika z konieczności wykonania analizy EFI dla pól rangi N , które z góry przekształcają się zgodnie z transformacjami, których parametry mogą zależeć lokalnie od położenia w przestrzeni bazowej B . Zatem na B określana jest *strukturalna grupa symetrii* G wspomnianej powyżej

⁸Zgodnie z uwagą uczynioną powyżej, skrót rezultatów metody EFI dla samych pól cechowania w elektrodynamice Maxwella znajduje się w Dodatku.

transformacji, która w teorii pola jest grupą Liego pola cechowania. Konstruuje się więc główną wiązkę włóknistą $E(B, G)$, a fizyczną estymację przeprowadza się na włóknach⁹.

Zauważmy, że skoro kwadrat pola cechowania jest elementem \mathcal{S} , więc dla określonej algebry grupy cechowania, wybór cechowania dokonuje częściowej foliacji¹⁰ przestrzeni \mathcal{S} na pierwsze warstwy. Dopiero w tym momencie, zasady informacyjne umożliwiają dokonanie wyboru kolejnych foliacji przestrzeni \mathcal{S} na warstwy związane z wszystkimi szczególnymi reprezentacjami grupy G [35], których wymiar jest ściśle związany z rangą pól N [7].

Sens powyższego rozważania ujmijmy tak: Cała procedura EFI musi nie tylko od początku wybrać typ amplitudy (tzn. q_n lub ψ_n), określić zasady informacyjne i warunki brzegowe, ale musi być wykonana od początku we właściwym układzie współrzędnych, który uwzględnia istnienie strukturalnej grupy symetrii G oraz związanych z nią pól cechowania. Pole cechowania umożliwia bowiem wybór podprzestrzeni $T_{\mathcal{P}}H$ przestrzeni stycznej $T_{\mathcal{P}}E$ wiązki głównej $E \equiv E(B, G)$ w każdym jej punkcie \mathcal{P} , tak, że $T_{\mathcal{P}}E = T_{\mathcal{P}}G \otimes T_{\mathcal{P}}H$, przy czym baza na $T_{\mathcal{P}}H$ może być zdefiniowana jako liniowa kombinacja cztero-wymiarowej bazy:

$$(\partial_{\mu}) \equiv \left(\frac{\partial}{\partial \mathbf{x}^{\mu}} \right) \equiv \left(\frac{\partial}{\partial (ct)}, \frac{\partial}{\partial \mathbf{x}^1}, \frac{\partial}{\partial \mathbf{x}^2}, \frac{\partial}{\partial \mathbf{x}^3} \right) \equiv \left(\frac{\partial}{\partial (ct)}, \vec{\nabla} \right) \quad (4.29)$$

oraz generatorów D_G infinitezymalnych transformacji grupy G .

Na przykład, w przypadku pól cechowania A_{μ} grupy $G = U(1)$ pochodna kowariantna ma postać:

$$D_{\mu} \equiv (D_0, D_l) = \partial_{\mu} - i \frac{e}{c\hbar} A_{\mu}, \quad (4.30)$$

gdzie e jest ładunkiem elektronu. Zatem, wprowadzając pochodną kowariantną do zasad informacyjnych i stosując metodę EFI, otrzymujemy równania ruchu, które rozwiązując dają bazę na znalezionych (pod)rozmaitościach przestrzeni statystycznej \mathcal{S} .

Jest wiele sposobów, na które współzmiennicza postać pojemności informacyjnej I może być odczytana. Poniżej skoncentrujemy się na dwóch z nich, jednej dla pól skalarnych rangi N (tzn. N -skalarów) oraz drugiej, dla pól fermionowych rangi N .

4.3.2.2 Równanie ruchu Diraca dla pola swobodnego rangi N .

Zgodnie z powyższymi rozważaniami, współzmiennicza forma (3.49) pojemności informacyjnej z pochodną kowariantną D_{μ} ma postać:

$$I = 4N \int d^4\mathbf{x} \sum_{n=1}^{N/2} \sum_{\mu=0}^3 (D_{\mu}\psi_n)^* D^{\mu}\psi_n. \quad (4.31)$$

⁹**Uwaga o głównej wiązce włóknistej:** Mając rozmaitość B oraz grupę Liego G , główna wiązka włóknista $E(B, G)$ jest rozmaitością taką, że:

1. Grupa G działa na E w sposób różniczkowalny i bez punktów stałych.
 2. Przestrzeń bazowa $B = E/G$, tzn. B jest przestrzenią ilorazową E względem G , oraz istnieje różniczkowalne odwzorowanie (nazywane rzutowaniem) $\pi : E \rightarrow B$.
 3. Dla każdej mapy $\{U_i\}$ w atlasie dla B , istnieje różniczkowalne i odwracalne odwzorowanie $\phi_j : \pi^{-1}(U_j) \rightarrow U_j \times G$ zadane przez $E \rightarrow (\pi(\mathcal{P}), f(\mathcal{P}))$ w każdym punkcie $\mathcal{P} \in E$, gdzie $f : \pi^{-1}(U_j) \rightarrow G$ spełnia warunek $f(g\mathcal{P}) = gf(\mathcal{P})$, dla każdego $g \in G$.
- Obraz π^{-1} , czyli $U_j \times G$, jest nazywany włóknem. Zatem każde włókno niesie z sobą kopię grupy strukturalnej G .

¹⁰Jako, że baza dla pozostałych pól nie jest jeszcze wybrana.

Zatem jedyna TFI (3.27) dostępna w metodzie EFI dla równania Kleina-Gordona i każdego pola typu ψ rangi N (skalarne czy fermionowe) ma postać:

$$K = K_{KG} \equiv 4N \int d^4\mathbf{x} \sum_{n=1}^{N/2} \sum_{\mu=0}^3 (D_\mu \psi_n)^* D^\mu \psi_n + Q. \quad (4.32)$$

Podobnie, wykorzystując (3.40) w miejsce (3.49) moglibyśmy zapisać K_{KG} dla pola bosonowego, do czego powrócimy jednak później.

Są dwie drogi, którymi analiza oparta o TFI zadaną przez (4.32) może podążać [7, 35]. Pierwsza związana jest z N -skalarami a druga z polami Diraca. Wtedy, gdy rozważamy pole N -skalara, to jak wiemy Q jest równe Q_S zadanemu przez (4.28). Fakt ten oznacza, że TFI dla równania ruchu N -skalara sprowadza się do postaci Kleina-Gordona, tzn.:

$$K = K_S = K_{KG} \quad \text{dla} \quad Q = Q_S, \quad (4.33)$$

a wariacyjna zasada informacyjna:

$$\delta_{(\psi^*)} K_S \equiv \delta_{(\psi^*)} (I + Q_S) = 0, \quad (4.34)$$

zadana przez (3.32) (lub (4.14)) prowadzi do równania Kleina-Gordona dla pola skalarne ([7, 16]).

Jak wiemy, dla pola Diraca, równanie Kleina-Gordona jest otrzymane drogą kwadratury równania Diraca [39]. Zatem, mogłoby się wydawać, że metoda informacyjna nie wybiera sama z siebie właściwej postaci TFI dla pola fermionowego. Sprawy mają się jednak następująco [35]. Zapiszmy (4.31) pól ψ , zarówno dla N -skalara jak i pola fermionowego rangi N , w następującej postaci:

$$I = 4N \sum_{n=1}^{N/2} \int d^4\mathbf{x} \sum_{\mu, \nu=0}^3 (D_\mu \psi_n^*(\mathbf{x})) \eta^{\mu\nu} (D_\nu \psi_n(\mathbf{x})), \quad (4.35)$$

która zapisana w tej postaci pozwala zauważyć, że jest ona niezmiennicza ze względu na transformacje izometrii działające w przestrzeni $\mathcal{C}^{N/2}$ pól $\psi \equiv (\psi_n(\mathbf{x}))$ rangi N . W przypadku N -skalarów ψ tworzących podprzestrzeń w $\mathcal{C}^{N/2}$ izometrie te są transformacjami identycznościowymi, pozostawiając pola ψ niezmiennymi. Natomiast dla pól fermionowych rangi N określonych na przestrzeni bazowej Minkowskiego, transformacje izometrii tworzą grupę Clifforda $Pin(1, 3)$ (będącą podzbiorem algebry Clifforda $C(1, 3)$), której elementy działają w podprzestrzeni $\mathcal{C}^{N/2}$ spinorów ψ . Okazuje się, że macierze Diraca γ^μ , $\mu = 0, 1, 2, 3$, (por. (4.41)-(4.43)), tworzą spinorową reprezentację ortogonalnej bazy w $C(1, 3)$, i spełniają tożsamości:

$$\eta^{\mu\nu} = 1/2 \{ \gamma^\mu, \gamma^\nu \}, \quad \mu, \nu = 0, 1, 2, 3, \quad (4.36)$$

które są podstawowym związkiem dla iloczynu Clifforda. W przypadku spinorów rangi $N = 8$ baza w $Pin(1, 3)$ jest $2^{N/2} = 16$ wymiarowa [39].

Dla trywialnego przypadku N -skalarów, postać (4.32) jest formą podstawową. Jednak w przypadku spinorów rangi N , po skorzystaniu z (4.36) w (4.35), można dokonać rozkładu fizycznej informacji K_{KG} , (4.32), na składowe. Postać jawnego rozkładu K_{KG} na składowe i ich faktoryzację, gdzie każdy z otrzymanych czynników jest elementem grupy Clifforda $Pin(1, 3)$, podał

Frieden [7]. Główny rezultat rozkładu fizycznej informacji K_{KG} , z uwzględnieniem pól cechowania w pochodnej kowariantnej D_μ , który czyni zadość “ograniczeniu” Klein-Gordon (co oznacza, że wymnożenie czynników i dodanie ich dałoby na powrót (4.32)), ma postać:

$$K = K_D \equiv 4N \int d^4\mathbf{x} \sum_{n=1}^{N/2} \sum_{\mu=0}^3 (D_\mu \psi_n)^* D^\mu \psi_n + Q, \quad (4.37)$$

gdzie informacja strukturalna Q jest równa:

$$Q = Q_D \equiv -4N \int d^4\mathbf{x} \sum_{n=1}^{N/2} \left[v_{1n} v_{2n} + \left(\frac{mc}{\hbar}\right)^2 \psi_n^* \psi_n \right] + (\text{pozostałe człony}), \quad (4.38)$$

a $\frac{N}{2} = 4$ wymiarowe wektory kolumnowe $v_i = (v_{i1}, v_{i2}, v_{i3}, v_{i4})^T$, $i = 1, 2$, mają postać:

$$v_{1n} = \sum_{n'=1}^4 \left(i D_0 - \beta \frac{mc}{\hbar} + \sum_{i=1}^3 \alpha^i i D_i \right)_{nn'} \psi_{n'}, \quad n = 1, 2, 3, 4, \quad (4.39)$$

oraz

$$v_{2n} = \sum_{n'=1}^4 \left(-i D_0 + \beta^* \frac{mc}{\hbar} + \sum_{l=1}^3 \alpha^{l*} i D_l \right)_{nn'} \psi_{n'}^*, \quad n = 1, 2, 3, 4, \quad (4.40)$$

gdzie macierze α^l , $l = 1, 2, 3$, oraz β są macierzami Diraca (4.42)¹¹.

W końcu, informacyjna zasada strukturalna (3.31) $\kappa = 1$ oraz zasada wariacyjna (3.32), mają postać:

$$I + Q_D = 0 \quad \text{dla } \kappa = 1, \quad \text{oraz} \quad \delta_{(\psi^*)} K_D = \delta_{(\psi^*)} (I + Q_D) = 0. \quad (4.44)$$

Powyższe zasady dają na poziomie obserwowanym warunki (4.13) oraz (4.14) metody EFI, czyli układ dwóch równań różniczkowych, które są rozwiązywane samo-spójnie. W wyniku otrzymujemy równanie Diraca:

$$v_1 = \left(i D_0 - \beta \frac{mc}{\hbar} + \sum_{l=1}^3 \alpha^l i D_l \right) \psi = 0, \quad \text{gdzie } \psi = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix}. \quad (4.45)$$

¹¹Macierze Diraca γ^μ , $\mu = 0, 1, 2, 3$, występujące w tzw. kowariantnej formie równania Diraca:

$$(i \gamma^\mu D_\mu - m) \psi = 0, \quad (4.41)$$

wyrażają się poprzez macierze Diraca β oraz α^l , $l = 1, 2, 3$, w sposób następujący: $\gamma^0 = \beta$ oraz $\gamma^l = \beta \alpha^l$, ($\gamma_\mu = \sum_{\nu=0}^3 \eta_{\mu\nu} \gamma^\nu$). Macierze Diraca $\vec{\alpha} \equiv (\alpha^1, \alpha^2, \alpha^3)$ oraz β mają postać:

$$\alpha^l = \begin{pmatrix} 0 & \sigma^l \\ \sigma^l & 0 \end{pmatrix}, \quad l = 1, 2, 3 \quad \beta = \begin{pmatrix} \mathbf{1} & 0 \\ 0 & -\mathbf{1} \end{pmatrix} \quad (4.42)$$

gdzie $\sigma^l = \sigma_l$, ($l = 1, 2, 3$) są macierzami Pauliego, a $\mathbf{1}$ jest macierzą jednostkową:

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \mathbf{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (4.43)$$

W równaniach (4.39) oraz (4.40) macierz jednostkowa $\mathbf{1}$, która stoi (nie wpisana) przy D_0 oraz $\frac{N}{2} \times \frac{N}{2} = 4 \times 4$ wymiarowe macierz Diraca β , α_l , są jednymi z elementów grupy Clifforda $Pin(1, 3)$. Jak wspomnieliśmy, Frieden [7] przeprowadził opisaną decompozycję K_{KG} , (4.32), do postaci K_D podanej w (4.37). Równocześnie pokazał, że wyrażenie oznaczone w (4.38) jako “*pozostałe człony*” zeruje się przy założeniu, że macierze β i α_l spełniają relacje algebry Clifforda.

Podsumowując, zauważyliśmy, że pojemność informacyjna I w TFI jest zadana przez (4.35) zarówno dla pól N -skalarów jak i pól fermionowych rangi N . Kluczową sprawą jest, że w strukturalną zasadę informacyjną, $I + Q = 0$, która jest jednoznaczna dla każdego równania ruchu, wchodzi tylko jemu charakterystyczna postać Q . Zauważyliśmy, że dla N -skalarów, jedynym członem, który tworzy strukturalną zasadę informacyjną jest K_{KG} w postaci (4.32), bez żadnego ukrytego rozkładu i faktoryzacji, pochodzących z nie-Fourierowskiego splątania. W przypadku równania Diraca, K_{KG} , (4.32), również wchodzi w strukturalną zasadę informacyjną, ale tylko jako jej szczególna (jak to można również zauważyć z porównania (4.37), (4.38) oraz (4.32)) część, splątana Fourierowsko w K_F (4.23). Fakt ten jest blisko związany z efektem EPR-Bohm’a opisanym w Rozdziale 5.2.2. Tak więc, w zgodzie z (4.38), szczegółowa postać Q dla równania ruchu Diraca okazała się być inna niż dla N -skalarów, i można ją zapisać następująco:

$$Q = Q_D \equiv S_q - 4N \int d^4\mathbf{x} \sum_{n=1}^{N/2} \left[\left(\frac{m c}{\hbar} \right)^2 \psi_n^* \psi_n \right], \quad (4.46)$$

gdzie S_q jest charakterystyczną częścią Diracowską działania dla kwadratury równania Diraca, wyznaczoną dla pola Diraca o randze $N = 8$ przy wzięciu pod uwagę wszystkich symetrii układu. Ponieważ jednak cała działania dla kwadratury specyficznej części Diracowskiej zachowuje się następująco:

$$S_q = -4N \int d^4x \sum_{n=1}^{N/2} (v_{1n} v_{2n}) + (\text{pozostałe człony}) = 0, \quad (4.47)$$

tzn. spełnia warunek zerowy, zatem $K \equiv K_D$ redukuje się i (w kwadraturze) określa postać K_{KG} dla (4.32)¹²:

$$K = K_{KG} \equiv 4N \int d^4\mathbf{x} \sum_{n=1}^{N/2} \left[\sum_{\mu=0}^3 (D_\mu \psi_n)^* D^\mu \psi_n - \left(\frac{m c}{\hbar} \right)^2 \psi_n^* \psi_n \right]. \quad (4.50)$$

Uwaga: Ważną sprawą jest zauważenie, że chociaż postać I dla N -skalarów oraz pól fermionowych rangi N wygląda “powierzchniowo” tak samo, to jednak układy te różnią się istotnie. W samej bowiem rzeczy, podczas gdy N -skalar jest rozwiązaniem równania ruchu, które wynika jedynie z wariacyjnej zasady informacyjnej (4.34), to pole fermionowe jest samo-spójnym rozwiązaniem zarówno wariacyjnej jak i strukturalnej zasady informacyjnej podanych w (4.44).

¹²Równanie Kleina-Gordona otrzymane z wariacji informacji (4.50) ma postać $(\vec{\nabla} = (\partial/\partial x^l), l = 1, 2, 3)$:

$$-c^2 \hbar^2 (\vec{\nabla} - \frac{ie\vec{A}}{c\hbar}) \cdot (\vec{\nabla} - \frac{ie\vec{A}}{c\hbar}) \psi_n + \hbar^2 (\frac{\partial}{\partial t} + \frac{ie\phi}{\hbar})^2 \psi_n + m^2 c^4 \psi_n = 0. \quad (4.48)$$

Dla pola swobodnego cztero-potencjał cechowania $A_\mu = (\phi, -\vec{A})$ jest równy zero i wtedy otrzymujemy:

$$-c^2 \hbar^2 \nabla^2 \psi_n + \hbar^2 \frac{\partial^2}{\partial t^2} \psi_n + m^2 c^4 \psi_n = 0. \quad (4.49)$$

4.3.3 Końcowe uwagi o wkładzie Q w zasadę strukturalną

Wyprowadzenie strukturalnej zasady informacyjnej, $I+Q = 0$, (3.28) [10], zakłada, że funkcja logarytmu wiarygodności $\ln P(\Theta)$ może być w obszarze interesujących nas wartości parametru położenia Θ rozwinięta w szereg Taylora. Oznacza to, że zakładamy jej analityczność. Tego typu są właśnie funkcje wiarygodności dla amplitud pól rangi N , spełniających równanie Diraca i/bądź Kleina-Gordona omawiane w Rozdziałach 4.3.1 oraz 4.3.2.

Jeśli jednak na układ nałożony jest dodatkowy warunek, który nie wynika wprost ze strukturalnej zasady informacyjnej $I + Q = 0$, wtedy liczba stopni swobody (*l.st.sw.*), która wchodzi w Q od strony zdefiniowanej w (3.4) struktury \mathfrak{tF} , może zostać zredukowana. Na przykład (omawiany w Dodatku) warunek Lorentza dla pola cechowania Maxwella, który jest równaniem typu równania ciągłości strumienia, pojawia się nie jako konsekwencja równań ruchu badanego pola, ale jako ograniczenie szukane na drodze niezależnej statystycznej estymacji. Najprawdopodobniej ten dodatkowy warunek¹³, pojawia się jako rezultat rozwinięcia w szereg Taylora “gołej” funkcji wiarygodności $P(\Theta)$ [10], podobnie jak wyprowadzone w Rozdziale 3.4 równanie master, co sygnalizowałoby spójność całej statystycznej metody estymacyjnej¹⁴. Gdy jest on nałożony na układ redukując jego symetrię, wtedy fakt ten pociąga za sobą pojawienie się zasady strukturalnej $I + \kappa Q = 0$ z czynnikiem efektywności κ mniejszym niż 1. Tak też jest w przypadku równań Maxwella omówionych w Dodatku 7.3.1, gdzie zwrócimy uwagę [35], że wartość $\kappa = 1/2$ pojawia się automatycznie w metodzie EFI na skutek samo-spójnego rozwiązania równania strukturalnego, wariacyjnego i nałożonego warunku Lorentza.

¹³Jednak to przypuszczenie należy pokazać.

¹⁴Z drugiej strony same równania master leżą w innej części klasycznej statystycznej estymacji, tzn. w obszarze działania teorii procesów stochastycznych [40].

Rozdział 5

Przykłady z fizyki statystycznej i ekonofizyki oraz efekt EPR-Bohm'a

5.1 Wyprowadzenie klasycznej fizyki statystycznej z informacji Fishera

Celem rozdziału jest wyprowadzenie klasycznej fizyki statystycznej z teorii pomiaru. Teoria pomiaru jest analityczną procedurą, zgodnie z którą wprowadza się zasadniczo wszystkie prawa nauk empirycznych, w tym fizyki. Podstawowym narzędziem obecnego skryptu jest jej szczególna aplikacja oparta o pojęcie informacji Fishera oraz zasady informacyjne. Oryginalnie przedstawiona została przez Friedena i Soffera pod nazwą EFI (ekstremalna fizyczna informacja). Jej podstawy były systematycznie wprowadzane w Rozdziałach 1-4. Metoda EFI pokazuje w jaki sposób wychodząc z pojemności informacyjnej I i dodatkowych fizycznych więzów nałożonych na układ (które mają podstawę statystyczną), otrzymać informację strukturalną opisywanego układu. Tak więc postulatem metody EFI jest stwierdzenie, że wszystkie teorie fizyczne można uzyskać z wariacyjnej analizy statystycznej z więzami typu strukturalnego oraz równań typu master (np. równań ciągłości), a wynikające z niej wyprowadzenia równań ruchu lub równań stanu polegają na potraktowaniu wszystkich warunków nałożonych na układ jako związków na odchylenia (fluktuacje) wartości mierzonych od wartości oczekiwanych. Przy tym, analizowane dane pojawiają się jako efekt pomiaru dokonanego przez układ.

Obecny rozdział jest jako ilustracja EFI poświęcony otrzymaniu równań generujących, z których wyprowadzone zostaną rozkład Boltzmanna dla energii a następnie rozkład Maxwella-Boltzmanna dla pędu. Analiza została przedstawiona po raz pierwszy w [7]. Rachunki w obecnym skrypcie idą jej śladem, jednak zostanie ona wykonana w oparciu o wprowadzoną w Rozdziale 3.1.1 strukturalną zasadę informacyjną [10]. Różnice interpretacyjne pomiędzy obu podejściami zostały przedstawione w Rozdziale 3.2. Dodatkowo omówiony zostanie warunek informacyjny na tempo wzrostu entropii Shannona [7].

5.1.1 Fizyczne sformułowanie zagadnienia

Rozważmy gaz składający się z M identycznych cząsteczek o masie m zamknięty w zbiorniku. Temperatura gazu ma stałą wartość T . Ruch cząsteczek jest losowy i oddziałują one ze sobą poprzez siły potencjalne zderzając się ze sobą i ściankami naczynia, przy czym zakładamy, że są

to zderzenia sprężyste. Średnia prędkość każdej cząsteczki jest równa zero.

Oznaczmy przez $\theta_\varphi = (\theta_\epsilon, \vec{\theta}_\varphi)$ czterowektor wartości oczekiwanej energii oraz pędu cząsteczki, gdzie indeks φ oznacza pęd. We współrzędnych kartezjańskich $\vec{\theta}_\varphi = (\theta_{\varphi 1}, \theta_{\varphi 2}, \theta_{\varphi 3})$. Podobnie jak w (3.36) wprowadzamy zmienną losową $\mathbf{Y} = (\mathbf{Y}_\epsilon \equiv \frac{E}{c}, \vec{\mathbf{Y}}_\varphi)$ przyjmującą wartości $\mathbf{y} = (\mathbf{y}_\epsilon \equiv \epsilon/c, \vec{\mathbf{y}}_\varphi)$, której składowe spełniają związki:

$$\mathbf{y}_\epsilon \equiv \frac{\epsilon}{c} = \theta_\epsilon + \mathbf{x}_\epsilon, \quad \frac{\epsilon_0}{c} \leq \mathbf{y}_\epsilon \leq \infty \quad (5.1)$$

$$\vec{\mathbf{y}}_\varphi = \vec{\theta}_\varphi + \vec{\mathbf{x}}_\varphi, \quad \vec{\mathbf{y}}_\varphi = (\mathbf{y}_{\varphi 1}, \mathbf{y}_{\varphi 2}, \mathbf{y}_{\varphi 3}), \quad (5.2)$$

gdzie

$$\vec{\mathbf{x}}_\varphi = \vec{\varphi} \quad (5.3)$$

oznacza fluktuację pędu, natomiast c oznacza prędkość światła. Wszystkie współrzędne energetyczne $(\mathbf{y}_\epsilon, \theta_\epsilon, \mathbf{x}_\epsilon)$ są wyrażone w jednostkach $1/c$ tak, aby miały taką samą jednostkę jak współrzędne pędowe. Parametry θ_ϵ oraz $\vec{\theta}_\varphi$ są odpowiednimi wartościami oczekiwanymi energii (z dokładnością do $1/c$) oraz pędu, a \mathbf{x}_ϵ oraz $\vec{\mathbf{x}}_\varphi$ fluktuacjami względem wartości oczekiwanych.

Znajdziemy rozkład prawdopodobieństwa dla fluktuacji energii \mathbf{x}_ϵ oraz pędu $\vec{\varphi}$ dla (jednej) dowolnej cząsteczki gazu w dowolnej chwili czasu t . Ponieważ wartość t nie musi być duża, zatem rozważamy gaz, który nie koniecznie jest w stanie równowagi. Będziemy więc szukać ogólnej postaci nierównowagowego rozkładu prawdopodobieństwa odpowiadającego tak postawionemu problemowi.

Jednak po pierwsze współrzędne czterowektora fluktuacji energii-pędu $(\mathbf{x}_\epsilon, \vec{\mathbf{x}}_\varphi)$ nie są statystycznie niezależne, tzn. nie są niezależnymi stopniami swobody układu. Po drugie, jak się okaże, spełniają one zasadę dyspersyjną typu (4.20), więc nie tworzą układu zmiennych Fishera (porównaj (2.127)). Zatem problem ten wymagałby estymacji odpowiednich równań generujących dla skomplikowanego czaso-przestrzennego zagadnienia, tzn. należałoby wyznaczyć łączny rozkład prawdopodobieństwa współrzędnych $(\mathbf{x}_\epsilon, \vec{\mathbf{x}}_\varphi)$, co wykracza poza zakres skryptu. Nie mniej np. w przypadku relatywistycznych zjawisk astrofizycznych taka estymacja może okazać się niezbędna. W skrypcie ograniczymy się jedynie do wyznaczenia brzegowych rozkładów prawdopodobieństwa dla \mathbf{x}_ϵ (i w konsekwencji dla ϵ) oraz dla $\vec{\varphi}$, co w nierelatywistycznej granicy jest uzasadnione.

5.1.2 Informacja kinetyczna i strukturalna oraz sformułowanie zasad informacyjnych

Tak więc, statystycznie jedna cząsteczka podlega łącznemu rozkładowi prawdopodobieństwa $p(\mathbf{x}_\epsilon, \vec{\mathbf{x}}_\varphi)$, przy czym współrzędne czterowektora pędu nie są niezależne. Uproszczona analiza skoncentruje się na rozkładach brzegowych, których analiza ze względu na wspomniany brak niezależności nie odtwarza analizy łącznej, chociaż jest słuszna w przybliżeniu nierelatywistycznym. Wyznamy więc brzegowe amplitudy prawdopodobieństwa $q(\mathbf{x}_\epsilon)$ oraz $q(\vec{\mathbf{x}}_\varphi)$:

$$q_n(\mathbf{x}_\epsilon), \quad n = 1, \dots, N_\epsilon \quad \text{oraz} \quad q_n(\vec{\mathbf{x}}_\varphi), \quad n = 1, \dots, N_\varphi, \quad (5.4)$$

a następnie powrócimy do brzegowych rozkładów prawdopodobieństwa $p(\mathbf{x}_\epsilon)$ oraz $p(\vec{\mathbf{x}}_\varphi)$. W końcu po wyznaczeniu $p(\mathbf{x}_\epsilon)$ skorzystamy z (5.1) aby uzyskać wymagany rozkład $p(\epsilon)$. Podobnie korzystając z (5.2) otrzymamy po wyznaczeniu $p(\vec{\mathbf{x}}_\varphi)$ rozkład prędkości $p(\vec{\mathbf{y}}_\varphi)$, przy czym te dwa ostatnie są równe w naszych rozważaniach, ze względu na na średnią wartość prędkości cząsteczki $\vec{\theta}_\varphi = 0$.

Chociaż $(\mathbf{x}_\epsilon, \vec{\mathbf{x}}_\varphi)$ nie jest czterowektorem we wspomnianym ujęciu Fishera to jest on czterowektorem w sensie Minkowskiego. Korzystamy więc z metryki czasoprzestrzeni Minkowskiego w postaci $(\eta_{\nu\mu}) = \text{diag}(1, -1, -1, -1)$ zgodnie z (2.125). Ponieważ w ogólnym przypadku rozkładu łącznego i jednoczesnego próbkowania czterowektora energii i pędu (przy zachowaniu nieoznaczoności wynikającej z nierówności Rao-Cramera), pojemność informacyjna miałaby postać (2.134)¹ dla parametru wektorowego $\Theta = ((\theta_{\nu n})_{\nu=0}^3)_{n=1}^N$:

$$I = \sum_{n=1}^N \int dy P(y|\Theta) \left[\left(\frac{\partial \ln P(y|\Theta)}{\partial \theta_{\epsilon n}} \right)^2 - \sum_{k=1}^3 \left(\frac{\partial \ln P(y|\Theta)}{\partial \theta_{\varphi k n}} \right)^2 \right], \quad (5.5)$$

gdzie $y = (\mathbf{y})_{\mathbf{n}=1}^N$ jest N -wymiarową próbą. Zatem informacje Fishera $I(\Theta_\epsilon)$ oraz $I(\Theta_{\vec{\varphi}})$ dla rozkładów brzegowych fluktuacji energii oraz pędu mają postać²:

$$I(\Theta_\epsilon) = 4 \int d\mathbf{x}_\epsilon \sum_{n=1}^{N_\epsilon} \left(\frac{dq_n(\mathbf{x}_\epsilon)}{d\mathbf{x}_\epsilon} \right)^2 \quad (5.6)$$

oraz

$$I(\Theta_{\vec{\varphi}}) = -4 \int d\vec{\mathbf{x}}_\varphi \sum_{n=1}^{N_\varphi} \sum_{k=1}^3 \left(\frac{dq_n(\vec{\mathbf{x}}_\varphi)}{dx_{\varphi k}} \right)^2, \quad (5.7)$$

gdzie, podkreślmy jeszcze raz, minus w (5.7) jest konsekwencją przyjętej metryki (2.125) odzwierciedlonej w (5.5) zgodnie z rozważaniami Rozdziału 2.7.1. Z poniższych rachunków przekonamy się, że nie uwzględnienie metryki Minkowskiego (gdy $\eta_{00} = 1$ to $\eta_{kk} = -1$ dla $k = 1, 2, 3$), doprowadziłoby w konsekwencji w termodynamicznych rozważaniach do błędnego rozkładu dla rozkładu prędkości³.

Aby skorzystać z zasad informacyjnych, wariacyjnej (3.32) oraz strukturalnej (3.28) dla energii:

$$\delta_{(q_n)} (I(\Theta_\epsilon) + Q(\Theta_\epsilon)) = 0, \quad I(\Theta_\epsilon) + \kappa Q(\Theta_\epsilon) = 0, \quad (5.8)$$

oraz pędu:

$$\delta_{(q_n)} (I(\Theta_{\vec{\varphi}}) + Q(\Theta_{\vec{\varphi}})) = 0, \quad I(\Theta_{\vec{\varphi}}) + \kappa Q(\Theta_{\vec{\varphi}}) = 0, \quad (5.9)$$

¹W zgodzie z ogólną własnością kontrakcji indeksu Minkowskiego dla czterowektora $x \equiv (x_\nu)_{\nu=0}^3 = (x_0, x_k)$:

$$\sum_{\nu\mu=0}^3 \eta_{\nu\mu} \frac{df}{dx_\nu} \frac{df}{dx_\mu} = \left(\frac{df}{dx_0} \right)^2 - \sum_{k=1}^3 \left(\frac{df}{dx_k} \right)^2.$$

W [7] metryka ma postać Euklidesową a zmienne mają współrzędną czasową urojoną (ix_0, \vec{x}_k) .

²Jak w Rozdziale 3.3 indeks n przy współrzędnej pominięto, korzystając z założenia, że rozproszenie zmiennej nie zależy od punktu pomiarowego próby.

³Odpowiedź na pytanie czy rozważania termodynamiczne są przyczyną metryki Minkowskiego niezbędnej w relatywistycznej teorii pola, wykracza poza obszar skryptu. Nie mniej autor skryptu uważa, że tak się istotnie sprawy mają, tzn. że przestrzeń Euklidesowa z transformacją Galileusza są pierwotne wobec przestrzeni Minkowskiego z transformacją Lorentza. Stąd podejście efektywnej teorii pola Logunova [34] do teorii grawitacji jest bliższe teorii pomiaru fizycznego Friedena-Soffera (którą jest EFI). Nieco więcej na ten temat można znaleźć w Dodatku.

należy wyznaczyć informacje strukturalne $Q(\Theta_\epsilon)$ oraz $Q(\Theta_{\bar{\theta}})$. W pierwszej kolejności rozważymy problemem (5.8) dla $p(\epsilon)$.

5.1.3 Rozkład Boltzmana dla energii

Założmy wstępnie, że wartość fluktuacji energii \mathbf{x}_ϵ zmienia się w pewnym zakresie $\langle \mathbf{x}_\epsilon^{min}, \mathbf{x}_\epsilon^{max} \rangle$:

$$\mathbf{x}_\epsilon^{min} \leq \mathbf{x}_\epsilon \leq \mathbf{x}_\epsilon^{max} . \quad (5.10)$$

W ten sposób wpierym uchwycimy ogólną zależność amplitudy rozkładu od b , a następnie dokonamy przejścia granicznego, przechodząc z górną granicą fluktuacji energii b do nieskończoności. Informacja Fishera ma postać (3.40):

$$I(\Theta_\epsilon) = 4 \int_{\mathbf{x}_\epsilon^{min}}^{\mathbf{x}_\epsilon^{max}} d\mathbf{x}_\epsilon \sum_{n=1}^N q_n'^2(\mathbf{x}_\epsilon) \quad \text{gdzie} \quad q_n'(\mathbf{x}_\epsilon) \equiv \frac{dq_n(\mathbf{x}_\epsilon)}{d\mathbf{x}_\epsilon} , \quad (5.11)$$

gdzie rozkłady prawdopodobieństwa p_n dla fluktuacji energii są powiązane z amplitudami q_n zależnością (2.75):

$$p_n(\mathbf{x}_\epsilon) = q_n^2(\mathbf{x}_\epsilon) . \quad (5.12)$$

Natomiast informacja strukturalna ma postać (3.21):

$$Q(\Theta_\epsilon) = \int d\mathbf{x}_\epsilon \sum_{n=1}^N q_n^2(\mathbf{x}_\epsilon) \mathbf{F}_n(q_n(\mathbf{x}_\epsilon)) \quad (5.13)$$

Wariacyjna IP przyjmuje zgodnie z (3.32) i (4.6), postać:

$$\delta_{(q_n)} K = \delta_{(q_n)} (I(\Theta_\epsilon) + Q(\Theta_\epsilon)) = \delta_{(q_n)} \left(\int d\mathbf{x}_\epsilon k \right) = 0 , \quad (5.14)$$

gdzie k jest mikroskopową informacją fizyczną (4.7):

$$k = 4 \sum_{n=1}^N \left(q_n'^2 + \frac{1}{4} q_n^2(\mathbf{x}_\epsilon) \mathbf{F}_n(q_n(\mathbf{x}_\epsilon)) \right) . \quad (5.15)$$

Rozwiązaniem problemu wariacyjnego (5.14) względem $q_n(\mathbf{x}_\epsilon)$ jest równanie Eulera-Lagrange'a (4.12):

$$\frac{d}{d\mathbf{x}_\epsilon} \left(\frac{\partial k}{\partial q_n'(\mathbf{x}_\epsilon)} \right) = \frac{\partial k}{\partial q_n(\mathbf{x}_\epsilon)} , \quad \text{dla } n = 1, 2, \dots, N , \quad (5.16)$$

gdzie 'prim' oznacza pochodną po \mathbf{x}_ϵ , tzn. $q_n'(\mathbf{x}_\epsilon) \equiv dq_n(\mathbf{x}_\epsilon)/d\mathbf{x}_\epsilon$. Równanie (5.16) dla naszego problemu przyjmuje postać:

$$2 q_n''(\mathbf{x}_\epsilon) = \frac{\partial(\frac{1}{4} q_n^2(\mathbf{x}_\epsilon) \mathbf{F}_n(q_n(\mathbf{x}_\epsilon)))}{\partial q_n} = \frac{d(\frac{1}{4} q_n^2(\mathbf{x}_\epsilon) \mathbf{F}_n(q_n(\mathbf{x}_\epsilon)))}{dq_n} . \quad (5.17)$$

Całkując (5.11) przez części, I przyjmuje postać:

$$I(\Theta_\epsilon) = -4 \int_a^b dx \sum_{n=1}^N q_n(\mathbf{x}_\epsilon) q_n''(\mathbf{x}_\epsilon) + I_C , \quad \text{gdzie} \quad I_C \equiv 4 \sum_{n=1}^N C_n , \quad (5.18)$$

oraz

$$C_n = q_n(\mathbf{x}_\epsilon^{max}) q_n'(\mathbf{x}_\epsilon^{max}) - q_n(\mathbf{x}_\epsilon^{min}) q_n'(\mathbf{x}_\epsilon^{min}) . \quad (5.19)$$

Biorąc pod uwagę (5.18), strukturalna IP ma postać (3.28):

$$I + \kappa Q = 4 \int d\mathbf{x}_\epsilon \sum_{n=1}^N \left(C_n - q_n(\mathbf{x}_\epsilon) q_n''(\mathbf{x}_\epsilon) + \kappa \frac{1}{4} q_n^2(\mathbf{x}_\epsilon) \mathfrak{F}_n(q_n(\mathbf{x}_\epsilon)) \right) = 0 , \quad (5.20)$$

zatem na poziomie mikroskopowym otrzymujemy strukturalną IP w postaci⁴:

$$C_n - q_n(\mathbf{x}_\epsilon) q_n''(\mathbf{x}_\epsilon) + \kappa \frac{1}{4} q_n^2(\mathbf{x}_\epsilon) \mathfrak{F}_n(q_n(\mathbf{x}_\epsilon)) = 0 \quad (5.22)$$

Korzystając z (5.17) w (5.22) otrzymujemy:

$$\frac{1}{2} q_n \frac{d(q_n^2(\mathbf{x}_\epsilon) \mathfrak{F}_n(q_n(\mathbf{x}_\epsilon)))}{dq_n} = \kappa q_n^2(\mathbf{x}_\epsilon) \mathfrak{F}_n(q_n(\mathbf{x}_\epsilon)) + 4 C_n . \quad (5.23)$$

Poniżej, dla uproszczenia zapisu pominiemy przy rozwiązywaniu równania(5.23) oznaczenie argumentu \mathbf{x}_ϵ w amplitudzie q_n . Zapiszmy (5.23) w postaci:

$$2 \frac{dq_n}{q_n} = \frac{d \left[\frac{1}{4} q_n^2 \mathfrak{F}_n(q_n) \right]}{\kappa \left[\frac{1}{4} q_n^2 \mathfrak{F}_n(q_n) \right] + C_n} . \quad (5.24)$$

Rozwiązując powyższe równanie różniczkowe otrzymujemy kolejno:

$$2 \ln q_n + \alpha_n' = \frac{1}{\kappa} \ln \left(\frac{\kappa}{4} q_n^2 \mathfrak{F}_n(q_n) + C_n \right)$$

i po przekształceniu:

$$2\kappa \ln \alpha_n'' q_n = \ln \left(\frac{\kappa}{4} q_n^2 \mathfrak{F}_n(q_n) + C_n \right) ,$$

gdzie α_n' jest w ogólności zespoloną stałą całkowania, oraz stała $\alpha_n'' = \exp(\alpha_n'/2)$, skąd:

$$q_n^2(\mathbf{x}_\epsilon) \mathfrak{F}_n(q_n) = \frac{4}{\kappa} (\alpha_n^2 q_n^{2\kappa} - C_n) \quad (5.25)$$

gdzie (w ogólności zespolona) stała $\alpha_n^2 = \alpha_n''^{2\kappa}$. Zatem w rezultacie otrzymaliśmy obserwowaną SI, $\mathfrak{F}_n(q_n(\mathbf{x}_\epsilon))$, jako znaną funkcję amplitudy $q_n(\mathbf{x}_\epsilon)$.

W końcu, korzystając z (5.17) oraz z (5.25) otrzymujemy równanie różniczkowe dla $q_n(\mathbf{x}_\epsilon)$:

$$q_n''(\mathbf{x}_\epsilon) = \alpha_n^2 q_n^{2\kappa-1}(\mathbf{x}_\epsilon) . \quad (5.26)$$

Podsumowując, wariacyjna oraz strukturalną mikroskopowa IP zostały zapisane w postaci układu równań (5.17) oraz (5.22), który rozwiązując, dał w rezultacie szukaną postać (5.26) *równania generującego* amplitudę $q_n(\mathbf{x}_\epsilon)$. Znalezienie postaci tego równania było celem metody EFI.

⁴Gdyby nie wycalkować w (5.18) I przez części, wtedy korzystając bezpośrednio z postulatu (3.10), strukturaln IP na poziomie mikroskopowym miała by postać (4.11):

$$q_n'^2(\mathbf{x}_\epsilon) + \frac{\kappa}{4} q_n^2(\mathbf{x}_\epsilon) \mathfrak{F}_n(q_n(\mathbf{x}_\epsilon)) = 0 . \quad (5.21)$$

Rozwiążmy równanie generujące (5.26). Z jego postaci widać, że skoro amplituda $q_n(\mathbf{x}_\epsilon)$ jest funkcją rzeczywistą, zatem stała α_n^2 musi być rzeczywista. Rozwiązanie znajdziemy dla przypadku, gdy:

$$\kappa = 1 \quad , \quad (5.27)$$

czyli dla przypadku gdy zasada strukturalna jest konsekwencją “trywialnego” rozwinięcia logarytmu funkcji wiarygodności układu w szereg Taylora. W przypadku tym mikroskopowa IS (5.25) przyjmuje prostą postać:

$$q_n^2(\mathbf{x}_\epsilon) \Phi_n(q_n(\mathbf{x}_\epsilon)) = 4(\alpha_n^2 q_n^2(\mathbf{x}_\epsilon) - C_n) \quad , \quad (5.28)$$

a całkowa postać IS jest następująca:

$$Q(\Theta_\epsilon) = 4 \int d\mathbf{x}_\epsilon \sum_{n=1}^N \alpha_n^2 q_n^2(\mathbf{x}_\epsilon) - 4 \sum_{n=1}^N C_n \quad , \quad (5.29)$$

co po skorzystaniu z unormowania kwadratu amplitudy:

$$\int d\mathbf{x}_\epsilon q_n^2(\mathbf{x}_\epsilon) = \int d\mathbf{x}_\epsilon p_n(\mathbf{x}_\epsilon) = 1 \quad , \quad (5.30)$$

daje:

$$Q(\Theta_\epsilon) = 4 \sum_{n=1}^N (\alpha_n^2 - C_n) = 4 \sum_{n=1}^N \alpha_n^2 - I_C \quad . \quad (5.31)$$

W ostatnim przejściu w (5.31) skorzystano z postaci stałej $I_C \equiv 4 \sum_{n=1}^N C_n$ wprowadzonej w (5.18).

Dla rozważanego przypadku $\kappa = 1$, równanie generujące (5.26) ma postać:

$$q_n''(\mathbf{x}_\epsilon) = \alpha_n^2 q_n(\mathbf{x}_\epsilon) \quad . \quad (5.32)$$

Wstawiając (5.32) do (5.18), otrzymujemy następującą postać pojemności informacyjnej:

$$I(\Theta_\epsilon) = -4 \sum_{n=1}^N \alpha_n^2 + I_C \quad , \quad (5.33)$$

co wraz z (5.31) oznacza sprawdzenie poprawności rachunku, poprzez spełnienie przez znalezione rozwiązanie wyjściowego strukturalnego warunku $I + Q = 0$.

Kolejnym etapem analizy jest rozwiązanie równania generującego (5.32). Najogólniejsza postać amplitudy $q_n(\mathbf{x}_\epsilon)$ będącej rozwiązaniem równania (5.32) jest przy warunku $\mathbf{x}_\epsilon^{\min} \leq \mathbf{x}_\epsilon \leq \mathbf{x}_\epsilon^{\max}$ jak w (5.10), następująca:

$$q_n(\mathbf{x}_\epsilon) = B_n \exp(\alpha_n \mathbf{x}_\epsilon) + D_n \exp(-\alpha_n \mathbf{x}_\epsilon) \quad , \quad \mathbf{x}_\epsilon^{\min} \leq x \leq \mathbf{x}_\epsilon^{\max} \quad , \quad B_n, D_n = const. \quad (5.34)$$

Ponieważ stała α_n^2 jest rzeczywista, zatem wprowadzając nową rzeczywistą stałą β_n , można α_n przedstawić jako $\alpha_n = \beta_n$, i wtedy rozwiązanie (5.34) ma charakter czysto eksponencjalny:

$$q_n(\mathbf{x}_\epsilon) = B_n \exp(\beta_n \mathbf{x}_\epsilon) + D_n \exp(-\beta_n \mathbf{x}_\epsilon) \quad , \quad (5.35)$$

bądź jako $\alpha_n = i\beta_n$, i wtedy rozwiązanie (5.34) ma charakter czysto trygonometryczny:

$$q_n(\mathbf{x}_\epsilon) = B_n \exp(i\beta_n \mathbf{x}_\epsilon) + D_n \exp(-i\beta_n \mathbf{x}_\epsilon) . \quad (5.36)$$

Tak określone funkcje muszą być dopuszczalne jako amplitudy prawdopodobieństwa, zatem muszą spełniać warunek normalizacji dla gęstości prawdopodobieństwa (5.30).

Załóżmy, że wartości fluktuacji energii \mathbf{x}_ϵ nie jest ograniczona od góry, co zrealizujemy jako dążenie $\mathbf{x}_\epsilon^{max}$ do nieskończoności. Jednak kwadrat funkcji trygonometrycznej nie może być unormowany do jedności dla $\mathbf{x}_\epsilon^{max} \rightarrow \infty$, zatem funkcja trygonometryczna (5.36) nie jest dopuszczalnym rozwiązaniem.

Pozostaje więc rozwiązanie eksponencjalne (5.35). Ponieważ jednak warunek unormowania (5.30) ma być zadany na przedziale otwartym $\mathbf{x}_\epsilon^{min} \leq \mathbf{x}_\epsilon < \infty$, zatem część rozwiązania z dodatnią eksponentą musi być odrzucona ze względu na jej rozbieżność do nieskończoności. Skąd otrzymujemy żądanie, że dla $\beta_n \geq 0$ stała $B_n = 0$.

Ostatecznie szukana postać amplitudy jest więc następująca:

$$q_n(\mathbf{x}_\epsilon) = D_n \exp(-\beta_n \mathbf{x}_\epsilon) , \quad \beta_n \in \mathbf{R}_+ , \quad \mathbf{x}_\epsilon^{min} \leq \mathbf{x}_\epsilon < \infty . \quad (5.37)$$

Z powyższego i z warunku normalizacji (5.30) $\int_{\mathbf{x}_\epsilon^{min}}^{\infty} d\mathbf{x}_\epsilon q_n^2(\mathbf{x}_\epsilon) = 1$, wyznaczamy stałą D_n , otrzymując:

$$D_n = \sqrt{2\beta_n} \exp(\beta_n \mathbf{x}_\epsilon^{min}) . \quad (5.38)$$

Rozwiązanie to zostało otrzymane dla przypadku $\beta_n = \alpha_n \in \mathbf{R}_+$, zatem ostateczną postacią (5.34) jest:

$$q_n(\mathbf{x}_\epsilon) = \sqrt{2\alpha_n} \exp[\alpha_n (\mathbf{x}_\epsilon^{min} - \mathbf{x}_\epsilon)] , \quad \alpha_n \in \mathbf{R}_+ . \quad (5.39)$$

Zauważmy, że α_n jest w jednostkach [*c/energia*].

W końcu możemy wyznaczyć pojemność informacyjną. Wstawiając (5.39) do (5.11), otrzymujemy:

$$I(\Theta_\epsilon) = 4 \sum_{n=1}^N \alpha_n^2 > 0 , \quad \alpha_n \in \mathbf{R}_+ , \quad (5.40)$$

co po porównaniu z (5.33) daje wartość stałej I_C równą:

$$I_C = 8 \sum_{n=1}^N \alpha_n^2 . \quad (5.41)$$

Uwaga: Warunek dodatniości pojemności informacyjnej otrzymany w (5.40), dla pojemności informacyjnej związanej z dodatnią częścią sygnatury metryki Minkowskiego, jest istotnym wynikiem z punktu widzenia z teorii pomiaru. W rozważanym przykładzie analizy estymacyjnej wartości oczekiwanej energii cząstki gazu, został on otrzymany na gruncie samospójnego rozwiązania równań różniczkowych zasady informacyjnej strukturalnej i wariacyjnej. Niespełnienie tego warunku oznacza niestabilność badanego układu. Z kolei, wstawiając otrzymaną wartość stałej I_C do (5.31) otrzymujemy:

$$Q(\Theta_\epsilon) = -4 \sum_{n=1}^N \alpha_n^2 < 0 . \quad (5.42)$$

Natomiast sam problem wariacyjny (5.14), który można wyrazić po skorzystaniu z postaci $Q(\Theta_\epsilon)$ w (5.31) w postaci:

$$\delta_{(q_n)} \left(I(\Theta_\epsilon) + 4 \sum_{n=1}^N \alpha_n^2 - I_C \right) = 0, \quad (5.43)$$

jest ze swojej natury nieczuły na wartość stałej I_C .

Istotnym zagadnieniem metody EFI jest wielkość próby N pobranej przez układ, tzn. ranga N amplitudy. Do sprawy liczby amplitud q_n wchodzących w opis układu podejmiemy w najprostszy z możliwych sposobów. Pierwsze ze stosowanych kryteriów wyboru N zakłada taki wybór N dla zrealizowanych w przyrodzie równań stanu (i równań ruchu), który minimalizuje pojemność informacyjną [7]. Sugeruje to, konieczność wzrostu informacji o układzie (zawartej w I) ale tylko na tyle na ile wymaga tego struktura układu (opisana informacją strukturalną Q). Tzn. przy zachowaniu warunku $I > 0$ ranga N może na tyle spaść, żeby samo-spójne rozwiązanie równań cząstkowych strukturalnej i wariacyjnej zasady informacyjnej było jeszcze rozwiązaniem stacjonarnym. Kryterium to nie oznacza jednak nie realizowania rozwiązań z większą niż minimalna liczbą N . Rozważany przykład rozkładu energii i kolejny, rozkładu prędkości cząsteczki gazu, pozwalają na dobrą ilustrację tego kryterium.

Otóż zauważmy, że ponieważ wszystkie α_n w (5.40) są rzeczywiste, zatem, przy założeniu braku wpływu nowych stopni swobody na poprzednie, pojemność informacyjna I wzrasta wraz ze wzrostem N . Zgodnie z powyższym kryterium, przyjmujemy dla rozważanego przypadku rozkładu fluktuacji energii, że:

$$N = 1, \quad (5.44)$$

a jedyny współczynnik α_1 oznaczmy α . Zatem z (5.39) mamy amplitudę:

$$q(\mathbf{x}_\epsilon) = \sqrt{2\alpha} \exp [\alpha (\mathbf{x}_\epsilon^{min} - \mathbf{x}_\epsilon)], \quad (5.45)$$

której odpowiada rozkład gęstości prawdopodobieństwa fluktuacji energii \mathbf{x}_ϵ :

$$p(\mathbf{x}_\epsilon) = q^2(\mathbf{x}_\epsilon) = 2\alpha \exp [2\alpha (\mathbf{x}_\epsilon^{min} - \mathbf{x}_\epsilon)], \quad \alpha \in \mathbf{R}_+. \quad (5.46)$$

Na koniec, wyznaczmy rozkład gęstości prawdopodobieństwa energii ϵ cząsteczki. Ponieważ zgodnie z (5.1) mamy:

$$\mathbf{y}_\epsilon \equiv \frac{\epsilon}{c} = \theta_\epsilon + \mathbf{x}_\epsilon, \quad \frac{\epsilon_0}{c} \leq \mathbf{y}_\epsilon < \infty, \quad \text{dla } \mathbf{x}_\epsilon^{min} \leq \mathbf{x}_\epsilon < \infty, \quad (5.47)$$

zatem $d\epsilon/d\mathbf{x}_\epsilon = c$, skąd rozkład dla zmiennej ϵ ma postać:

$$p(\epsilon) = p(\mathbf{x}_\epsilon) \frac{1}{|d\epsilon/d\mathbf{x}_\epsilon|} = 2\frac{\alpha}{c} \exp [-2\alpha (\epsilon - \epsilon_0)/c], \quad \epsilon_0 \leq \epsilon < \infty, \quad (5.48)$$

gdzie $\epsilon_0/c = \theta_\epsilon + \mathbf{x}_\epsilon^{min}$.

Pozostało jeszcze określenie stałej α . Ponieważ wartość oczekiwana energii wynosi:

$$\langle E \rangle \equiv \theta_E = \int_{\epsilon_0}^{+\infty} d\epsilon p(\epsilon) \epsilon, \quad (5.49)$$

więc wstawiając (5.48) do (5.49) otrzymuje się:

$$2\alpha = c(\langle E \rangle - \epsilon_0)^{-1} . \quad (5.50)$$

Podsumowując powyższe rachunki, szukany rozkład gęstości prawdopodobieństwa ma postać:

$$p(\epsilon) = \begin{cases} (\langle E \rangle - \epsilon_0)^{-1} \exp[-(\epsilon - \epsilon_0)/(\langle E \rangle - \epsilon_0)] & \text{dla } \epsilon \geq \epsilon_0 \\ 0 & \text{dla } \epsilon < \epsilon_0 \end{cases} , \quad (5.51)$$

gdzie w drugiej linii po prawej stronie zaznaczono fakt nie występowania cząstek z energią mniejszą niż ϵ_0 . Rozkład (5.51) jest końcowym rezultatem metody EFI. Jego postać daje zasadniczo rozkład Boltzmana dla energii cząsteczki w gazie.

Uwaga: Aby domknąć temat od strony fizycznej zauważmy, że energia całkowita $\epsilon = \epsilon_{kin} + V$, gdzie ϵ_{kin} jest energią kinetyczną cząsteczki a V jej energią potencjalną. Do potencjału V możemy dodać pewną stałą np. ϵ_0 , nie zmieniając przy tym fizycznego opisu zjawiska, zatem po przesunięciu ϵ o ϵ_0 otrzymujemy:

$$\epsilon_0 = 0 , \quad \text{oraz } \epsilon \geq 0 . \quad (5.52)$$

Z kolei, dla cząstki gazu poruszającej się bez obrotu, twierdzenie o ekwipartycji energii mówi, że:

$$\langle E \rangle = \frac{3kT}{2} , \quad (5.53)$$

gdzie T jest temperaturą bezwzględną gazu. Wstawiając (5.52) wraz z (5.53) do (5.51) otrzymujemy:

$$p(\epsilon) = (3kT/2)^{-1} e^{-2\epsilon/3kT} , \quad \epsilon \geq 0 . \quad (5.54)$$

czyli właściwą postać rozkładu Boltzmana dla energii cząsteczki w gazie o temperaturze T .

Pozostało drugie kryterium wyboru rangi amplitudy. Wiąże się ono z wyborem takiej wartości N , dla której otrzymane rozwiązanie ma realizację zauważoną w obserwacji zjawiska. Tak się składa, że we współczesnych teoriach fizyki statystycznej oraz teorii pola realizowane są rozwiązania z niskimi wartościami rangi. Zauważyliśmy ten fakt już w Rozdziale 4 dla modeli teorii pola. W kolejnym z rozdziałów znajdziemy rozwiązanie EFI dla rozkładu prędkości cząsteczki gazu. Jednak obok stacjonarnego rozwiązania Maxwella-Boltzmana z $N = 1$, wskażemy również na fizyczną interpretację rozwiązań z $N > 1$.

Uwaga o równowadze statystycznej w metodzie EFI: Metoda EFI, biorąc pod uwagę ograniczenia narzucone na normalizację (5.30) i skończoność wartości oczekiwanej (5.49) rozkładu, wyznacza rozkład przy narzuceniu warunków informacyjnych. Dokonuje ona (dla gładkiej funkcji wiarygodności) poprzez zasadę strukturalną, separacji członu estymacyjnego dla parametru θ_ϵ związanego z krzywizną przestrzeni statystycznej od członu strukturalnego, co jest odbite w zasadzie Macha. Natomiast, poprzez zasadę wariacyjną, dokonuje stabilizacji rozwiązania, przy możliwie najmniejszej odległości stanów układu od innych możliwych.

5.1.4 Model Aoki-Yoshikawy dla produktywności branż

Model Aoki i Yoshikawy (AYM) został opracowany w celu opisu produktywności branż kraju [41, 42]. Rozważmy g ekonomicznych sektorów. Sektor i -ty jest scharakteryzowany przez *czynnik wielkości produkcji* n_i (tzn. liczebność) oraz zmienną poziomu produktywności A (tzn. wydajność jednostkowa) przyjmującą wartości a_i .

Niech n_i będzie *liczbą aktywnych pracowników w i -tym sektorze*, co oznacza, że praca (robocizna) jest jedynym czynnikiem produkcyjnym. Zatem zmienną losową w AYM jest poziom produktywności A , której rozkład jest określony poprzez parę (a_i, n_i) , $i = 1, 2, \dots, g$. Załóżmy, że całkowity zasób czynnika wielkości produkcji (tzn. liczba dostępnych pracowników) jest w ekonomii zadany jako wielkość egzogeniczna, czyli nie kontrolowana od wewnątrz lecz zadana z zewnątrz. Niech jego wielkość jest równa n , co traktujemy jako *pierwsze ograniczenie* w modelu, tak, że zachodzi:

$$\sum_{i=1}^g n_i = n . \quad (5.55)$$

Uporządkujmy wielkość produktywności a_i w sektorach od najmniejszej do największej:

$$a_1 < a_2 < \dots < a_g . \quad (5.56)$$

Ponieważ a_i jest poziomem produktywności i -tego sektora, zatem uzysk (wartość produkcji) w i -tym sektorze wynosi:

$$z_i = a_i n_i . \quad (5.57)$$

Zatem całkowity uzysk z w ekonomii kraju wynosi:

$$z = \sum_{i=1}^g z_i = \sum_{i=1}^g a_i n_i . \quad (5.58)$$

Wielkość z jest interpretowana jako *produkt krajowy brutto* (PKB).

Założeniem AYM dla wartości PKB jest ustalenie wartości z poprzez egzogenicznie zadany agregatowy popyt D (demand), tzn.:

$$z = D . \quad (5.59)$$

Zatem *drugie ograniczenie* w modelu ma postać:

$$\sum_{i=1}^g a_i n_i = D . \quad (5.60)$$

Celem AYM jest wyznaczenie teoretycznego rozkładu liczebności dla zmiennej poziomu produktywności A , tzn. określenie *wektora obsadzeń*:

$$\mathbf{n} = (n_1, n_2, \dots, n_g) . \quad (5.61)$$

Zagadnienie to jest odpowiednikiem powyższego zagadnienia związanego z określeniem rozkładu prawdopodobieństwa energii cząstki gazu, które można wyrazić jako zagadnienie rozmieszczenia n cząstek gazu na g poziomach energetycznych ϵ_i w warunkach równowagi statystycznej, w taki sposób, że zachowane są, liczba cząstek:

$$\sum_{i=1}^g n_i = n , \quad (5.62)$$

oraz całkowita energia gazu \mathcal{E} :

$$\sum_{i=1}^g \epsilon_i n_i = \mathcal{E} \quad \text{lub} \quad \langle E \rangle = \sum_{i=1}^g \frac{n_i}{n} \epsilon_i = \frac{\mathcal{E}}{n} \quad (5.63)$$

Zatem poziom produktywności a_i jest analogiem poziomu energetycznego ϵ_i , natomiast ograniczenie, które dał w AYM popyt D jest analogiem ograniczenia pochodzącego od wartości całkowitej energii \mathcal{E} gazu.

Dokonajmy następującego przyporządkowania pomiędzy wielkościami opisującymi rozkład energii cząstki gazu oraz rozkład poziomu produktywności. Po lewej stronie przyporządkowania " \leftrightarrow " jest wielkość dla energii cząstki, po prawej dla produktywności pracownika. Strałki \rightarrow w nawiasach oznaczają przejście od ciągłego do dyskretnego rozkładu zmiennej (lub na odwrót).

Zatem, zmiennej energii cząstki E odpowiada poziom produktywności pracownika A :

$$E = (\epsilon \rightarrow \epsilon_i) \leftrightarrow A = (a_i \rightarrow a) , \quad (5.64)$$

Rozkładowi prawdopodobieństwa zmiennej energii cząstki $p(\epsilon)$ odpowiada rozkład poziom produktywności pracownika $p(a)$:

$$\left(p(\epsilon) \rightarrow p_{\epsilon_i} = \frac{n_i}{n} \right) \leftrightarrow \left(p_i = \frac{n_i}{n} \rightarrow p(a) \right) , \quad (5.65)$$

Normalizacje rozkładów są sobie przyporządkowane następująco:

$$\left(\int_{\mathcal{Y}_\epsilon} d\epsilon p(\epsilon) = 1 \rightarrow \sum_{i=1}^g p_{\epsilon_i} = 1 \right) \leftrightarrow \left(\sum_{i=1}^g p_i = 1 \rightarrow \int_{\mathcal{X}_a} da p(a) = 1 \right) \quad (5.66)$$

Wartości oczekiwanej (średniej) energii cząstki odpowiada wartość oczekiwana wartość produktywności pracownika:

$$\theta_E \equiv \langle E \rangle = \left(\int_{\mathcal{Y}_\epsilon} d\epsilon p(\epsilon) \epsilon \rightarrow \sum_{i=1}^g \frac{n_i}{n} \epsilon_i \right) \leftrightarrow \left(\sum_{i=1}^g p_i a_i \rightarrow \int_{\mathcal{X}_a} da p(a) a \right) = \langle A \rangle \equiv \theta_E , \quad (5.67)$$

przy czym w AYM wartość oczekiwana produktywności jest zadana jako:

$$\langle A \rangle = D/n . \quad (5.68)$$

Aby analiza w AYM mogła przebiegać dokładnie tak samo jak dla rozkładu Boltzmana, musimy dokonać jeszcze jednego przejścia po stronie produktywności, a mianowicie przejść od poziomu produktywności A do jej fluktuacji \mathbf{X}_a od wartości oczekiwanej $\langle A \rangle$, tzn. dokonać addytywnego rozkładu: $\mathbf{Y}_a \equiv A = \langle A \rangle + \mathbf{X}_a$. Odpowiedni analog pomiędzy fluktuacjami energii i produktywności ma więc następującą postać: Dla energii cząstki zachodzi (5.1):

$$\mathbf{y}_\epsilon = \frac{\epsilon}{c} = \theta_\epsilon + \mathbf{x}_\epsilon , \quad \frac{\epsilon_0}{c} \leq \mathbf{y}_\epsilon \leq \infty , \quad \mathbf{x}_\epsilon^{\min} \leq \mathbf{x}_\epsilon < \infty \quad (5.69)$$

gdzie skorzystano z założenia o nieograniczoności od góry fluktuacji energii (por. (5.37) wraz z dyskusją zawartą powyżej), natomiast dla produktywności pracownika zachodzi:

$$\mathbf{y}_a \equiv a = \theta_A + \mathbf{x}_a , \quad a_0 \leq \mathbf{y}_a \leq \infty , \quad \mathbf{x}_a^{\min} = a_0 - \theta_A \leq \mathbf{x}_a < \infty . \quad (5.70)$$

Zatem analiza, która w Rozdziale 5.1.3 doprowadziła do amplitudy rangi $N = 1$ określonej przez (5.71) oraz (5.50), daje teraz amplitudę produktyjności:

$$q(\mathbf{x}_a) = \frac{1}{\sqrt{D/n}} \exp \left[-\frac{(D/n) - a_0 + \mathbf{x}_a}{2D/n} \right] \quad \text{dla} \quad \mathbf{x}_a^{min} = a_0 - \theta_A \leq \mathbf{x}_a < \infty. \quad (5.71)$$

oraz rozkład gęstości produktyjności:

$$p(a) = \begin{cases} \frac{1}{(D/n) - a_0} \exp \left(-\frac{a - a_0}{(D/n) - a_0} \right) & \text{dla} \quad a \geq a_0 \\ 0 & \text{dla} \quad a < a_0 \end{cases}, \quad (5.72)$$

gdzie w drugiej linii po prawej stronie zaznaczono fakt nie występowania produktyjności mniejszej niż a_0 . Rozkład (5.72) jest końcowym rezultatem metody EFI dla modelu AYM produktyjności. Aby jego postać oddała w pełni wynik Aoki i Yoshikawy, należy powrócić do dyskretyzacji wartości $a \rightarrow a_i$ zmiennej A .

Na końcu tego punktu zamieszczono omówienie metody Aoki i Yoshikawy. Rozwiązali oni powyższy problem stosując metodę mnożników Lagrange'a z warunkami ograniczającymi (5.57) oraz (5.60). Następnie założyli [41], że wartości produktyjności są dyskretne i tworzą ciąg arytmetyczny:

$$a_i = i a_0 \quad \text{gdzie} \quad i = 1, 2, \dots, g, \quad (5.73)$$

gdzie a_0 jest najmniejszą produktyjnością. W końcu, przy założeniach, po pierwsze, że liczba dostępnych sektorów produkcyjnych jest bardzo duża, tzn. $g \gg 1$, oraz po drugie, że:

$$r \equiv \frac{D/n}{a_0}, \quad (5.74)$$

tzn. agregatowy popyt przypadający na jednego pracownika D/n odniesiony do najmniejszej produktyjności a_0 , jest bardzo duży, Aoki i Yoshikawy otrzymali wynik [41, 42]:

$$P(i|\mathbf{n}^*) = \frac{n_i^*}{n} \approx \frac{1}{r-1} \left(\frac{r-1}{r} \right)^i \approx \left(\frac{1}{r} + \frac{1}{r^2} \right) e^{-\frac{i}{r}}, \quad i = 1, 2, \dots, \quad r \gg 1, \quad (5.75)$$

gdzie n_i^* , $i = 1, 2, \dots, g$, są współrzędnymi n_i wektora obsadzeń (5.61), przy których prawdopodobieństwo pojawienia się tego wektora obsadzeń jest maksymalne. Rezultat (5.75) podaje prawdopodobieństwo, że losowy wyprany pracownik jest w i -tym sektorze produktyjności, o ile gospodarka znajduje się w stanie określonym wektorem obsadzeń $\mathbf{n}^* = (n_1^*, n_2^*, \dots, n_g^*)$.

Aby porównać wynik (5.75) otrzymany w AYM z wynikiem (5.72) otrzymanym w metodzie EFI, przejdźmy w (5.72) do rozkładu dyskretnego. W tym celu musimy wyciąkować wynik EFI w przedziale (a_i, a_{i+1}) , przy czym od razu założymy, że zachodzi $a_i = i a_0$, (5.73). W rezultacie otrzymujemy⁵:

$$P(i) = \int_{ia_0}^{(i+1)a_0} da p(a) = \left(1 - e^{-1/(r-1)} \right) e^{-(i-1)/(r-1)} \quad \text{dla} \quad i = 1, 2, \dots. \quad (5.77)$$

⁵Przyjmując, że $r \gg 1$, otrzymujemy, w podanych granicach, następujący rozkład prawdopodobieństwa produktyjności pracownika:

$$P(i) \approx \left(\frac{1}{r} + \frac{1}{2r^2} \right) \left(e^{-\frac{i}{r}} + \frac{1}{r} \right) \quad \text{dla} \quad i = 1, 2, \dots, \quad \text{oraz} \quad a_0 > 0, \quad r \gg 1. \quad (5.76)$$

Porównanie modeli dla $a_0 = 0$. Niech δa jest stałą szerokością sektorów produkcyjności. Wtedy, w przypadku gdy $a_0 = 0$ wzór (5.51) metody EFI prowadzi w miejsce (5.77) do rozkładu:

$$P(i) = \int_{(i-1)\delta a}^{i\delta a} da p(a) = \left(-1 + e^{1/\tilde{r}}\right) e^{-i/\tilde{r}}, \quad i = 1, 2, \dots, \quad \text{dla } a_0 = 0, \quad (5.78)$$

gdzie w zamiast (5.74) wprowadziliśmy:

$$\tilde{r} \equiv \frac{D/n}{\delta a}, \quad (5.79)$$

jako agregatowy popyt przypadający na jednego pracownika D/n odniesiony do szerokości sektora produkcyjności δa . W granicy $\tilde{r} \gg 1$ z (5.78) otrzymujemy:

$$P(i) \approx \left(\frac{1}{\tilde{r}} + \frac{1}{2\tilde{r}^2}\right) e^{-i/\tilde{r}}, \quad i = 1, 2, \dots, \quad \text{dla } a_0 = 0, \quad \tilde{r} \gg 1. \quad (5.80)$$

W granicy dużych \tilde{r} oba wyniki się schodzą. Jednakże granicę niskiej produktywności metoda EFI ujmuje inaczej niż AYM, tzn. formuła EFI (5.80) daje dla $a_0 = 0$ inną kwadratową poprawkę niż rezultat (5.75) otrzymany w modelu AYM.

Uwaga: Ponadto, wynik (5.78) jest dokładny, natomiast w AYM dyskretyzacja poziomu produktywności jest tylko wybiegiem technicznym, gdyż zmienna ta jest z natury ciągła, z czego i tak korzystamy w AYM przy wyznaczaniu mnożników Lagrange'a przechodząc (z powodów rachunkowych) w (5.57) oraz (5.60) z g do nieskończoności.

Uzupełnienie: *Analiza Aoki i Yoshikawy dla produkcyjności.* Podajmy sposób wyprowadzenia rozkładu wektora obsadzeń w AYM metodą czynników Lagrangea [42]. Wprowadźmy wektor $\mathbf{H}^{(n)}$ indywidualnych przypisań, posiadający tyle składowych ilu jest pracowników w całej gospodarce kraju:

$$\mathbf{H}^{(n)} \equiv (H_1, H_2, \dots, H_n) = \mathbf{h} \equiv (h_1, h_2, \dots, h_n). \quad (5.81)$$

Każda ze współrzędnych H_i , $i = 1, 2, \dots, n$, może przyjmować wartości $h_i = s$, gdzie $s \in \{1, 2, \dots, g\}$, co oznacza, że i -ty pracownik jest aktywny w s -tym sektorze gospodarki. Zatem wektor \mathbf{h} podaje jedną konfigurację indywidualnych przypisań. Jeśli ustalimy wektorem obsadzeń \mathbf{n} , (5.61), to liczba W różnych \mathbf{h} (indywidualnych konfiguracji przypisań), realizujących ten sam ustalony wektor obsadzeń \mathbf{n} , wynosi [43]:

$$W(\mathbf{H}|\mathbf{n}) = \frac{n!}{\prod_{i=1}^g n_i!}. \quad (5.82)$$

Boltzman zauważył, że gdy układ znajduje się w równowadze statystycznej to, prawdopodobieństwo $\pi(\mathbf{n})$, że znajduje się on w stanie o określonym wektorze obsadzeń \mathbf{n} (tzn. że pojawił się taki właśnie wektor obsadzeń), jest proporcjonalne do liczby jego możliwych realizacji tzn. do $W(\mathbf{H}|\mathbf{n})$. Zatem:

$$\pi(\mathbf{n}) = W(\mathbf{H}|\mathbf{n}) P(\mathbf{H}|\mathbf{n}) = \frac{n!}{\prod_{i=1}^g n_i!} \prod_{i=1}^g p^{n_i} = \frac{n!}{\prod_{i=1}^g n_i!} p^n = \mathcal{K} \frac{n!}{\prod_{i=1}^g n_i!}, \quad (5.83)$$

gdzie \mathcal{K} jest właściwą normalizacyjną stałą, a p jest prawdopodobieństwem zajęcia przez i -tego pracownika określonego s -tego sektora obsadzeń, które zostało przyjęte jako takie samo dla wszystkich indywidualnych konfiguracji tych obsadzeń. W celu rozwiązania postawionego problem maksymalizacji prawdopodobieństwa $\pi(\mathbf{n})$ z warunkami (5.57) oraz (5.60), rozwiązujemy poniższy układ g równań:

$$\frac{\partial}{\partial n_i} \left[\ln \pi(\mathbf{n}) + \nu \left(\sum_{i=1}^g n_i - n \right) - \beta \left(\sum_{i=1}^g a_i n_i - D \right) \right] = 0, \quad (5.84)$$

otrzymując jako rozwiązanie wektor obsadzeń \mathbf{n} :

$$n_i = n_i^* = e^\nu e^{-\beta a_i}, \quad i = 1, 2, \dots, g. \quad (5.85)$$

Stałe ν oraz β otrzymujemy wykorzystując (5.85) w (5.57) oraz (5.60).

Uwaga: Aby zapisać $\prod_{i=1}^g n_i!$ w formie nadającej się do minimalizacji, korzystamy z przybliżenia Stirling'a:

$$\ln \left[\prod_{i=1}^g n_i! \right] \approx \sum_{i=1}^g n_i (\ln n_i - 1), \quad (5.86)$$

słusznego dla dużego układu z $n \gg 1$ oraz $n_i \gg 1$, $i = 1, 2, \dots, g$.

Przedstawiona w tym przypisie metoda znajdowania rozkładu w równowadze statystycznej zakłada spełnienie hipotezy wyrażonej równaniem (5.83) i mówiącej, że wszystkie stany opisane wektorem indywidualnych przypisań $\mathbf{H}^{(n)}$ a spełniające warunki (5.57) oraz (5.60) są równie prawdopodobne. Rozkład opisany wektorem \mathbf{n}^* jest w tym ujęciu sednem definicji rozkładu będącego w równowadze statystycznej.

5.1.5 Rozkład Maxwella-Boltzmana dla prędkości

Poniżej wyznaczmy rozkładu prędkości cząsteczki w gazie. Wychodząc z informacji Fishera (5.7) dla pędu oraz posługując się wariacyjną i strukturalną zasadą wariacyjną (5.9) otrzymamy równanie generujące i znajdziemy jego rozwiązania, tzn. postać amplitud dla rozkładu prędkości. Ze względu spójność rozważań dla czterowektora pędu, przyjęcie współczynnika $\kappa = 1$ w rozważaniach dla energii skutkuje przyjęciem $\kappa = 1$ w analizie dla rozkładu pędu.

Strukturalna zasada informacyjna ma postać:

$$K = I(\Theta_{\vec{\rho}}) + Q(\Theta_{\vec{\rho}}) = 4 \int d\vec{x}_\varphi \sum_{n=1}^N \left[-\sum_{k=1}^3 \left(\frac{\partial q_n(\vec{x}_\varphi)}{\partial x_{\varphi k}} \right)^2 + \frac{1}{4} q_n^2(\vec{x}_\varphi) \mathfrak{F}_n(q_n(\vec{x}_\varphi), \vec{x}_\varphi) \right] = 0 \quad (5.87)$$

gdzie znaczenie znaku “minus” w definicji pojemności informacyjnej dla części pędowej było omówione w Rozdziale 5.1.2. Wprowadzenie do obserwowanej informacji strukturalnej $\mathfrak{F}_n(q_n(\vec{x}_\varphi), \vec{x}_\varphi)$ jej jawnej zależności nie tylko od amplitud q_n , ale również od pędu \vec{x}_φ pozwala na rozważenie szerszego zakresu zagadnień niż to miało miejsce dla przypadku rozkładu energii, mianowicie pozwoli na rozważenie rozwiązań nierównowagowych.

Wariacyjna zasada informacyjna ma postać:

$$\delta_{(q_n)}(I(\Theta_{\vec{\rho}}) + Q(\Theta_{\vec{\rho}})) = 4 \int d\vec{x}_\varphi \sum_{n=1}^N \left[-\sum_{k=1}^3 \left(\frac{\partial q_n(\vec{x}_\varphi)}{\partial x_{\varphi k}} \right)^2 + \frac{1}{4} q_n^2(\vec{x}_\varphi) \mathfrak{F}_n(q_n(\vec{x}_\varphi), \vec{x}_\varphi) \right] = 0 \quad (5.88)$$

Podobnie jak to uczyniliśmy w (5.18) dla energii, tak i teraz dla pojemności informacyjnej (5.7) dokonamy całkowania przez części. Zakładając dodatkowo, że amplitudy dla prędkości $q_n(\vec{x}_\varphi)$ znikają w \pm nieskończoności, zapisujemy strukturalną zasadę informacyjną (5.87) w postaci:

$$I(\Theta_{\vec{\rho}}) + Q(\Theta_{\vec{\rho}}) = 4 \int d\vec{x}_\varphi \sum_{n=1}^N \left[q_n(\vec{x}_\varphi) \sum_{k=1}^3 \frac{\partial^2 q_n(\vec{x}_\varphi)}{\partial x_{\varphi k}^2} + \frac{1}{4} q_n^2(\vec{x}_\varphi) \mathfrak{F}_n(q_n(\vec{x}_\varphi), \vec{x}_\varphi) \right] = 0 \quad (5.89)$$

Rozwiązaniem N -funkcyjnego problemu wariacyjnego (5.88) jest układ równań Eulera-Lagrange’a:

$$\sum_{m=1}^3 \frac{\partial}{\partial x_{\varphi m}} \left(\frac{\partial k(q_n(\vec{x}_\varphi), \vec{x}_\varphi)}{\partial \left(\frac{\partial q_n(\vec{x}_\varphi)}{\partial x_{\varphi m}} \right)} \right) = \frac{\partial k(q_n(\vec{x}_\varphi), \vec{x}_\varphi)}{\partial q_n(\vec{x}_\varphi)}, \quad n = 1, 2, \dots, N, \quad (5.90)$$

gdzie

$$k(q_n(\vec{x}_\varphi), \vec{x}_\varphi) = 4 \left[-\sum_{k=1}^3 \left(\frac{\partial q_n(\vec{x}_\varphi)}{\partial x_{\varphi k}} \right)^2 + \frac{1}{4} q_n^2(\vec{x}_\varphi) \mathfrak{F}_n(q_n(\vec{x}_\varphi), \vec{x}_\varphi) \right], \quad (5.91)$$

jest gęstością informacji fizycznej (4.7) zdefiniowaną w Rozdziale 4.2. Zauważmy, że z (5.91) i z żądania nieujemności informacji fizycznej k na poziomie mikroskopowym, wynika:

$$\mathfrak{F}_n(q_n(\vec{x}_\varphi), \vec{x}_\varphi) \geq 0, \quad (5.92)$$

tzn. *nieujemność mikroskopowej informacji strukturalnej w analizie estymacyjnej wartości oczekiwanej pędu cząsteczki gazu.*

Poniżej, ze względu na uproszczenie zapisu, pominiemy zaznaczenie fluktuacji pędu \vec{x}_φ w argumencie amplitudy $q_n(\vec{x}_\varphi)$.

Zatem każde z N równań Eulera-Lagrange'a (5.90) ma postać następującego równania różniczkowego:

$$\sum_{m=1}^3 \frac{\partial^2 q_n}{\partial x_{\varphi_m}^2} = -\frac{1}{2} \frac{\partial(\frac{1}{4}q_n^2 \mathfrak{F}_n(q_n, \vec{x}_\varphi))}{\partial q_n} \quad (5.93)$$

Natomiast zasada strukturalna (5.89) na poziomie mikroskopowym ma postać:

$$q_n \sum_{m=1}^3 \frac{\partial^2 q_n}{\partial x_{\varphi_m}^2} + \frac{1}{4}q_n^2 \mathfrak{F}_n(q_n, \vec{x}_\varphi) = 0. \quad (5.94)$$

Równania (5.93) oraz (5.94) pozwalają wyeliminować sumę po ich lewej stronie, dając równanie:

$$\frac{1}{2} \frac{\partial(\frac{1}{4}q_n^2 \mathfrak{F}_n(q_n, \vec{x}_\varphi))}{\partial q_n} = \frac{\frac{1}{4}q_n^2 \mathfrak{F}_n(q_n, \vec{x}_\varphi)}{q_n}, \quad (5.95)$$

które po obustronnym scałkowaniu prowadzi do rozwiązania:

$$\frac{1}{4}q_n^2 \mathfrak{F}_n(q_n, \vec{x}_\varphi) = q_n^2 f_n(\vec{x}_\varphi), \quad (5.96)$$

lub

$$\mathfrak{F}_n(q_n, \vec{x}_\varphi) = 4f_n(\vec{x}_\varphi) \geq 0. \quad (5.97)$$

Funkcja $f_n(\vec{x}_\varphi)$ nie zależy od amplitudy q_n i pojawiła się w wyniku całkowania równania (5.95) jako pewna stała całkowania w znaczeniu jej niezależności od amplitudy q_n . Zaznaczona nieujemność funkcji $f_n(\vec{x}_\varphi)$ w jej dziedzinie wynika z nieujemności $\mathfrak{F}_n(q_n, \vec{x}_\varphi)$ otrzymanej w (5.92). Wykorzystując (5.96) w (5.94) eliminujemy obserwowaną informację strukturalną, otrzymując *równanie generujące* dla amplitudy rozkładu (fluktuacji) prędkości \vec{x}_φ cząsteczki w gazie:

$$\nabla^2 q_n(\vec{x}_\varphi) = -q_n(\vec{x}_\varphi) f_n(\vec{x}_\varphi), \quad (5.98)$$

gdzie $\nabla^2 \equiv \sum_{m=1}^3 \partial^2 / \partial x_{\varphi_m}^2$ jest operatorem Laplace'a ze względu na współrzędne pędowe x_{φ_m} . Równanie (5.98) podobnie jak (5.32) pojawiło się jako rozwiązanie strukturalnej i wariacyjnej zasady informacyjnej. Rozwiązanie q_n równania (5.98) jest rozwiązaniem sprzężonego układu równań różniczkowych (5.94) i (5.93) utworzonych przez parę zasad informacyjnych i jako takie jest tzw. samospójnym.

Poniżej rozwiążemy równanie generujące (5.98), czyniąc kilka fizycznych założeń odnośnie postaci funkcji $f_n(\vec{x}_\varphi)$. Jej postać wpływa na postać otrzymanych amplitud $q_n(\vec{x}_\varphi)$, a zatem również na rozkład $p(\vec{x}_\varphi)$.

Założmy, że każda gęstość rozkładu prawdopodobieństwa $p_n(\vec{x})$ jest taką samą funkcją parzystą od każdej współrzędnej x_{φ_i} (fluktuacji) pędu \vec{x}_φ . W konsekwencji układ jest izotropowy, tzn. rozkład prawdopodobieństwa dla pędu nie zależy od kierunku w bazowej przestrzeni położeń. Następnie zakładamy nierelatywistyczne przybliżenie, co oznacza, że prędkości cząsteczek są dużo mniejsze od prędkości światła c , czyli (fluktuacja) pędu cząsteczki x_{φ_i} jest również mała w porównaniu z mc . Z powyższych założeń wynika ogólna postać funkcji $f_n(\vec{x}_\varphi)$. Otóż jej rozwinięcie w szereg potęgowy ma tylko składowe parzyste modułu (fluktuacji) pędu $|\vec{x}_\varphi|$, a ponieważ wartość $|\vec{x}_\varphi|$ jest mała, zatem szereg ten obetniemy na drugim wyrazie:

$$f_n(\vec{x}_\varphi) = A_n + B |\vec{x}_\varphi|^2, \quad A_n, B = const. \quad (5.99)$$

Podstawiając (5.99) do (5.98) otrzymujemy następującą postać równania generującego:

$$\nabla^2 q_n(\vec{x}_\varphi) + (A_n + B|\vec{x}_\varphi|^2) q_n(\vec{x}_\varphi) = 0 \quad \text{dla } n = 1, 2, \dots, N. \quad (5.100)$$

Przejdźmy do nowego indeksu:

$$n' := n - 1 = 0, 1, \dots, N - 1. \quad (5.101)$$

Wtedy w miejsce (5.100) otrzymujemy równanie:

$$\nabla^2 q_{n'}(\vec{x}_\varphi) + (A_{n'} + B|\vec{x}_\varphi|^2) q_{n'}(\vec{x}_\varphi) = 0 \quad \text{dla } n' = 1, 2, \dots, N - 1, \quad (5.102)$$

które po dokonaniu separacji zmiennych kartezjańskich i faktoryzacji amplitudy:

$$q_{n'}(\vec{x}_\varphi) = q_{n'_1}(x_{\varphi_1}) q_{n'_2}(x_{\varphi_2}) q_{n'_3}(x_{\varphi_3}), \quad (5.103)$$

przechodzi w równoważny mu układ trzech równań różniczkowych:

$$q_{n'_i}''(x_{\varphi_i}) + (A_{n'_i} + B x_{\varphi_i}^2) q_{n'_i}(x_{\varphi_i}) = 0, \quad i = 1, 2, 3, \quad \sum_{i=1}^3 A_{n'_i} \equiv A_{n'}. \quad (5.104)$$

Gdy stałe równania (5.104) mają postać:

$$A_{n'_i} = \frac{n'_i + 1/2}{a_0^2}, \quad B = -\frac{1}{4a_0^4}, \quad a_0 = \text{const.}, \quad n'_i = 0, 1, \dots, \quad (5.105)$$

wtedy ma ono rozwiązanie. Ponieważ jest to równanie Helmholtz'a, zatem jego rozwiązaniami są paraboliczno-cylindryczne funkcje [44, 45]:

$$q_{n'_i}(x_{\varphi_i}) = e^{-x_{\varphi_i}^2/(4a_0^2)} 2^{-n'_i/2} H_{n'_i} \left(\frac{x_{\varphi_i}}{a_0\sqrt{2}} \right), \quad \text{gdzie } n'_i = 0, 1, \dots, \quad i = 1, 2, 3. \quad (5.106)$$

gdzie $H_{n'_i}$ są wielomianami Hermite'a:

$$H_j(t) = j! \sum_{m=0}^{[j/2]} (-1)^m \frac{(2t)^{j-2m}}{m!(j-2m)!}, \quad j = 0, 1, \dots \quad (5.107)$$

gdzie $[j/2]$ oznacza liczbę całkowitą nie przekraczającą $j/2$. Wstawiając (5.106) do (5.103) otrzymujemy szukaną postać amplitudy (fluktuacji) pędu:

$$q_{n'}(\vec{x}_\varphi) = e^{-|\vec{x}_\varphi|^2/(4a_0^2)} 2^{-n'/2} \cdot \sum_{\substack{ijk \\ i+j+k=n'}} a_{n'_{ijk}} H_i \left(\frac{x_{\varphi_1}}{a_0\sqrt{2}} \right) H_j \left(\frac{x_{\varphi_2}}{a_0\sqrt{2}} \right) H_k \left(\frac{x_{\varphi_3}}{a_0\sqrt{2}} \right), \quad \text{gdzie } a_{n'_{ijk}} = \text{const.}, \quad (5.108)$$

przy czym:

$$\sum_{i=1}^3 n'_i = n' = 0, 1, \dots, N - 1. \quad (5.109)$$

Funkcja rozkładu prawdopodobieństwa ma więc zgodnie z (3.50) postać:

$$p(\vec{x}_\varphi) = \frac{1}{N} \sum_{n'=0}^{N-1} q_{n'}^2(\vec{x}_\varphi) = p_0 e^{-|\vec{x}_\varphi|^2/(2a_0^2)} \cdot \left\{ 1 + \sum_{n'=1}^{N-1} 2^{-n'} \left[\sum_{\substack{ijk \\ i+j+k=n'}} b_{n'ijk} H_i\left(\frac{x_{\varphi_1}}{a_0\sqrt{2}}\right) H_j\left(\frac{x_{\varphi_2}}{a_0\sqrt{2}}\right) H_k\left(\frac{x_{\varphi_3}}{a_0\sqrt{2}}\right) \right]^2 \right\}, \quad (5.110)$$

gdzie $p_0 = a_{0000}^2/N$, a liczba 1 w nawiasie klamrowym pochodzi z $n' = 0$ w sumie w pierwszej równości, natomiast nowe stałe $b_{n'ijk}$ są proporcjonalne do stałych $a_{n'ijk}$. Po raz pierwszy takie wyprowadzenie postaci funkcji rozkładu gęstości prawdopodobieństwa dla (fluktuacji) pędu zostało podane w [7].

Wartość $N = 1$, dla której (5.110) jest rozkładem Gaussa, daje równowagowy *rozkład Maxwella-Boltzmannna dla (fluktuacji) pędu cząsteczki gazu*. Pozostałe rozwiązania dla $N \geq 2$ dają rozwiązania nierównowagowe [7]. Jednak i one są stacjonarne ze względu na fakt bycia samospójnymi rozwiązaniami układu sprzężonych równań różniczkowych zasady strukturalnej (5.94) i wariacyjnej (5.93). Zwróćmy uwagę, że zawierają one interferencję pomiędzy wyrażeniami iloczynowymi występującymi w amplitudach (5.108). Interferencja pojawiła się więc jako cecha charakterystyczna dla rozwiązywanego równania różniczkowego oraz wprowadzenia amplitud do opisu układu (a nie jako cecha charakterystyczna wyłącznie mechaniki kwantowej). Wcześniej, rozkłady (5.110) z $N \geq 2$ zostały odkryte przez Rumer'a i Ryskin'a [46] jako rozwiązania równania transportu Boltzmann'a.

Uogólnienie fizyczne dla $N = 1$: Założyliśmy na wstępie, że średnia prędkość cząsteczki wynosi zero (wystarczy przyjąć, że zbiornik zawierający cząsteczki gazu jest w spoczynku w pewnym inercjalnym układzie współrzędnych). Zatem $\vec{\theta}_\varphi = 0$ i z (5.2) mamy, $\vec{y}_\varphi = \vec{x}_\varphi = \vec{\varphi}$. Zatem dla $N = 1$ rozkład gęstości prawdopodobieństwa pędu (5.110) przyjmuje postać:

$$p(\vec{\varphi}) = p_0 e^{-|\vec{\varphi}|^2/(2a_0^2)} = p_0 e^{-\varphi_1^2/(2a_0^2)} e^{-\varphi_2^2/(2a_0^2)} e^{-\varphi_3^2/(2a_0^2)}, \quad (5.111)$$

gdzie $\vec{\varphi} = (\varphi_1, \varphi_2, \varphi_3)$ oraz $-\infty < \varphi_i < \infty$, $i = 1, 2, 3$. Wyznamy wartość stałej a_0 dla przypadku zerowej energii potencjalnej. Z warunku normalizacji $\int d\vec{\varphi} p(\vec{\varphi}) = 1$, wyznaczamy, że stała p_0 w (5.111) wynosi:

$$p_0 = \frac{1}{(2\pi)^{3/2} a_0^3} \quad (5.112)$$

Ponieważ wtedy $\langle E \rangle = \langle \vec{\varphi}^2 \rangle / (2m)$ a z twierdzenia o ekwipartycji energii wiemy, że $\langle E \rangle = 3kT/2$, zatem

$$\langle \vec{\varphi}^2 \rangle = 3m kT. \quad (5.113)$$

Mając wartość p_0 , wyznaczamy wartość oczekiwaną:

$$\langle \vec{\varphi}^2 \rangle = \int d\vec{\varphi} p(\vec{\varphi}) \vec{\varphi}^2, \quad (5.114)$$

i przyrównajmy ją do (5.113). W rezultacie otrzymujemy:

$$a_0^2 = mc^2 kT . \quad (5.115)$$

Kolejnym krokiem jest wyznaczenie rozkładu prawdopodobieństwa wartości pędu $\wp = |\vec{\wp}|$. Przechodząc od współrzędnych kartezjańskich pędu (\wp_1, \wp_2, \wp_3) do współrzędnych sferycznych (\wp, θ, ϕ) , otrzymujemy:

$$p(\wp, \theta, \phi) = |J| p(\wp_1, \wp_2, \wp_3) , \quad J = \wp^2 \sin \theta , \quad (5.116)$$

gdzie J jest jakobianem przejścia. Wstawiając teraz (5.112) wraz z (5.115) do (5.111) otrzymujemy rozkład $p(\wp_1, \wp_2, \wp_3)$ ze znanymi już stałymi a_0 oraz p_0 . Wynik ten wstawiając do (5.116) i wyciąkując po zmiennych θ oraz ϕ otrzymujemy szukany rozkład gęstości prawdopodobieństwa dla wartości pędu:

$$p(\wp) = \sqrt{2/\pi} (mkT)^{-3/2} \wp^2 e^{-\wp^2/2mkT} . \quad (5.117)$$

Rozkład (5.117) wyraża tzw. prawo Maxwella-Boltzmann, a podstawiając $\wp = mv$ w (5.117) w miejsce wartości pędu, otrzymujemy rozkład Maxwella-Boltzmann dla wartości prędkości v .

5.1.6 Informacja Fishera jako ograniczenie dla wzrostu entropii

Rozważmy układ zawierający *jedną* lub *więcej* cząstek poruszających się w sposób losowy w pewnym zamkniętym obszarze. Niech rozważany obszar będzie izolowany, tzn. żadna cząstka ani z obszaru nie ucieka ani do niego nie przechodzi. Brzeg (\vec{B}) obszaru jest zadany wektorem wodzącym $\vec{b} \in \vec{B}$. W pewnej chwili układ dokonuje pomiaru jego czero-wektora położenia t, \vec{y} , zgodnie z rozkładem gęstości prawdopodobieństwa $p(\vec{y}, t)$ ⁶. Niech (y^1, y^2, y^3) są współrzędnymi kartezjańskimi wektora \vec{y} położenia przestrzennego, a $r = |\vec{y}|$ jego długością. Ponieważ założyliśmy, że cząstka znajduje się gdzieś wewnątrz obszaru więc:

$$p(t) = \int d\vec{y} p(\vec{y}, t) = 1. \quad (5.118)$$

Oznaczmy przez $S(t)$ entropię Shannona układu (2.60) w chwili czasu t :

$$S(t) = - \int d\vec{y} p(\vec{y}, t) \ln p(\vec{y}, t). \quad (5.119)$$

Można pokazać (por. Dodatek), że entropia Shannona spełnia drugą zasadę termodynamiki:

$$\frac{dS(t)}{dt} \geq 0, \quad (5.120)$$

która określa dolną granicę tempa zmiany entropii. Poniższe rozważania poświęcone są znalezieniu ograniczenia na jego granicę górną.

Ponieważ żadna cząstka nie opuszcza ani nie wpływa do obszaru, zatem spełnione jest równanie ciągłości strumienia prawdopodobieństwa:

$$\frac{\partial p(\vec{y}, t)}{\partial t} + \vec{\nabla} \cdot \vec{P}(\vec{y}, t) = 0, \quad (5.121)$$

gdzie $\vec{P}(\vec{y}, t)$ jest prądem prawdopodobieństwa, którego konkretna postać zależy od rozważanego układu.

Z kolei określmy warunki brzegowe spełniane przez gęstość prawdopodobieństwa i jego prąd. Będą nam one przydatne dla rozwiązania postawionego sobie zadania.

Ponieważ żadne cząstki nie przechodzą przez granicę obszaru, zatem \vec{P} spełnia warunek brzegowy Dirichleta:

$$\vec{P}(\vec{y}, t) \Big|_{\vec{y} \in \vec{B}} = 0. \quad (5.122)$$

Założmy dodatkowo, że jeśli brzeg obszaru znajduje się w nieskończeniu, to:

$$\lim_{\vec{y} \rightarrow \infty} \vec{P}(\vec{y}, t) \rightarrow 0, \quad \text{szybciej niż } 1/r^2. \quad (5.123)$$

Ponieważ obszar jest odizolowany zatem prawdopodobieństwo, że cząstka znajduje się na jego granicy zanika, co oznacza, że $p(\vec{y}, t)$ również spełnia warunek brzegowy Dirichleta:

$$p(\vec{y}, t) \Big|_{\vec{y} \in \vec{B}} = 0. \quad (5.124)$$

⁶Jeśli układ opisany jest funkcją falową $\psi(\vec{y}, t)$, wtedy $p(\vec{y}, t) = |\psi(\vec{y}, t)|^2$

Ze względu na normalizację:

$$\int d\vec{y} p(\vec{y}, t) = 1, \quad (5.125)$$

zakładamy, że w przypadku brzegu nieskończenie oddalonego, $p(\vec{y}, t)$ spełnia warunek:

$$\lim_{\vec{y} \rightarrow \infty} p(\vec{y} | t) \rightarrow 0, \quad \text{szybciej niż } 1/r^3. \quad (5.126)$$

W końcu dla domknięcia potrzebnych warunków brzegowych określamy:

$$\vec{P} \ln p \Big|_{\vec{y} \in \vec{B}} = 0, \quad (5.127)$$

Przystąpmy teraz do sedna rachunków. Różniczkowanie (5.119) po $\frac{\partial}{\partial t}$ daje:

$$\frac{\partial S}{\partial t} = -\frac{\partial}{\partial t} \int d\vec{y} p \ln p = -\int d\vec{y} \frac{\partial p}{\partial t} \ln p - \int d\vec{y} p \frac{1}{p} \frac{\partial p}{\partial t}. \quad (5.128)$$

Druga całka po prawej stronie daje po skorzystaniu z warunku unormowania:

$$\frac{\partial}{\partial t} \int d\vec{y} p = 0. \quad (5.129)$$

Oznaczmy przez (P^1, P^2, P^3) kartezyjańskie składowe prądu \vec{P} . Podstawiając (5.121) do pierwszej całki po prawej stronie w (5.128) otrzymujemy:

$$\frac{\partial S}{\partial t} = \int d\vec{y} \vec{\nabla} \cdot \vec{P} \ln p = \int \int \int dy^3 dy^2 dy^1 \left[\frac{\partial}{\partial y^1} P^1 + \frac{\partial}{\partial y^2} P^2 + \frac{\partial}{\partial y^3} P^3 \right] \ln p. \quad (5.130)$$

Całkując przez części wewnętrzne całki $\int dy^i \frac{\partial}{\partial y^i} P^i$ w (5.130) dla trzech składników $i = 1, 2, 3$ otrzymujemy dla każdego z nich:

$$\int dy^i \left(\frac{\partial}{\partial y^i} P^i \right) \ln p = P^i \ln p \Big|_{\vec{y} \in \vec{B}} - \int dy^i \frac{P^i}{p} \frac{\partial p}{\partial y^i} = - \int dy^i P^i \frac{\partial p}{\partial y^i} \frac{1}{p}, \quad i = 1, 2, 3, \quad (5.131)$$

gdzie skorzystano z (5.127). Zatem:

$$\frac{\partial S}{\partial t} = - \int dy^3 dy^2 dy^1 \left(P^1 \frac{\partial p}{\partial y^1} + P^2 \frac{\partial p}{\partial y^2} + P^3 \frac{\partial p}{\partial y^3} \right) \frac{1}{p} = - \int d\vec{y} \left(\vec{P} \cdot \vec{\nabla} p \right) \frac{1}{p}, \quad (5.132)$$

skąd po prostych przekształceniach:

$$\left(\frac{\partial S}{\partial t} \right)^2 = \left[\int d\vec{y} \left(\frac{\vec{P}}{\sqrt{p}} \right) \cdot \left(\frac{\sqrt{p} \vec{\nabla} p}{p} \right) \right]^2 = \left[\sum_{i=1}^3 \int d\vec{y} \left(\frac{P^i}{\sqrt{p}} \right) \left(\frac{\sqrt{p} \nabla_i p}{p} \right) \right]^2, \quad (5.133)$$

gdzie $\nabla_i = \frac{\partial}{\partial y^i}$. Z nierówności Schwartza otrzymujemy:

$$\left[\sum_{i=1}^3 \int d\vec{y} \left(\frac{P^i}{\sqrt{p}} \right) \left(\frac{\sqrt{p} \nabla_i p}{p} \right) \right]^2 \leq \left[\sum_{i=1}^3 \int d\vec{y} \left(\frac{P^i}{\sqrt{p}} \right)^2 \right] \left[\sum_{i=1}^3 \int d\vec{y} \left(\frac{\sqrt{p} \nabla_i p}{p} \right)^2 \right]. \quad (5.134)$$

Ostatecznie z (5.133) i (5.134) oraz po zastąpieniu sumowań po i znakiem iloczynów skalarnych, dostajemy:

$$\left(\frac{\partial S}{\partial t} \right)^2 \leq \int d\vec{y} \frac{\vec{P} \cdot \vec{P}}{p} \int d\vec{y} \frac{\vec{\nabla} p \cdot \vec{\nabla} p}{p}. \quad (5.135)$$

Ponieważ gradient $\vec{\nabla} \equiv (\frac{\partial}{\partial y^1}, \frac{\partial}{\partial y^2}, \frac{\partial}{\partial y^3})$ zawiera różniczkowanie po wartościach pomiarowych y^i a nie fluktuacjach x_i , jak to jest w kinematycznej postaci pojemności informacyjnej Friedena-Soffera, zatem musimy uczynić dodatkowe założenie.

Uwaga o założeniu niezmienniczości rozkładu ze względu na przesunięcie: Przejdźmy do addytywnych przesunięć, $\vec{x} = \vec{y} - \vec{\theta}$. Ponieważ $\vec{x} = (x_i)_{i=1}^3$ oraz $\vec{y} = (y_i)_{i=1}^3$ różnią się o stałą wartość oczekiwaną położenia $\vec{\theta} = (\theta_i)_{i=1}^3$, zatem zakładając dodatkowo, że rozkład p jest niezmienniczy ze względu na przesunięcie $\vec{\theta}$, otrzymujemy:

$$\frac{\partial p}{\partial y^i} = \frac{\partial p}{\partial x_i}, \quad i = 1, 2, 3. \quad (5.136)$$

Przy powyższym założeniu, zachodzi:

$$\int d\vec{y} \frac{\vec{\nabla} p(\vec{y}, t) \cdot \vec{\nabla} p(\vec{y}, t)}{p(\vec{y}, t)} = \int d\vec{x} \frac{\vec{\nabla} p(\vec{x}, t) \cdot \vec{\nabla} p(\vec{x}, t)}{p(\vec{x}, t)}, \quad (5.137)$$

i okazuje się, że druga całka po prawej stronie nierówności w (5.135) jest pojemnością informacyjną trzech kanałów przestrzennych dla $N = 1$ w chwili czasu t :

$$I \equiv I(t) = \int d\vec{r} \frac{\nabla p(\vec{r}|t) \cdot \nabla p(\vec{r}|t)}{p(\vec{r}|t)} \quad (5.138)$$

Możemy więc zapisać (5.135) następująco:

$$\left(\frac{\partial S}{\partial t} \right)^2 \leq I(t) \int d\vec{r} \frac{\vec{P} \cdot \vec{P}}{p}, \quad \text{lub} \quad \left(\frac{\partial S}{\partial t} \right) \leq \sqrt{I(t)} \sqrt{\int d\vec{r} \frac{\vec{P} \cdot \vec{P}}{p}}, \quad (5.139)$$

Pokazaliśmy zatem, że (przy założeniu niezmienniczości rozkładu za względu na przesunięcie) tempo wzrostu entropii jest obustronnie ograniczone. Ograniczenie dolne (5.120) wyraża zasadę niemalenia entropii Shannona w czasie. Jej termodynamicznym odpowiednikiem jest twierdzenie H Boltzmannna.

Natomiast twierdzenie (5.139) stwierdzające, że ograniczenie górne *tempa wzrostu entropii jest proporcjonalne do pierwiastka z pojemności informacyjnej* (5.138) dla pomiaru położenia \vec{y} , jest jednym z nowych wyników teorii pomiaru otrzymanym przez Friedena, Soffera, Plastino i Plastino [7]. Jego termodynamiczne konsekwencje czekają na weryfikację. W [7] podano przykłady zastosowania tego twierdzenia dla strumienia cząstek klasycznych, strumienia w elektrodynamice klasycznej i strumienia cząstek ze spinem 1/2. Poniżej zostanie podany wynik analizy dla cząstki ze spinem 1/2.

5.1.6.1 Wynik dla strumienia cząstek ze spinem 1/2

W Rozdziale 4.3.2.2 pokazaliśmy, że metoda EFI daje dla relatywistycznej cząstki o spinie półkowym równanie ruchu Diraca (4.45). Pod nieobecność pola elektromagnetycznego równanie ciągłości z niego wynikające ma postać:

$$\frac{\partial}{\partial t} p(\vec{y}, t) + \vec{\nabla} \cdot \vec{P}(\vec{y}, t) = 0, \quad (5.140)$$

gdzie gęstość prawdopodobieństwa p oraz gęstość prądu prawdopodobieństwa są równe:

$$p(\vec{y}, t) = \psi^\dagger \psi, \quad \psi \equiv \psi(\vec{y}, t), \quad (5.141)$$

oraz

$$\vec{P}(\vec{y}, t) = c\psi^\dagger \vec{\alpha} \psi, \quad (5.142)$$

gdzie $\vec{\alpha} \equiv (\alpha^1, \alpha^2, \alpha^3)$ są macierzami Diraca (4.42), natomiast $\psi^\dagger = (\psi_1, \psi_2, \psi_3, \psi_4)^*$ jest polem sprzężonym hermitowsko do bispinora Diraca ψ (Rozdział 4.3.2.2).

W [7] pokazano, że tempo wzrostu entropii (5.139) ma w tym przypadku postać:

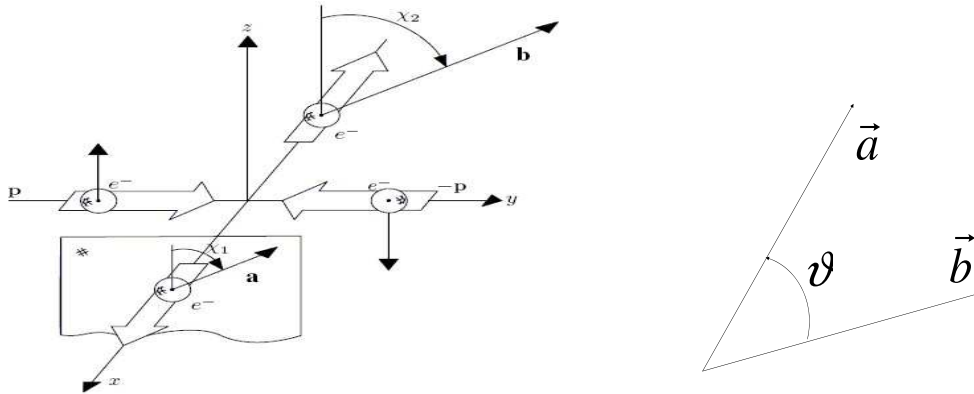
$$\frac{\partial S}{\partial t} \leq c\sqrt{I(t)}. \quad (5.143)$$

Nierówność ta oznacza, że wzrost entropii Shannona rozkładu układu w jednostce czasu (czyli tempo spadku informacji) jest ograniczony przez skończoną prędkość światła jak również przez pojemność informacyjną jaką posiada układ (np. elektron) o swoich współrzędnych czasoprzestrzennych. Inny sposób interpretacji (5.143) polega na zauważeniu, że dostarcza ona definicji prędkości światła c jako górnego ograniczenia stosunku tempa zmiany entropii do pierwiastka informacji Fishera, która jest przecież na wskroś statystycznym pojęciem informacyjnym.

5.2 Zastosowanie wprowadzonego formalizmu do analizy paradoksu EPR

Przedstawiona w tym rozdziale analiza paradoksu EPR pochodzi od Friedena [7]. Zamieszczamy ją ze względu na to, że efekt ten jest dość powszechnie utożsamiany z własnością teorii kwantowych, ale również z powodu uporządkowania warunków brzegowych zagadnienia zawartego w pracy oryginalnej [7] a dokonanego w pracy [16] oraz wnioskowania uzgodnionego z przedstawioną w skrypcie fizyczną interpretacją informacji fizycznej K [9, 10].

Rozpocznijmy od omówienia eksperymentu EPR-Bohm'a. Rozważmy źródło molekuł o spinie zero, które rozpadają się na parę identycznych cząstek o spinie $1/2$ lecących w przeciwnych kierunkach. Taka początkowa dla rozważań eksperymentu EPR-Bohm'a konfiguracja dwucząstkowej molekuly może być efektywnie przygotowana jako stan końcowy w rozpraszaniu $e^-e^- \rightarrow e^-e^-$, gdzie spiny początkowych elektronów procesu są ustawione przeciwnie (równoległe i antyrównoległe względem osi z), a ich początkowe pędy wzdłuż osi y wynoszą \vec{p} oraz $-\vec{p}$ [47]. Istnieje niezerowa amplituda, że dwie rozproszone cząstki (tu końcowe elektrony), poruszają się z pędami wzdłuż osi x jak na Rysunku 5.1. W eksperymencie EPR-Bohm'a dokonywany jest pomiar spinu



Rysunek 5.1: Panel lewy: Eksperyment EPR-Bohma [47]. Panel prawy: Kąt między urządzeniami Sterna-Gerlacha umieszczonymi wzdłuż kierunków \vec{a} oraz \vec{b} wynosi ϑ .

rozproszonych cząstek, spinu cząstki 1 wzdłuż wektora jednostkowego \vec{a} tworzącego z osią z kąt χ_1 oraz spinu cząstki 2 wzdłuż wektora jednostkowego \vec{b} tworzącego z osią z kąt χ_2 . Ponieważ rozważamy cząstkę o spinie $1/2$, w wyniku pomiaru możemy uzyskać dwie możliwości rzutu spinu na wybrany kierunek: zgodny lub przeciwny z tym kierunkiem.

W szczególności niech analizator „a” będący urządzeniem typu Sterna-Gerlacha mierzy rzut $S_a = \vec{a} \cdot \vec{S}_1$ spinu cząstki 1 na kierunek \vec{a} i podobnie analizator „b” mierzy rzut $S_b = \vec{b} \cdot \vec{S}_2$ cząstki 2 na kierunek \vec{b} . Płaszczyzny \vec{a} oraz \vec{b} są względem siebie odwrócone o kąt $\vartheta = \chi_1 - \chi_2$, $0 \leq \vartheta \leq 2\pi$.

Poniżej wyprowadzimy warunki brzegowe dla metody EFI, uważając aby nie odwoływać się do widzenia rzeczywistości przez pryzmat mechaniki kwantowej ⁷.

⁷Mając na względzie ogólną uwagę: Niech P_{12} jest łącznym rozkładem prawdopodobieństwa pewnych dwóch zmiennych losowych „1” oraz „2”, a P_1 oraz P_2 ich rozkładami brzegowymi. Jeśli w ogólności nie są one względem

5.2.1 Warunki brzegowe

Założenie o istnieniu dokładnie dwóch możliwych rzutów spinu cząstki ze spinem $\hbar/2$ na dowolny wybrany kierunek w przestrzeni: Niech “+” oznacza obserwowaną wartość rzutu spinu $S_a = +\hbar/2$ natomiast “-” oznacza $S_a = -\hbar/2$.

Uwaga: Jest to jedyne miejsce, w którym uciekamy się do opisu fenomenologicznego odwodując się do nieklasycznej fizyki zjawiska, mianowicie zakładamy, że rzutowanie spinu cząstki na określony kierunek przestrzenny jest skwantowane zgodnie z wymiarem reprezentacji grupy obrotów (fakt istnienia takich reprezentacji wynika z rozważań entropijnych [7]). Jednak nie oznacza to, że nie istnieje model teoriopowy, który by takie kwantowanie rzutowania opisywał. Musiałby on jedynie zakładać, że po pierwsze spinowe stopnie swobody (z ciągłym rozkładem jego kierunku przed pomiarem) nie mają (prostego) charakteru czaso-przestrzennego [48], a po drugie, że bezwładność cząstki związana ze spinowymi stopniami swobody jest bardzo mała w porównaniu z siłą sprzężenia tych spinowych stopni swobody z otoczeniem (np. aparaturą traktowaną jako rezerwuuar), którego moment pędu ignorujemy. Wtedy przejście cząstki ze spinem przez jakąkolwiek aparaturę pomiarową typu Sterna-Gerlacha porządkowałoby jej spin w sposób dyskretny. Oczywiście zmiana momentu pędu aparatury już nas “nie interesuje”.

Niech także:

$$S_a S_b \equiv S_{ab} \in \Omega_{ab} = \{S_{++}, S_{--}, S_{+-}, S_{-+}\} \equiv \{(++), (--), (+-), (-+)\} \quad (5.145)$$

opisuje łączną przestrzeń zdarzeń Ω_{ab} stanów spinowych pary cząstek 1 i 2.

Oznaczmy przez $P(S_{ab}|\vartheta)$ cztery możliwe łączne warunkowe prawdopodobieństwa:

$$P(++|\vartheta), \quad P(--|\vartheta), \quad P(+ -|\vartheta), \quad P(- +|\vartheta). \quad (5.146)$$

W analizie estymacyjnej postulujemy, że prawdopodobieństwo $P(S_{ab}|\vartheta) \Big|_{\vartheta=\hat{\vartheta}}$ jest funkcją estymatora $\hat{\vartheta}$ parametru ϑ . Poprzez analizę tego prawdopodobieństwa, korzystając z proporcjonalności prawdopodobieństwa do obserwowanej liczebności zdarzeń w detektorze, oszacowuje się wartość kąta ϑ . Jednak aby to uczynić, trzeba mieć model analitycznych formuł na prawdopodobieństwa (5.146). Poniżej wyprowadzimy je metodą EFI.

W celu rozwiązania równań różniczkowych EFI konieczne jest ustalenie warunków brzegowych. Wynikają one z zasad zachowania i symetrii przestrzennej badanego układu.

Zauważmy, że warunek normalizacji:

$$\sum_{ab} P(S_{ab}) = 1, \quad (5.147)$$

jest spełniony dla każdego ϑ .

Kolejne warunki wynikają z symetrii i warunku zachowania całkowitego spinu dla rozważanego siebie niezależne (łączne prawdopodobieństwa nie jest iloczynem prawdopodobieństw brzegowych), wtedy:

$$I_F(P_{12}) \geq I_F(P_1) + I_F(P_2) \equiv \tilde{C} \quad (5.144)$$

gdzie \tilde{C} jest pojemnością informacyjną złożonego układu. Relacja (5.144) oznacza, że jeśli występują jakiekolwiek korelacje między zmiennymi to, jeśli znamy wynik doświadczenia dla pierwszej zmiennej to maleje ilość informacji koniecznej do określenia wyniku doświadczenia dla drugiej z nich, tzn. istnienie korelacji w układzie zwiększa informację Fishera I_F o parametrach charakteryzujących rozkład układu.

eksperymentu, przy czym w eksperymencie *względny orbitalny moment pędu wynosi zero*.

Wpierw rozważmy prosty przypadek $\vartheta = 0$, gdy obie płaszczyzny, w których ustawione są urządzenia Sterna-Gerlacha są tak samo zorientowane. Z warunku zachowania całkowitego spinu wynika:

$$P(+ + | 0) = P(- - | 0) = 0, \quad \vartheta = 0 \quad (5.148)$$

co oznacza, że w tym ustawieniu aparatury nigdy nie zobaczymy obu spinów jednocześnie skierowanych w górę czy w dół. Analogicznie jeśli kąt $\vartheta = \pi$, to warunek:

$$P(+ - | \pi) = P(- + | \pi) = 0, \quad \vartheta = \pi, \quad (5.149)$$

oznacza, że w tym przypadku nigdy nie zobaczymy spinów ustawionych jeden w górę drugi w dół. W konsekwencji z zasady zachowania całkowitego spinu otrzymujemy, że jeśli $\vartheta = 0$ lub $\vartheta = \pi$, to zawsze obserwacja jednego spinu daje nam całkowitą wiedzę o drugim. W tym przypadku stany spinów wyraźnie nie są niezależne, są stanami skorelowanymi. Wniosek ten jest intuicyjnie zawarty w sposobie ich przygotowania.

Następnie, ponieważ $P(S_b|\vartheta)$ jest prawdopodobieństwem brzegowym wystąpienia określonej wartości rzutu spinu cząstki 2, zatem nie zależy ono od S_a , czyli od orientacji rzutu spinu cząstki 1, więc nie zależy również od kąta ϑ pomiędzy wektorami \vec{a} oraz \vec{b} , czyli:

$$P(S_b|\vartheta) = C = const. \quad (5.150)$$

Stąd

$$P(S_b) = \int_0^{2\pi} P(S_b|\vartheta) r(\vartheta) d\vartheta = C \int_0^{2\pi} r(\vartheta) d\vartheta = C \quad \text{dla } S_b = +, - \quad (5.151)$$

gdzie

$$r(\vartheta) = \frac{1}{2\pi}, \quad 0 \leq \vartheta < 2\pi \quad (5.152)$$

co oznacza, że w zakresie aparaturowej zmienności ustawienia wartości kąta $\vartheta \in (0, 2\pi)$, z powodu naszej niewiedzy, jest możliwa w równym stopniu każda jego wartość. Tak czy owak wartość ϑ jest związana z ustawieniem aparatury pomiarowej Sterna-Gerlacha a i b , i trudno ją (na serio) traktować jako wielkość posiadającą rozproszenie. Zasadnicza ϑ jest parametrem charakterystycznym dla przeprowadzonego eksperymentu. Nie mniej przez niektórych $r(\vartheta)$ jest widziane jako „prawdopodobieństwo” znane a priori, co oznacza, że wyprowadzoną w ten sposób mechanikę kwantową należałoby traktować jako statystyczną teorię Bayesowską [49].

Z warunku unormowania, prawdopodobieństwa zdarzenia pewnego, mamy

$$P(S_b = +) + P(S_b = -) = C + C = 2C = 1 \quad \text{czyli } C = \frac{1}{2} \quad (5.153)$$

Zatem z (5.150) otrzymujemy, że⁸:

$$P(S_b|\vartheta) = \frac{1}{2}. \quad (5.155)$$

⁸Podobnie dla S_a , wychodząc z $P(S_a|\vartheta) = C' = const.$ i postępując analogicznie jak przy przejściu od (5.150) do (5.153), otrzymujemy:

$$P(S_a|\vartheta) = \frac{1}{2}. \quad (5.154)$$

Inną ważną własnością symetrii przestrzennej eksperymentu jest brak preferencji występowania spinu skierowanego w górę czy dół, tzn:

$$P(S_{+-}|\vartheta) = P(S_{-+}|\vartheta) \quad \text{oraz} \quad P(S_{++}|\vartheta) = P(S_{--}|\vartheta) \quad (5.156)$$

Co oznacza, że jeśli obserwowalibyśmy eksperyment dla układu odwróconego względem osi ϑ o kąt π , statystyczny wynik byłby dokładnie taki sam. Z (5.156) otrzymujemy

$$P(S_{++}) = \int P(S_{++}|\vartheta) r(\vartheta) d\vartheta = \int P(S_{--}|\vartheta) r(\vartheta) d\vartheta = P(S_{--}) \quad (5.157)$$

oraz w analogiczny sposób

$$P(S_{+-}) = P(S_{-+}) \quad (5.158)$$

Dodatkowo otrzymujemy:

$$\begin{aligned} P(S_{+-}) &= \int P(S_{+-}|\vartheta) r(\vartheta) d\vartheta = \int P(S_{+-}|\vartheta + \pi) r(\vartheta + \pi) d\vartheta \\ &= \int P(S_{++}|\vartheta) r(\vartheta + \pi) d\vartheta = P(S_{++}) , \end{aligned} \quad (5.159)$$

gdzie w drugiej równości skorzystano ze zwykłej zamiany zmiennych $\vartheta \rightarrow \vartheta + \pi$, w trzeciej z tego, że $P(S_{+-}|\vartheta + \pi) = P(S_{++}|\vartheta)$ a w ostatniej z $r(\vartheta) = 1/(2\pi) = r(\vartheta + \pi)$. Jako konsekwencja (5.157)-(5.159) otrzymujemy:

$$P(S_{-+}) = P(S_{--}) \quad (5.160)$$

Ponieważ zdarzenia (5.145) wykluczają się wzajemnie i tworzą całą przestrzeń zdarzeń, zatem:

$$P(S_{++}) + P(S_{+-}) + P(S_{-+}) + P(S_{--}) = 1 \quad (5.161)$$

zatem w świetle (5.157)-(5.161) otrzymujemy:

$$P(S_{ab}) = \frac{1}{4} . \quad (5.162)$$

W końcu wzór Bayes'a na *prawdopodobieństwo warunkowe* daje:

$$P(S_a|S_b) = \frac{P(S_{ab})}{P(S_b)} = \frac{1/4}{1/2} = \frac{1}{2} \quad \text{dla każdego } S_a, S_b . \quad (5.163)$$

Zauważmy, że jak dotąd w wyprowadzeniu zależności (5.148) do (5.163) nie skorzystano z EFI. Zależności te wynikały z początkowego założenia o istnieniu dla cząstki o spinie połówkowym dokładnie dwóch możliwych rzutów spinu na dowolny kierunek, zasady zachowania całkowitego momentu pędu oraz z symetrii układu.

5.2.2 Pojemność informacyjna dla zagadnienia EPR-Bohm'a

Poniżej wyprowadzimy wyrażenia dla prawdopodobieństw (5.146) jako wynik estymacji rozkładów metodą EFI. Podobnie jak poprzednio układ sam próbkuje swoimi Fisherowskimi kinetycznymi stopniami swobody dostępną mu przestrzeń położeń. Tym razem jest to jednowymiarowa przestrzeń kąta ϑ . Obserwator zewnętrzny mierzy jedynie wartość rzutu spinu S_a jednej cząstki, *powiedzmy, że cząstki 1*, natomiast nie mierzy wartości rzutu spinu drugiej cząstki S_b ,

jest ona traktowana jako wielkość nieznana, ale ustalona. Również wartość S_b nie jest estymowana z obserwacji S_a . Na podstawie danych pomiarowych wartości rzutu spinu S_a estymujemy kąt ϑ pomiędzy dwoma płaszczyznami zaznaczonymi na rysunku 5.1. Kąt ten jest więc traktowany jako nieznanymi parametr i jest estymowany z obserwacji S_a . Nie estymujemy natomiast wartości S_b z obserwacji S_a .

Ponieważ dokonujemy pomiaru S_a , a S_b oraz ϑ są nieznanymi, ale ustalonymi wielkościami, zatem funkcję wiarygodności dla powyższej postawionego problemu zapiszmy następująco:

$$P(S_a|S_b, \vartheta) \rightarrow \text{funkcja wiarygodności próby dla } \vartheta. \quad (5.164)$$

Jej postaci szukamy wykorzystując metodę EFI. Jednak postać informacji Fishera, która byłaby miarą precyzji estymacji kąta ϑ opartej o pomiar S_a , wynikać będzie z jej pierwotnej postaci (2.99) dla wymiaru próby $N = 1$. Skoro więc ϑ jest estymowanym parametrem, zatem pełni on teraz taką funkcję jak parametrem θ w wzorze (2.99), w którym obok podstawienia $\theta \rightarrow \vartheta$ należy dokonać następujących podstawień:

$$n = N = 1, \quad \text{oraz} \quad \int dy^1 \rightarrow \sum_a. \quad (5.165)$$

Wprowadzając jak zwykle amplitudy prawdopodobieństwa q_{ab} :

$$P(S_a|S_b, \vartheta) = q_{ab}^2(\vartheta), \quad (5.166)$$

otrzymujemy:

$$\begin{aligned} I_b &= \sum_{a=-}^+ \frac{1}{P(S_a|S_b, \vartheta)} \left(\frac{\partial P(S_a|S_b, \vartheta)}{\partial \vartheta} \right)^2 = \sum_{a=-}^+ \frac{1}{q_{ab}'^2} \left(\frac{\partial q_{ab}'^2}{\partial \vartheta} \right)^2 = \sum_{a=-}^+ \frac{1}{q_{ab}'^2} \left(2q_{ab} \frac{\partial q_{ab}}{\partial \vartheta} \right)^2 \\ &= 4 \sum_{a=-}^+ \left(\frac{\partial q_{ab}}{\partial \vartheta} \right)^2, \end{aligned} \quad (5.167)$$

lub podsumowując:

$$I_b \equiv I(S_b, \vartheta) = 4 \sum_{a=-}^+ q_{ab}'^2(\vartheta), \quad S_b = (+, -), \quad \text{gdzie} \quad q_{ab}' \equiv \frac{dq_{ab}}{d\vartheta}, \quad (5.168)$$

gdzie sumowanie po a odpowiada całkowaniu w informacji Fishera po przestrzeni bazowej.

Uwaga o informacji Fishera parametru ϑ : Zwróćmy uwagę, że ϑ w powyższej formule na $I_b \equiv I(S_b, \vartheta)$ jest wyznaczone dla konkretnej wartości ϑ i konkretnej wartości S_b rzutu spinu cząstki b . Zatem jest to pojemność informacyjna jednego kanału (ϑ, S_b) czyli *informacja Fishera parametru ϑ* w tym kanale:

$$IF(\vartheta) = I_b. \quad (5.169)$$

Powróćmy do warunków (5.165). Pierwszy z nich oznacza, że ranga amplitudy pola jest równa $N = 1$, drugi oznacza, że *przestrzeń bazowa to teraz dwupunktowy zbiór wartości rzutu spinu cząstki 1 na kierunek 'a'*. W poniższych uwagach odniesiemy się do tych elementów analizy.

Uwaga o randze $N = 1$. Zachodzi pytanie: Jeśli okaże się, że tradycyjne formuły mechaniki kwantowej dla eksperymentu EPR-Bohm'a pojawią się dla powyżej określonej estymacji z $N = 1$ (a tak się istotnie stanie), to czyżby pomiar oddziaływania cząstki z aparaturą Sterna-Gerlach'a miał mieć charakter termodynamiczny oddziaływania małego układu z termostatem, w rozumieniu podanym na samym początku Rozdziału 5.2.1? W końcu jak wiemy z poprzednich rozdziałów, rozwiązania z $N = 1$ odnoszą się do zjawisk termodynamicznych.

Uwaga o sumowaniu po a : Jako, że ϑ jest parametrem, a pomiary w próbie są związane z obserwacjami S_a , zatem zapis w (5.166) wymaga pewnego wyjaśnienia. Wyrażenie (5.168) reprezentuje dwa równania dla dwóch możliwych wartości S_b . Ponieważ zmienną w pomiarze jest rzut spinu cząstki 1, zatem sumowanie przebiega po dwóch możliwych wartościach spinu $S_a = +, -$. Tak więc zapis $I(S_b, \vartheta)$ wskazuje na informację zawartą w obserwacji S_a dla cząstki 1 na temat nieznanego kąta ϑ w obecności pewnej nieznannej, lecz ustalonej wartości rzutu spinu S_b cząstki 2. Ponieważ kąt ϑ jest parametrem, którego wartość może się zmieniać w sposób ciągły w przedziale $(0, 2\pi)$, stąd zgodnie z (5.168) mamy nieskończoną liczbę kanałów informacji Fishera związanych z wartościami ϑ . Dla każdego z tej nieskończonej liczby kanałów są jeszcze dwa kanały informacyjne związane z możliwymi wartościami dla S_b . Aby poradzić sobie z taką sytuacją metoda EFI wykorzystuje pojedynczą, skalarną informację (oznaczy ją I_{1_a}), nazywaną pojemnością informacyjną wprowadzoną w Rozdziale 2.7. Wielkość tą konstruujemy dokonując sumowania informacji po wszystkich możliwych kanałach. Zatem, suma przebiegać będzie po wszystkich wartościach kąta ϑ_k przy nieznanym rzucie spinu S_b cząstki (2).

Tak więc, po pierwsze, pojemność informacyjna I_{nk} dla jednego kanału, wprowadzona po raz pierwszy w (2.113), przyjmuje postać:

$$I_{bk} = I(S_b, \vartheta_k) . \quad (5.170)$$

Po drugie, w wyniku sumowania po wszystkich możliwych wartościach kąta ϑ_k oraz rzutach spinu S_b cząstki (2), otrzymujemy (całkowitą) pojemność informacyjną dla parametru ϑ :

$$\begin{aligned} I_{1_a} &\equiv \sum_{b=-}^{+} \sum_k I(S_b, \vartheta_k) \rightarrow \\ &\rightarrow I_{1_a} = \sum_{b=-}^{+} \int_0^{2\pi} d\vartheta I(S_b, \vartheta) = 4 \sum_{b=-}^{+} \sum_{a=-}^{+} \int_0^{2\pi} d\vartheta q_{ab}'^2(\vartheta) \equiv 4 \sum_{ab} \int_0^{2\pi} d\vartheta q_{ab}'^2(\vartheta) , \end{aligned} \quad (5.171)$$

gdzie całkowanie pojawia się z powodu zastąpienia sumy dla dyskretnego indeksu k całkowaniem po ciągłym zbiorze wartości parametru ϑ (tzn. po zrozumieniu czym jest kanał związany z ϑ powracamy do pierwotnego całkowania), które przebiega zgodnie z (5.152) od 0 do 2π . W ostatniej równości wykorzystano (5.168). Sumowanie w (5.171) po ab przebiega po wszystkich możliwych kombinacjach łącznej przestrzeni stanów S_{ab} określonej w (5.145). Indeks '1_a' w I_{1_a} oznacza, że jest to pojemność informacyjna jednej cząstki, gdzie wyróżniliśmy cząstkę a . Pojemność informacyjna I_{1_a} wchodzi do estymacyjnej procedury EFI.

5.2.3 Informacja strukturalna. Amplituda prawdopodobieństwa

Biorąc powyższe uwagi odnośnie postaci I_{1_a} w (5.171), zauważamy, że informację strukturalną Q_{1_a} dla obserwowanej cząstki jako

$$Q_{1_a} \equiv 4 \sum_{ab} \int_0^{2\pi} d\vartheta \frac{1}{4} q_{ab}^2(\vartheta) \mathbf{F}_{ab}(q_{ab}) \quad (5.172)$$

Poniższa analiza związana z wyprowadzeniem amplitud q_{ab} jest podobna do analizy dla rozkładu Boltzmanna z rozdziału 5.1.

Korzystając z (5.171) i (5.172) wariacyjna zasada informacyjna ma zatem postać:

$$\delta_{(q_{ab})} K \equiv \delta_{(q_{ab})} (I_{1_a} + Q_{1_a}) = \delta_{(q_{ab})} \left(4 \sum_{ab} \int_0^{2\pi} d\vartheta \left(q_{ab}'^2 + \frac{1}{4} q_{ab}^2(\vartheta) \mathbf{F}_{ab}(q_{ab}) \right) \right) = 0, \quad (5.173)$$

natomiast strukturalną zasadę informacyjną zapisujemy następująco:

$$I_{1_a} + \kappa Q_{1_a} = I_{1_a} + Q_{1_a} = 4 \sum_{ab} \int_0^{2\pi} d\vartheta \left(q_{ab}'^2 + \frac{1}{4} q_{ab}^2(\vartheta) \mathbf{F}_{ab}(q_{ab}) \right) = 0, \quad (5.174)$$

gdzie jak w przypadku wszystkich problemów bez więzów związanych z równaniem ciągłości (nazwijmy je 'kwantowych') przyjęliśmy $\kappa = 1$, czyli:

$$I_{1_a} = -Q_{1_a}. \quad (5.175)$$

Szukamy wspólnych rozwiązań na amplitudę q_{ab} oraz obserwowaną informację strukturalną $\mathbf{F}_{ab}(q_{ab})$ dla układu równań (5.173) oraz (5.174). Rozwiązaniem problemu wariacyjnego (5.173) jest (porównaj (5.16)):

$$\frac{d}{d\vartheta} \left(\frac{\partial k}{\partial q_{ab}'(\vartheta)} \right) = \frac{\partial k}{\partial q_{ab}}, \quad (5.176)$$

gdzie

$$k = 4 \sum_{ab} \left(q_{ab}'^2 + \frac{1}{4} q_{ab}^2(\vartheta) \mathbf{F}_{ab}(q_{ab}) \right) \quad (5.177)$$

jest funkcją podcałkową w (5.173). Z równania (5.176) otrzymujemy:

$$q_{ab}'' = \frac{1}{2} \frac{d \left(\frac{1}{4} q_{ab}^2 \mathbf{F}_{ab}(q_{ab}) \right)}{dq_{ab}}. \quad (5.178)$$

Ponieważ $q_{ab}^2(\vartheta) \mathbf{F}_{ab}(q_{ab})$ jest jawnie jedynie funkcją q_{ab} , więc różniczka zupełna zastąpiła pochodną cząstkową po q_{ab} występującą w (5.176).

Aby znaleźć rozwiązanie zasady strukturalnej (5.174), wykonajmy w (5.171) całkowanie przez części:

$$I_{1_a} = 4 \sum_{ab} \int_0^{2\pi} d\vartheta \left(C_{ab} - q_{ab} q_{ab}'' \right), \quad (5.179)$$

gdzie

$$C_{ab} = \frac{1}{2\pi} \left(q_{ab}(2\pi) q'_{ab}(2\pi) - q_{ab}(0) q'_{ab}(0) \right). \quad (5.180)$$

Podstawiając (5.179) do (5.174) otrzymujemy:

$$I_{1_a} + Q_{1_a} = 4 \sum_{ab} \int_0^{2\pi} d\vartheta \left(C_{ab} - q_{ab} q''_{ab} + \frac{1}{4} q_{ab}^2 \mathbf{F}_{ab}(q_{ab}) \right) = 0. \quad (5.181)$$

Przenosząc spełnienie (5.181) na poziom obserwowanej zasady strukturalnej dla każdego (dwu)-cząstkowego łącznego stanu spinowego S_{ab} , otrzymujemy:

$$C_{ab} - q_{ab} q''_{ab} + \frac{1}{4} q_{ab}^2 \mathbf{F}_{ab}(q_{ab}) = 0. \quad (5.182)$$

Podstawiając do (5.182) związek (5.178) otrzymany z wariacyjnej zasady informacyjnej, otrzymujemy:

$$\frac{1}{2} q_{ab} \frac{d \left(\frac{1}{4} q_{ab}^2 \mathbf{F}_{ab}(q_{ab}) \right)}{dq_{ab}} = \frac{1}{4} q_{ab}^2 \mathbf{F}_{ab}(q_{ab}) + C_{ab}. \quad (5.183)$$

Zapiszmy powyższe równanie w wygodniejszej formie:

$$\frac{2dq_{ab}}{q_{ab}} = \frac{d \left(\frac{1}{4} q_{ab}^2 \mathbf{F}_{ab}(q_{ab}) \right)}{\frac{1}{4} q_{ab}^2 \mathbf{F}_{ab}(q_{ab}) + C_{ab}}, \quad (5.184)$$

z której po obustronnym wycalkowaniu otrzymujemy:

$$\frac{1}{4} q_{ab}^2(\vartheta) \mathbf{F}_{ab}(q_{ab}) = \frac{q_{ab}^2(\vartheta)}{A_{ab}^2} - C_{ab}, \quad (5.185)$$

gdzie A_{ab}^2 jest w ogólności zespoloną stałą. Podstawiając (5.185) do (5.178) otrzymujemy szukane różniczkowe równanie generujące dla amplitud:

$$q_{ab}''(\vartheta) = \frac{q_{ab}(\vartheta)}{A_{ab}^2}. \quad (5.186)$$

Przypomnijmy, że jest ono konsekwencją obu zasad informacyjnych, strukturalnej i wariacyjnej. Ponieważ amplituda q_{ab} jest rzeczywista więc A_{ab}^2 też musi być rzeczywiste. Podobnie jak w rozdziale 5.1.3, ponieważ stała A_{ab}^2 jest rzeczywista, więc można ją przedstawić za pomocą innej rzeczywistej stałej a_{ab} jako $A_{ab} = a_{ab}$ lub $A_{ab} = i a_{ab}$. Zatem istnieją dwie klasy rozwiązań równania (5.186).

Dla $A_{ab} = a_{ab}$, rozwiązanie (5.186) ma charakter czysto eksponencjalny:

$$q_{ab}(\vartheta) = B_{ab} \exp\left(-\frac{\vartheta}{a_{ab}}\right) + C_{ab} \exp\left(\frac{\vartheta}{a_{ab}}\right), \quad A_{ab} = a_{ab}, \quad (5.187)$$

gdzie stałe B_{ab} oraz C_{ab} są rzeczywiste.

Natomiast dla $A_{ab} = i a_{ab}$, rozwiązanie (5.186) ma charakter trygonometryczny:

$$q_{ab}(\vartheta) = B'_{ab} \sin\left(\frac{\vartheta}{a_{ab}}\right) + C'_{ab} \cos\left(\frac{\vartheta}{a_{ab}}\right), \quad A_{ab} = i a_{ab}, \quad (5.188)$$

gdzie a_{ab} , B'_{ab} , C'_{ab} są rzeczywistymi stałymi.

W rozważanym obecnie przypadku⁹ wartości ϑ są kątem z ograniczonego zbioru $\langle 0, 2\pi \rangle$. Zatem z geometrycznej symetrii układu przy obrocie o kąt 2π wynika, że rozkład $P(S_a|S_b, \vartheta)$ jest również funkcją okresową zmiennej ϑ . Za względu na (5.166) oznacza to, że funkcja $q_{ab}(\vartheta)$ powinna być również okresowa, *zatem wybieramy rozwiązanie o charakterze trygonometrycznym*, przy czym Funkcje \sin oraz \cos w (5.188) są funkcjami bazowymi tworzącymi amplitudę prawdopodobieństwa $q_{ab}(\vartheta)$.

Funkcje bazowe na przestrzeni amplitud powinny być ortogonalne. Z postaci q_a w (5.188) wynika, że funkcjami bazowymi są funkcje 'sin' oraz 'cos'. Ponieważ q_a jest określone na przestrzeni parametru $\vartheta \in \langle 0, 2\pi \rangle$, zatem warunek ortogonalności funkcji bazowych na tej przestrzeni:

$$\int_0^{2\pi} d\vartheta \sin(\vartheta/a_{ab}) \cos(\vartheta/a_{ab}) = 0 \quad (5.189)$$

daje po wyciągnięciu stałych a_{ab} :

$$a_{ab} = \frac{2}{n_{ab}}, \quad \text{gdzie } n_{ab} = 1, 2, \dots \quad (5.190)$$

Warunek (5.190) jest warunkiem na dopuszczalne wartości $a_{ab} = -iA_{ab}$ (por. (5.188)), ale jak widać nie wskazuje on na żadną z nich jednoznacznie. Może to nastąpić, gdy ustalimy wartość n_{ab} , co uczynimy ograniczając rozważania do zapostulowanego w Rozdziale 2.7.1 warunku minimalizacji informacji kinetycznej $I \rightarrow \min$. Wykorzystajmy zasadę strukturalną (5.180), która po wstawieniu do niej (5.188) oraz (5.190) daje postać stałej C_{ab} :

$$C_{ab} = \frac{1}{2\pi} B'_{ab} C'_{ab} \frac{n_{ab}}{2} [\cos^2(n_{ab}\pi) - 1] = 0, \quad \text{gdzie } n_{ab} = 1, 2, \dots \quad (5.191)$$

Po eliminacji w (5.179) wyrażenia q''_{ab} korzystając z równania generującego (5.186), następnie po skrzestaniu z postaci stałej C_{ab} podanej w (5.191), otrzymujemy użyteczną postać pojemności informacyjnej:

$$I_1 = -4 \sum_{ab} \frac{1}{A_{ab}^2} \int_0^{2\pi} d\vartheta q_{ab}^2(\vartheta) \quad (5.192)$$

Należy jeszcze wyznaczyć ostatnią całkę:

$$\begin{aligned} \int_0^{2\pi} d\vartheta q_{ab}^2(\vartheta) &\equiv \int_0^{2\pi} d\vartheta P(S_a|S_b, \vartheta) = \int_0^{2\pi} d\vartheta \frac{P(S_a, S_b, \vartheta)}{p(\vartheta, S_b)} = \int_0^{2\pi} d\vartheta \frac{p(\vartheta, S_a|S_b) P(S_b)}{p(\vartheta, S_b)} \\ &= \int_0^{2\pi} d\vartheta \frac{p(\vartheta, S_a|S_b)^{\frac{1}{2}}}{p(S_b|\vartheta) r(\vartheta)} = \int_0^{2\pi} d\vartheta \frac{p(\vartheta, S_a|S_b)^{\frac{1}{2}}}{\frac{1}{2} \frac{1}{2\pi}} \\ &= 2\pi \int_0^{2\pi} d\vartheta p(\vartheta, S_a|S_b) = 2\pi P(S_a|S_b) = 2\pi \frac{1}{2} = \pi, \end{aligned} \quad (5.193)$$

⁹Przypomnijmy sobie analizę przeprowadzoną w rozdziale (5.1.3) dla rozkładu energii cząsteczki gazu. Wtedy ze względu na nieograniczony zakres argumentu amplitudy wybraliśmy rozwiązanie o charakterze eksponencjalnym.

gdzie w pierwszej linii i) w pierw skorzystano z definicji amplitudy, $q_{ab}^2(\vartheta) = P(S_a | S_b, \vartheta)$, (5.166), ii) następnie z definicji prawdopodobieństwa warunkowego, iii) znowu z definicji prawdopodobieństwa warunkowego wprowadzając:

$$p(\vartheta, S_a | S_b) = \frac{P(S_a, S_b, \vartheta)}{P(S_b)}, \quad (5.194)$$

a w kolejnej linii z wyrażen (5.152),(5.155), i w ostatniej z:

$$P(S_a | S_b) = \int_0^{2\pi} p(\vartheta, S_a | S_b) d\vartheta, \quad (5.195)$$

i w końcu z (5.163). Jesteśmy już w stanie wyrazić informację I_1 tylko poprzez n_{ab} . Dodatkowo korzystając z wewnętrznej zasady informacyjnej, $I_1 = -Q_1$, (5.175), z wyniku (5.193), otrzymujemy dla $A_{ab} = ia_{ab}$ i po skorzystaniu z (5.190):

$$I_{1_a} = -Q_{1_a} = -4 \sum_{ab} \frac{1}{A_{ab}^2} \int_0^{2\pi} d\vartheta q_{ab}^2 = 4\pi \sum_{ab} \frac{1}{a_{ab}^2} = \pi \sum_{ab} n_{ab}^2, \quad \text{gdzie } n_{ab} = 1, 2, \dots \quad (5.196)$$

Z powyższego związku wynika, że warunek $I_{1_a} \rightarrow \min$ będzie spełniony dla $n_{ab} = 1$ dla dowolnych S_a oraz S_b . Warunek $n_{ab} = 1$ zgodnie z (5.190) odpowiada wartości a_{ab} :

$$a_{ab} = 2 \quad \text{dla dowolnych } S_a \text{ oraz } S_b. \quad (5.197)$$

Sumowanie w (5.196) przebiega po wszystkich S_a i S_b , zatem dla $n_{ab} = 1$, otrzymujemy wartość minimalną I_{1_a} :

$$I_{1_a}(\min) = \pi \sum_{a=-}^{+} \sum_{b=-}^{+} n_{ab}^2 = 4\pi, \quad \text{gdzie } n_{ab} = 1. \quad (5.198)$$

W wyrażeniu (5.188), które dla $a_{ab} = 2$ przyjmuje postać:

$$q_{ab}(\vartheta) = B'_{ab} \sin\left(\frac{\vartheta}{2}\right) + C'_{ab} \cos\left(\frac{\vartheta}{2}\right), \quad (5.199)$$

występują jeszcze stałe B'_{ab} oraz C'_{ab} , które również musimy wyznaczyć. Wyznaczymy je korzystając z wcześniej dla $\vartheta = 0$ ustalonych w (5.148) wartości łącznego prawdopodobieństwa $P(S_{ab}|\vartheta)$, które wykorzystując (5.155) możemy wyrazić przez amplitudę q_{ab} w następujący sposób:

$$\begin{aligned} P(S_{ab}|\vartheta) &\equiv P(S_a S_b|\vartheta) = \frac{P(S_a S_b, \vartheta)}{r(\vartheta)} = P(S_a | S_b, \vartheta) \frac{P(S_b, \vartheta)}{r(\vartheta)} \\ &= P(S_a | S_b, \vartheta) P(S_b|\vartheta) = q_{ab}^2(\vartheta) P(S_b|\vartheta) = \frac{1}{2} q_{ab}^2(\vartheta). \end{aligned} \quad (5.200)$$

Równanie (5.200) pokazuje, że:

$$q_{ab}(\vartheta) = 0 \quad \text{jeśli } P(S_{ab}|\vartheta) = 0. \quad (5.201)$$

Wstawiając (5.199) dla $\vartheta = 0$ do (5.200) i porównując otrzymany wynik z (5.148), $P(++|0) = P(--|0) = 0$, otrzymujemy:

$$C'_{++} = C'_{--} = 0. \quad (5.202)$$

Centralne założenie statystyczne dla eksperymentu EPR-Bohm'a: Rozkład prawdopodobieństwa (5.146), którego szukamy określony jest na przestrzeni zdarzeń $S_a S_b \equiv S_{ab} \in \Omega_{ab}$, (5.145), zatem zbiór możliwych wyników to $i = (++) , (--) , (+-) , (-+)$. Zatem ze związku (2.80) określającego metrykę Rao-Fishera $g_{ab} = \sum_{i=1}^N \frac{\partial q^i}{\partial \theta^a} \frac{\partial q^i}{\partial \theta^b}$, otrzymujemy w rozważanym obecnie zagadnieniu, po skorzystaniu z pochodnej z (5.199):

$$\frac{\partial q_{ab}}{\partial \vartheta} = \frac{1}{2} \left(B'_{ab} \cos \left(\frac{\vartheta}{2} \right) - C'_{ab} \sin \left(\frac{\vartheta}{2} \right) \right), \quad (5.203)$$

następująca postać tej metryki na jednowymiarowej podprzestrzeni statystycznej parametryzowanej w bazie ϑ a indukowanej z rozkładu (5.146):

$$\begin{aligned} g_{\vartheta \vartheta}(\vartheta) &= \sum_{ab} \frac{\partial q_{ab}}{\partial \vartheta} \frac{\partial q_{ab}}{\partial \vartheta} \\ &= \frac{1}{4} \sum_{ab} \left((B'_{ab})^2 + ((B'_{ab})^2 - (C'_{ab})^2) \sin^2 \left(\frac{\vartheta}{2} \right) - B'_{ab} C'_{ab} \sin^2 \vartheta \right). \end{aligned} \quad (5.204)$$

Założmy niezależność metryki Rao-Fishera $g_{\vartheta \vartheta}$ od wartości parametru ϑ . Wtedy (5.206) prowadzi do warunków:

$$g_{\vartheta \vartheta} = g_{\vartheta \vartheta}(\vartheta = 0) = \frac{1}{4} \sum_{ab} (B'_{ab})^2 = \frac{1}{4} (2(B'_{++})^2 + 2(B'_{+-})^2) \quad \text{dla } \vartheta = 0, \quad (5.205)$$

gdzie skorzystano z symetrii (5.156) otrzymując $B'_{++} = B'_{--}$ i $B'_{+-} = B'_{-+}$, oraz:

$$\begin{aligned} g_{\vartheta \vartheta} &= g_{\vartheta \vartheta}(\vartheta = \pi) = \frac{1}{4} \sum_{ab} \left(2(B'_{ab})^2 - (C'_{ab})^2 \right) \\ &= \frac{1}{4} \left(4(B'_{++})^2 + 4(B'_{+-})^2 - 2(C'_{+-})^2 \right) \quad \text{dla } \vartheta = \pi, \end{aligned} \quad (5.206)$$

gdzie skorzystano z (5.202). Z (5.205) oraz (5.206) otrzymujemy:

$$(B'_{++})^2 + (B'_{+-})^2 = (C'_{+-})^2 \geq 0. \quad (5.207)$$

Warunek ten oznacza, że $C_{+-} \neq 0$ gdyż w przeciwnym wypadku, tzn. dla $C_{+-} = 0$, z warunku (5.207) oraz z (5.202) otrzymalibyśmy zerowanie się wszystkich współczynników B'_{ab} oraz C'_{ab} , co odpowiadało by trywialnemu przypadkowi braku rozwiązania dla zagadnienia EPR. Zatem otrzymujemy:

$$C \equiv C_{+-} = C_{-+} \neq 0, \quad (5.208)$$

gdzie dla równości współczynników odwołano się również do symetrii (5.156).

W końcu warunek niezależności $g_{\vartheta \vartheta}$ od ϑ dla dowolnej wartości ϑ daje po uwzględnieniu (5.205):

$$\sum_{ab} \left((B'_{ab})^2 - (C'_{ab})^2 \right) = 2((B'_{++})^2 + (B'_{+-})^2 - (C'_{+-})^2) = 0, \quad (5.209)$$

czyli warunek automatycznie spełniony na podstawie (5.206), oraz

$$\sum_{ab} B'_{ab} C'_{ab} = 2(B'_{+-} C'_{+-}) = 0. \quad (5.210)$$

Zatem warunek ten wraz z (5.208) oznacza:

$$B'_{+-} = 0 = B'_{-+} . \quad (5.211)$$

Ze względu na (5.200) druga z powyższych równości wynika z symetrii (5.156).

W końcu zauważmy, że ze względu na (5.211) oraz (5.207) warunek istnienia nietrywialnego rozwiązania oznacza, że:

$$B \equiv B'_{++} = B'_{--} \neq 0, \quad (5.212)$$

gdzie ponownie w pierwszej równości skorzystano z symetrii (5.156).

Podstawiając otrzymane wyniki do (5.199) otrzymujemy:

$$q_{++}(\vartheta) = q_{--}(\vartheta) = B \sin(\vartheta/2) , \quad q_{-+}(\vartheta) = q_{+-}(\vartheta) = C \cos(\vartheta/2) . \quad (5.213)$$

Jak widać równości współczynników w związkach (5.208) oraz (5.212) jest odbiciem równości odpowiednich amplitud wynikającej z symetrii odbicia przestrzennego (5.156) oraz (5.200).

Musimy jeszcze wyznaczyć stałe B oraz C . Zwodu wykluczania się różnych zdarzeń S_{ab} , (5.145), oraz warunku normalizacji $P(S_a S_b | \vartheta)$:

$$\sum_{ab} P(S_a S_b | \vartheta) = P\left(\bigcup_{ab} S_a S_b | \vartheta\right) = 1 . \quad (5.214)$$

otrzymujemy ze względu na (5.200) równanie:

$$\frac{1}{2} (q_{++}^2(\vartheta) + q_{--}^2(\vartheta) + q_{-+}^2(\vartheta) + q_{+-}^2(\vartheta)) = 1 . \quad (5.215)$$

Korzystając z (5.213) i (5.215) mamy:

$$B^2 \sin^2(\vartheta/2) + C^2 \cos^2(\vartheta/2) = (B^2 - C^2) \sin^2(\vartheta/2) + C^2 = 1 . \quad (5.216)$$

Porównując współczynniki stojące przy odpowiednich funkcjach zmiennej ϑ po lewej i prawej stronie drugiej równości powyższego wyrażenia otrzymujemy:

$$B^2 = C^2 = 1 , \quad (5.217)$$

co po wstawieniu do (5.213) daje ostatecznie:

$$q_{++}(\vartheta) = q_{--}(\vartheta) = \pm \sin(\vartheta/2) , \quad q_{-+}(\vartheta) = q_{+-}(\vartheta) = \pm \cos(\vartheta/2) . \quad (5.218)$$

Podstawiając amplitudy (5.218) do $P(S_{ab} | \vartheta) = \frac{1}{2} q_{ab}^2(\vartheta)$, (5.200), daje wynik na łączne prawdopodobieństwo otrzymania określonej kombinacji rzutów spinów przy zadanej wartości kąta ϑ :

$$P(++ | \vartheta) = P(-- | \vartheta) = \frac{1}{2} \sin^2(\vartheta/2) , \quad P(+- | \vartheta) = P(-+ | \vartheta) = \frac{1}{2} \cos^2(\vartheta/2) , \quad (5.219)$$

który jest przewidywaniem mechaniki kwantowej [47]. Wynik ten w ramach metody EFI został oryginalnie wyprowadzony w [7], jednakże powyższy sposób wyprowadzenia różni się z dwóch miejscach. Po pierwsze warunki brzegowe zostały ujęte w sposób bardziej przejrzysty [16], a po drugie w [7] odwołano się do własności ortogonalności wprowadzonych tam kwantowych amplitud, o czym wspomniemy na końcu rozdziału, a czego w powyższym wyprowadzeniu uniknięto wprowadzając w to miejsce *warunek niezależności metryki Rao-Fishera od wartości parametru ϑ* .

Wcześniej w (5.162) wyznaczono łączne prawdopodobieństwo $P(S_{ab}) = \frac{1}{4}$. Z drugiej strony z (5.150) i (5.153) otrzymano, że $P(S_b) = 1/2$ (analogicznie $P(S_a) = 1/2$), a więc wynika z tego, że:

$$P(S_{ab}) = P(S_a)P(S_b) , \quad (5.220)$$

warunek, który oznacza niezależność spinowych zmiennych S_a oraz S_b . Zatem powstaje pytanie, czy (5.220) nie przeczy założeniom eksperymentu EPR-Bohm'a? Zauważmy jednak, że prawdopodobieństwo $P(S_{ab})$ w (5.220) z definicji określa równoczesne pojawienie się kombinacji rzutów spinu S_{ab} przy braku informacji o zmiennej ϑ (w (5.220) nie ma zależności od ϑ). W konsekwencji takie wydarzenie opisane równością (5.220) pojawia się z wyniku eksperymentu EPR-Bohm'a, w którym ϑ nie jest znane, a jedynie wiemy, że zawiera się w przedziale $(0, 2\pi)$, czyli w sytuacji (5.152). Zatem można oczekiwać, że korelacja zmniejszy się po uśrednieniu po wszystkich kątach ϑ , a z takim uśrednieniem mamy do czynienia, gdy ϑ nie jest znane. Z (5.220) wynika, że po uśrednieniu po całym zakresie kąta ϑ korelacja nie tylko zmniejsza się, ale całkowicie znika. Tzn. pod warunkiem nieznanności ϑ zmienne spinowe S_a oraz S_b są niezależne.

Powyższy wniosek ten ma interesującą fizyczną interpretację. Mianowicie twierdzenie Ehrenfesta mówi, że średnie wartości kwantowych operatorów są równe ich klasycznym odpowiednikom. W jego świetle wyrażenie (5.220) mówi, że w przypadku uśrednienia po kącie ϑ stany splątane EPR-Bohm'a, na powrót ulegają klasycznej separacji.

Uwaga: Natomiast z (5.218) widać, że efekt korelacji spinu zmienia się bardzo mocno wraz z wartością kąta ϑ , i w samej rzeczy dają one warunek (por. (5.154) i (5.155)):

$$P(S_{ab}|\vartheta) \neq P(S_a|\vartheta)P(S_b|\vartheta) . \quad (5.221)$$

Uwaga: Warunek niezależności wyprowadzenia formuł (5.213) od mechaniki kwantowej można by nieco osłabić, tzn. na tyle aby statystyczność teorii była dalej widoczna. Frieden uczynił to jak następuje [7]. Pokażmy, że $q_{ab}(\vartheta)$ jest proporcjonalna do (*kwantowej*) *amplitudy* $\psi_{ab}(\vartheta)$ parametryzowanej parametrem ϑ dla stanów (5.145). Niech $P(S_{ab}, \vartheta)$ jest prawdopodobieństwem pojawienia się konfiguracji (a, b) rzutu spinów, *podczas gdy* parametr wynosi ϑ . Nie jest to więc prawdopodobieństwo łączne zajścia zdarzenia: "pojawiła się konfiguracji (a, b) rzutu spinów, oraz kąt ϑ ", gdyż ten ostatni nie jest zmienną losową. Funkcję $\psi_{ab}(\vartheta)$ określimy tak, że kwadrat jej modułu spełnia związek:

$$|\psi_{ab}(\vartheta)|^2 \equiv |\psi_{ab}(\vartheta)|^2 (\vartheta | S_{ab}) \equiv p(\vartheta | S_{ab}) = \frac{P(S_{ab}, \vartheta)}{P(S_{ab})} , \quad (5.222)$$

tzn. jest prawdopodobieństwem, że skoro pojawiłaby się łączna konfiguracja spinów S_{ab} to wartość kąta wynosi ϑ . komplikacja polega na tym, że $q_{ab}(\vartheta)$ opisuje losowe zachowanie zmiennej spinowej S_a , a nie ϑ , co oznacza, że $q_{ab}(\vartheta)$ jest amplitudą typu $\psi_{\vartheta}(ab)$ a nie $\psi_{ab}(\vartheta)$.

Aby pokazać, że $q_{ab}(\vartheta)$ jest proporcjonalna do $\psi_{ab}(\vartheta)$ skorzystajmy (5.222) a następnie z twierdzenia Bayesa oraz definicji prawdopodobieństwa warunkowego:

$$\begin{aligned} |\psi_{ab}(\vartheta)|^2 &= \frac{P(S_{ab}, \vartheta)}{P(S_{ab})} = \frac{P(S_{ab}|\vartheta)r(\vartheta)}{P(S_{ab}|S_b)P(S_b)} = \frac{P(S_{ab}|\vartheta)r(\vartheta)}{P(S_a|S_b)P(S_b)} = \frac{P(S_a|S_b, \vartheta)P(S_b)r(\vartheta)}{P(S_a|S_b)P(S_b)} \\ &= \frac{q_{ab}^2(\vartheta)r(\vartheta)}{P(S_a|S_b)} . \end{aligned} \quad (5.223)$$

Korzystając z (5.152), (5.163) oraz (5.166) otrzymujemy więc z (5.223), że

$$|\psi_{ab}(\vartheta)|^2 = \frac{1}{\pi} q_{ab}^2(\vartheta), \quad (5.224)$$

skąd

$$\psi_{ab}(\vartheta) = \frac{e^{i\alpha}}{\pi} q_{ab}(\vartheta), \quad \alpha \in \mathbf{R}. \quad (5.225)$$

Zatem, $\psi_{ab}(\vartheta) \propto q_{ab}(\vartheta)$, co oznacza, że¹⁰, skoro pojawiła się łączna konfiguracja spinów S_{ab} to amplituda prawdopodobieństwa $\psi_{ab}(\vartheta)$, że wartość kąta wynosi ϑ , jest proporcjonalna do amplitudy prawdopodobieństwa $q_{ab}(\vartheta)$ zaobserwowania rzutu spinu S_a cząstki 1 pod warunkiem, że rzut spinu cząstki 2 wynosiłby S_b a kąt ϑ . Zdanie to jest wyrazem splątania, które pojawiło się w kwantowym-mechaniczno opisie eksperymentu EPR-Bohm'a.

W końcu, ze względu na (5.224) amplitudy $q_{ab}(\vartheta)$ są proporcjonalna do amplitud $\psi_{ab}(\vartheta)$. Zatem Frieden zarządał ortogonalności amplitud kwantowych ψ_{++} i ψ_{+-} , skąd automatycznie wynika ortogonalność amplitud q_{++} oraz q_{+-} :

$$\int_0^{2\pi} d\vartheta q_{++} q_{+-} = \int_0^{2\pi} d\vartheta B'_{++} \sin(\vartheta/2) \left[B'_{+-} \sin(\vartheta/2) + C'_{+-} \cos(\vartheta/2) \right] = 0, \quad (5.226)$$

co pozwoliło na wyprowadzenie zerowania się B'_{+-} oraz B'_{-+} , (5.211), a dalej postępując już jak powyżej otrzymał formuły (5.219).

5.2.4 Niepewność wyznaczenia kąta

Jedna sprawa dotycząca analizy estymacyjnej opisanej powyższymi rachunkami wymaga jeszcze podkreślenia. Otóż w rzeczywistym pomiarze badacz otrzymuje wartości rzutu spinu S_a z określonymi częstościami, które są oszacowaniami rozkładu prawdopodobieństwa (5.219). Natomiast analiza, która doprowadziła do (5.219) jest, tak jak w całej analizie EFI (nie uwzględniającej wpływu urządzenia pomiarowego), wewnętrzną sprawą układu dokonującego estymacji parametru zgodnie z metodą EFI, zatem jest sprawą modelu EFI. Sytuacja ta ma dwie następujące konsekwencje.

Pierwsza, związana z (wewnętrznym) błędem przedstawionej metody (fizycznej estymacji) EFI, określa ten błąd jak następuje. Zauważmy, że nierówność Rao-Cramera ma postać, $\sigma_{\hat{\theta}}^2 T \geq 1/I(\theta)$, (2.29), gdzie tym razem estymator $T \equiv \hat{\vartheta}$. Z nierówności tej otrzymujemy warunek na wariancję estymatora $\hat{\vartheta}$ kąta ϑ :

$$\sigma^2(\hat{\vartheta}) \geq \frac{1}{IF(\vartheta)}, \quad (5.227)$$

gdzie $IF(\vartheta) = I_b$ jest informacją Fishera (5.169), a nie pojemnością informacyjną I_{1_a} podaną w (5.198). Bez analizowania postaci estymatora $\hat{\vartheta}$, nierówność Rao-Cramera (5.227) określa DORC dla jego wariancji, o ile tylko estymator ten jest nieobciążony.

Ponieważ pojemność informacyjna jest sumą po kanałach z ich informacji Fishera, więc:

$$I_{1_a} \geq IF(\vartheta). \quad (5.228)$$

¹⁰W mechanice kwantowej powiedzielibyśmy, że $\psi_{ab}(\vartheta)$ oznacza amplituda prawdopodobieństwa zdarzenia, że wartość kąta wynosi ϑ o ile pojawiłaby się łączna konfiguracja spinów S_{ab} .

Z (5.228) oraz z (5.227) otrzymujemy:

$$\frac{1}{I_{1_a}} \leq \frac{1}{IF(\vartheta)} \leq \sigma^2(\hat{\vartheta}) . \quad (5.229)$$

Ponieważ pojemność kanału I_{1_a} według (5.198) wynosi:

$$I_{1_a} = 4\pi , \quad (5.230)$$

więc podstawiając tą wartość do nierówności (5.229) :

$$\sigma^2(\hat{\vartheta}) \geq \frac{1}{4\pi} \approx 0.08 \text{ rad}^2 . \quad (5.231)$$

Nierówność (5.231) stwierdza, że obserwacja jednej wartości rzutu spinu przy *kompletnej nieznanomości kąta* ϑ (stąd (5.152)) daje o nim małą, lecz skończoną informację. Błąd jego estymacji $\sqrt{0,08} \text{ rad} = 0,28 \text{ rad}$ jest dość duży, co jest związane z płaską funkcją “niewiedzy” $r(\vartheta)$ określoną w (5.152).

5.2.4.1 Wpływ zaszumienia pomiaru

Analiza EFI w obliczu pomiaru i płynącego z tego faktu zaszumienia danych wewnętrznych EFI nie jest przedmiotem tego skryptu. Zainteresowany czytelnik znajdzie omówienie tego tematu w [7, 16]. Poniżej zawieram wniosek z analizy Mroziakiewicz [16] dotyczącą tego problemu, dla powyższej analizy eksperymentu EPR-Bohm’a: W pomiarze stanu układu przez zewnętrznego obserwatora, otrzymujemy dane zaszumione przez układ pomiarowy. Tzn. dane pomiarowe rzutu spinu (oznaczymy je \bar{S}_a) są generowane przez prawdziwe wartości wielkości S_a, S_b, ϑ w obecności szumu aparatury pomiarowej. Szum ten powstaje co prawda w urządzeniach Sterna-Gerlacha a oraz b ale, ze względu na założenie podane na samym początku Rozdziału 5.2.1, z innych przyczyn niż (przyjęte jako zero) fluktuacje rzutu spinu. Uzyskana przy tych danych informacja Fishera I_{zasz} o parametrze ϑ , uwzględniająca zaszumienie pomiarowe, jest generowana przez informację Fishera $IF(\vartheta) = I_b$ (5.169). W tym sensie informacja Fishera $IF(\vartheta)$, (5.169), procedury estymacyjnej EFI jest częścią informacji, która przejawia się w pomiarze. Zatem obok nierówności (5.228) zachodzi również:

$$IF(\vartheta) \geq I_{zasz}(\vartheta) . \quad (5.232)$$

Ze względu na to, że $I_{zasz}(\vartheta)$ wchodzi w (5.229) w miejsce $IF(\vartheta)$, warunek ten oznacza jednak pogorszenie jakości estymacji w porównaniu z (5.231).

5.2.5 Informacja Q jako miara splątania

Tak jak we wszystkich problemach estymacyjnych rozwiązanych metodą EFI, tak i w eksperymencie EPR-Bohm’a wykorzystano przy estymacji kąta ϑ , informacyjną zasadę strukturalną $Q_{1_a} = -I_{1_a}$, (5.174). Ponieważ nie mierzony był stan cząstki 2, tzn. spin S_b w czasie naszej analizy mógł przyjąć wartość $+$ lub $-$, zatem pojemność informacyjna I_{1_a} jest informacją o kącie ϑ , zawartą w obserwacji (zgodnej z dokładnością do szumu, z rzutem) spinu cząstki 1, dokonanej przez zewnętrznego obserwatora¹¹.

¹¹Zaszumienie z aparatury Sterna-Gerlacha, o którym wspomniano powyżej, z uproszczenia pomijamy.

Zatem informacja strukturalna Q_{1_a} (5.172) dla cząstki 1, jako pozostała część informacji fizycznej K , jest informacją zawartą w nieobserwowanej wewnętrznej strukturze informacyjnej układu. Sugeruje to istnienie łącznej, *nie separowalnej* informacji strukturalnej dla cząstek 1 i 2. Słowem, informacja o łącznej strukturze układu, to nie to samo co suma informacji o jego składowych.

Zatem, ze względu na $Q_{1_a} = -I_{1_a}$, przestrzeń danych (tzn. obserwacje częstości rzutu spinu cząstki 1), jest na poziomie informacji, związana ze źródłową przestrzenią konfiguracyjną odbitą w Q_{1_a} . Co prawda, ze względu na brak bezpośredniego wglądu w relacje wewnątrz układu, możemy jedynie ze skończoną dokładnością (Rozdział 5.2.4) wnioskować o kącie ϑ , jednak sam fakt możliwości takiego wnioskowania określa sytuacja nazywana *splątaniem stanów* obu cząstek.

Wynik ten prowadzi do stwierdzenia, że wewnętrzna zasada informacyjna $I = -\kappa Q$, stosowana dla każdego rozważanego problemu EFI, opisuje (w przypadku braku separowalności I oraz Q na sumy pojemności informacyjnych oraz informacji strukturalnych właściwych dla podukładów) splątanie przestrzeni danych obserwowanych¹² z nieobserwowaną konfiguracją układu cząstek¹³. Postuluje to wykorzystywanie EFI jako formalizmu do predykcji wystąpienia stanów splątanych i to nie tylko w przypadku problemu EPR-Bohm'a. Tak więc informacja strukturalna Q jest informacją o „splątaniu” widocznym w korelacji danych przestrzeni pomiarowej (o wartościach w przestrzeni bazowej) z przestrzenią konfiguracyjną układu.

Na koniec wspomnijmy, że inne traktowanie informacji fizycznej K w [7] niż w obecnym skrypcie, sprawia, że wnioski tam otrzymane są inne¹⁴.

¹²Powyżej częstości rejestracji określonych rzutów spinu S_a .

¹³Powyżej kąt ϑ .

¹⁴Wnioski w pracy [7] są następujące: Przekaz [związanej] informacji J do cząstki obserwowanej wzbudza wartość spinu S_a do poziomu danej w próbce, zgodnie z końcowym prawem (5.219). Oznacza to, że przekaz informacji sprawia, że cząstka spełnia końcowe prawo (5.219) [opisujące zachowanie się] spinu. Przyczyną jest jakiś nieznan mechanizm oddziaływania, być może pewna „bogata i złożona struktura, która może odpowiadać na informację i kierować zgodnie z tym swoim własnym ruchem”. W ten sposób, przekaz [związanej] informacji J wiąże z sobą stany spinowe nieobserwowanej i obserwowanej cząstki. A zatem J (zgodnie z wymogiem) „nadaje postać” własnościom spinu nieobserwowanej cząstki.

Rozdział 6

Zakończenie

Celem skryptu była prezentacja uogólnienia MNW oraz zastosowania IF w jej części wnioskowania statystycznego dotyczącego estymacji niepartametrycznej metodą EFI zaproponowaną przez Friedena i Soffera do opisu zjawisk fizycznych i (ekono)fizycznych. Podstawowy przeprowadzenia takiej analizy statystycznego opisu zjawisk zostały wprowadzonych w latach 20 ubiegłego wieku przez Fishera. Początkowo Fisher wprowadził MNW poszerzoną o pojęcie IF w celu (jakiegoś) rozwiązania problemu estymacji punktowej i przedziałowej parametru rozkładu zmiennej losowej w sytuacji małej próby. Jego statystyczna metoda doboru modeli, rozwijana niezależnie od ówczesnie rozwijanych teorii fizycznych, okazała się jednak sięgać o wiele dalej, w obszar estymacji równań fizycznych teorii pola, przy czym okazało się, że wielkość (małej) próby jest w tej estymacji cechą charakterystyczną modeli. Zagadnieniu temu poświęcona jest główna część niniejszego skryptu.

Jesteśmy w punkcie, w którym można już dać pewne wstępne podsumowanie metody EFI jako procedury statystycznej budowania modeli, w tym modeli fizycznych i ekonofizycznych. Jest więc EFI metodą statystycznej estymacją równań ruchu teorii pola lub równań generujących rozkłady fizyki statystycznej. Gdy już szukane równanie ruchu wyestymowane metodą EFI zostało otrzymane, wtedy poszukiwanie rozkładu prawdopodobieństwa na analitycznych warstwach przestrzeni statystycznej \mathcal{S} , tzn. na (pod)rozmaitości w \mathcal{S} z metryką Rao-Fishera, może być dalej prowadzone. Może się zdarzyć, że funkcja łącznego rozkładu prawdopodobieństwa (*pdf*) jest analizowana przez EFI na całej położeniowo-pędowej przestrzeni fazowej układu¹ i to dla więcej niż jednego pola. Ma to np. miejsce, gdy obok pola fermionowego Diraca obecne jest również pole cechowania (Rozdział 4.3.2.1). Wtedy na warstwach rozkładów fermionowych, w ich pochodnych kowariantnych, pole cechowania (dla brzegowego *pdf* pola cechowania wyznaczonego z łącznego *pdf* wszystkich pól) musi być samo-spójnie wzięte pod uwagę.

Jednak sedno metody EFI jest zawarte w ogólnej postaci informacji fizycznej K zadanej przez równania (4.6) oraz (4.7) (lub (4.8)), która jest zarówno funkcją $\mathbf{q}_n(q_n(\mathbf{x}))$ jak i amplitud $q_n(\mathbf{x})$ (lub $\psi_n(\mathbf{x})$) wraz z ich pochodnymi. Przy tak ogólnym zrozumieniu K , różnorodność równań EFI jest konsekwencją różnych warunków wstępnych dyktowanych przez fizykę. Mogą się one wyrażać np. poprzez równania ciągłości (które same są wynikiem estymacji statystycznej [10], Rozdział 3.4), pewne symetrie (Rozdział 2.7.1) charakterystyczne dla zjawiska oraz warunki normalizacyjne. Wstępne założenia fizyczne mogą również wskazywać (poza ewentualnością nałożenia dodatkowych równań ciągłości) na wykorzystanie tylko wariacyjnej zasadę informacyjnej, jak

¹Gdyby na przykład chcieć wyznaczyć metodą EFI łączny rozkład energii i pędu cząstki gazu.

to ma miejsce w przypadku równania Kleina-Gordona dla pola skalarnego (Rozdział 4.3.1), bądź na obie zasady i ich samo-spójne rozwiązanie, jak to miało miejsce w pozostałych przypadkach.

Podsumowując, przeszliśmy w skrypcie przez trzy etapy rozbudowy działania MNW oraz pojęcia IF w statystyce. Na pierwszym, wstępnym etapie, przedstawiono zastosowanie MNW oraz pojęcia IF w statystyce (Rozdział 1), na drugim wprowadzono pojęcie entropii względnej i pojemności informacyjnej oraz wprowadzono strukturalną i wariacyjną zasadę informacyjną (Rozdział 2-3), a w trzecim przedstawiono wykorzystanie tych pojęć do wyprowadzenia podstawowych równań modeli fizycznych (Rozdziały 4-5, Dodatek), włączając w to przykład ekonofizycznego opisu zjawiska (Rozdział 5.1.4).

Rozdział 7

Dodatki

7.1 Dodatek: Zasada nieoznaczoności Heisenberga

Poniższe wyprowadzenie zasady nieoznaczoności Heisenberga zostało przedstawione w [7, 18]. Zasada Heisenberga stwierdza, że w określonej chwili czasu t nie można z dowolną dokładnością wyznaczyć jednocześnie położenia \mathbf{y} i pędu \mathbf{y}_p cząstki, tzn. że położenie i pęd cząstki są *rozmyte*, co możemy zapisać następująco:

$$\sigma_\theta^2 \cdot \sigma_p^2 \geq \left(\frac{\hbar}{2}\right)^2, \quad (7.1)$$

gdzie σ_θ^2 i σ_p^2 są wariancjami położenia i pędu cząstki względem ich wartości oczekiwanych, odpowiednio θ oraz θ_p .

Wyprowadzenie nierówności (7.1) dla pola rangi $N = 2$

Załóżmy, że dokonujemy estymacji tylko w jednym przestrzennym kanale informacyjnym przy założeniu, że pozostałe (czaso-przestrzenne) parametry rozkładu są znane. Powyższa relacja może być wyprowadzona przy odwołaniu się do własności informacji Fishera. Wartości położenia zmiennej \mathbf{Y} są *rozmyte* (bądź fluktuują) wokół jej wartości oczekiwanej θ , skąd wartości \mathbf{x} zmiennej odchyłek \mathbf{X} spełniają (jak zwykle w skrypcie), związek:

$$\mathbf{x} = \mathbf{y} - \theta. \quad (7.2)$$

Załóżmy, że jest spełniona nierówność Rao-Cramera (2.29):

$$\sigma_\theta^2 \cdot I_F(\theta) \geq 1, \quad \text{gdzie} \quad \sigma_\theta^2 \equiv E \left(\left(\hat{\theta}(y) - \theta \right)^2 \right), \quad (7.3)$$

gdzie $I_F(\theta)$ jest informacją Fishera parametru θ dla *pojedynczego pomiaru* (z powodu odwołania się do skalarnej wersji twierdzenia Rao-Cramera), która po skorzystaniu z postaci kinematycznej (3.40) w Rozdziale 3.3, ma zatem postać:

$$I_F = 4 \int_{\mathcal{X}} d\mathbf{x} \left(\frac{\partial q(\mathbf{x})}{\partial \mathbf{x}} \right)^2. \quad (7.4)$$

Rozważmy zespoloną amplitudę ψ , (3.42), dla pola rangi $N = 2$:

$$\psi(\mathbf{x}) = \psi_1(\mathbf{x}) = \frac{1}{\sqrt{2}} (q_1(\mathbf{x}) + i q_2(\mathbf{x})), \quad (7.5)$$

która ma jako część rzeczywistą i urojoną dwie rzeczywiste amplitudy q_i , $i = 1, 2$. Zespoloną funkcję $\psi(\mathbf{x})$ można zapisać poprzez jej transformatę Fouriera, przechodząc z reprezentacji położeniowej \mathbf{x} do pędowej \mathbf{p} :

$$\psi(\mathbf{x}) = \frac{1}{\sqrt{2\pi\hbar}} \int d\mathbf{p} \phi(\mathbf{p}) \exp(i\mathbf{p}\mathbf{x}), \quad (7.6)$$

gdzie \mathbf{p} jest rozmyciem (fluktuacją) pędu \mathbf{y}_p wokół jego wartości oczekiwanej θ_p , tzn.:

$$\mathbf{p} = \mathbf{y}_p - \theta_p, \quad (7.7)$$

a pęd cząstki jest mierzony w tej samej chwili co jego położenie.

Dla amplitudy $N = 2$ pojemność informacyjna I , (3.49), ma postać:

$$I = 8 \int d\mathbf{x} \frac{d\psi^*(\mathbf{x})}{d\mathbf{x}} \frac{d\psi(\mathbf{x})}{d\mathbf{x}} = 8 \int d\mathbf{x} \left| \frac{d\psi(\mathbf{x})}{d\mathbf{x}} \right|^2. \quad (7.8)$$

Przedstawiając ψ w postaci:

$$\psi(\mathbf{x}) = |\psi(\mathbf{x})| \exp(iS(\mathbf{x})) \quad (7.9)$$

i wykonując różniczkowanie:

$$\frac{d\psi(\mathbf{x})}{d\mathbf{x}} = \frac{d|\psi(\mathbf{x})|}{d\mathbf{x}} e^{iS(\mathbf{x})} + i|\psi(\mathbf{x})| e^{iS(\mathbf{x})} \frac{dS(\mathbf{x})}{d\mathbf{x}}. \quad (7.10)$$

możemy (7.8) przekształcić następująco:

$$\begin{aligned} I &= 8 \int d\mathbf{x} \left| \frac{d\psi(\mathbf{x})}{d\mathbf{x}} \right|^2 = 8 \int d\mathbf{x} \frac{d\psi^*(\mathbf{x})}{d\mathbf{x}} \frac{d\psi(\mathbf{x})}{d\mathbf{x}} = \\ &= 8 \int d\mathbf{x} \left(\frac{d|\psi(\mathbf{x})|}{d\mathbf{x}} e^{-iS(\mathbf{x})} - i|\psi(\mathbf{x})| e^{-iS(\mathbf{x})} \frac{dS(\mathbf{x})}{d\mathbf{x}} \right) \left(\frac{d|\psi(\mathbf{x})|}{d\mathbf{x}} e^{iS(\mathbf{x})} + i|\psi(\mathbf{x})| e^{iS(\mathbf{x})} \frac{dS(\mathbf{x})}{d\mathbf{x}} \right) = \\ &= 8 \int d\mathbf{x} \left[\left(\frac{d|\psi(\mathbf{x})|}{d\mathbf{x}} \right)^2 + i|\psi(\mathbf{x})| e^{iS(\mathbf{x})} \frac{d|\psi(\mathbf{x})|}{d\mathbf{x}} \frac{dS(\mathbf{x})}{d\mathbf{x}} - i|\psi(\mathbf{x})| e^{iS(\mathbf{x})} \frac{dS(\mathbf{x})}{d\mathbf{x}} \frac{d|\psi(\mathbf{x})|}{d\mathbf{x}} + \right. \\ &\quad \left. + |\psi(\mathbf{x})|^2 \left(\frac{dS(\mathbf{x})}{d\mathbf{x}} \right)^2 \right] = 8 \int d\mathbf{x} \left[\left(\frac{d|\psi(\mathbf{x})|}{d\mathbf{x}} \right)^2 + |\psi(\mathbf{x})|^2 \left(\frac{dS(\mathbf{x})}{d\mathbf{x}} \right)^2 \right]. \end{aligned} \quad (7.11)$$

Zajmijmy się teraz pierwszym składnikiem pod ostatnią całką w (7.11). Ponieważ norma amplitudy ψ jest równa:

$$|\psi| = \sqrt{\frac{1}{2} (q_1^2 + q_2^2)}, \quad (7.12)$$

zatem:

$$\left(\frac{d|\psi(\mathbf{x})|}{d\mathbf{x}} \right)^2 = \frac{1}{2} \left(\frac{d}{d\mathbf{x}} \sqrt{q_1^2 + q_2^2} \right)^2 = |q_1 = q_2 = q| = \left(\frac{dq}{d\mathbf{x}} \right)^2. \quad (7.13)$$

gdzie skorzystano z założenia, że $N = 2$ wymiarowa próba dla zmiennej odchyłeń \mathbf{X} jest prosta, skąd rozkłady $q_i(\mathbf{x})$, $i = 1, 2$, są takie same.

Podstawiając powyższy wynik do całki (7.11) i korzystając z (7.4), otrzymamy następującą równość:

$$\begin{aligned} I &= 8 \int d\mathbf{x} \left| \frac{d\psi(\mathbf{x})}{d\mathbf{x}} \right|^2 = 8 \int d\mathbf{x} \left[\left(\frac{d|\psi(\mathbf{x})|}{d\mathbf{x}} \right)^2 + |\psi(\mathbf{x})|^2 \left(\frac{dS(\mathbf{x})}{d\mathbf{x}} \right)^2 \right] = \\ &= 8 \int d\mathbf{x} \left(\frac{dq}{d\mathbf{x}} \right)^2 + 8 \int d\mathbf{x} |\psi(\mathbf{x})|^2 \left(\frac{dS(\mathbf{x})}{d\mathbf{x}} \right)^2 = 2I_F + 8E \left[\left(\frac{dS(\mathbf{x})}{d\mathbf{x}} \right)^2 \right] \end{aligned} \quad (7.14)$$

Z (7.14) oraz (7.8) wynika więc następująca nierówność:

$$2I_F \leq I \quad \text{lub} \quad I_F \leq 4 \int d\mathbf{x} \left| \frac{d\psi(\mathbf{x})}{d\mathbf{x}} \right|^2 . \quad (7.15)$$

Wykorzystując transformatę Fouriera (7.6), otrzymujemy z (7.15) po przejściu do reprezentacji pędowej:

$$I_F \leq \frac{4}{\hbar^2} \int d\mathbf{p} |\phi(\mathbf{p})|^2 \mathbf{p}^2 . \quad (7.16)$$

gdzie $|\phi(\mathbf{p})|^2$ jest brzegową gęstością prawdopodobieństwa $P(\mathbf{p})$ odchyień pędu. Zatem całka po prawej stronie (7.16) jest wartością oczekiwaną $E(\mathbf{p}^2)$ dla \mathbf{p}^2 :

$$I_F \leq \frac{4}{\hbar^2} E(\mathbf{p}^2) = \left(\frac{2}{\hbar} \right)^2 E[(\mathbf{y}_p - \theta_p)^2] \equiv \left(\frac{2}{\hbar} \right)^2 \sigma_p^2 . \quad (7.17)$$

gdzie σ_p^2 jest wariancją pędu cząstki a pierwsza równość wynika z tego, że \mathbf{p} jest odchyleniem (fluktuacją) pędu od jego wartości oczekiwanej θ_p , (7.7), (a przesunięcie o stałą wartości zmiennej nie zmienia jej wariancji).

Podstawiając powyższy wynik do nierówności Rao-Cramera, $\sigma_\theta^2 \cdot I_F(\theta) \geq 1$, (7.3), otrzymujemy (7.1):

$$\sigma_\theta^2 \cdot \sigma_p^2 \geq \left(\hbar/2 \right)^2 , \quad \text{c.n.d.} \quad . \quad (7.18)$$

7.2 Dodatek: Równanie Schrödingera

Równania Schrödingera wyprowadzimy jako nierelatywistyczną granicę równania Kleina-Gordona. Wyprowadzenie go jako nierelatywistycznej granicy równania Diraca byłoby merytorycznie bardziej uzasadnione [50], jednak celem poniższego wyprowadzenia jest zwrócenie uwagi na relatywistyczne pochodzenie spinu elektronu.

Rozpoczniemy od separacji zmiennych czasowych i przestrzennych występujących w równaniu (4.49) dla pola rangi $N = 2$, zapisując po lewej stronie wszystkie wyrazy zawierające pochodną czasową a po prawej pochodną po współrzędnych przestrzennych:

$$\begin{aligned} & \left(\hbar^2 \frac{\partial^2}{\partial t^2} + 2i\hbar e\phi \frac{\partial}{\partial t} + i\hbar e \frac{\partial \phi}{\partial t} - e^2 \phi^2 \right) \psi \\ & = \left(c^2 \hbar^2 \vec{\nabla}^2 - 2ie\hbar c (\vec{A} \cdot \vec{\nabla}) - ie\hbar c (\vec{\nabla} \cdot \vec{A}) - e^2 \vec{A}^2 - m^2 c^4 \right) \psi, \end{aligned} \quad (7.19)$$

z wyrażenia

$$\nabla \cdot (\vec{A}\psi) = \psi \nabla \cdot \vec{A} + \vec{A} \cdot \nabla \psi \quad (7.20)$$

Następnie skorzystajmy z nierelatywistycznej reprezentacji funkcji falowej [50]:

$$\psi(\mathbf{x}, t) = \tilde{\psi}(\mathbf{x}, t) e^{-imc^2 t/\hbar} \quad (7.21)$$

gdzie wydzieliliśmy z ψ wyraz zawierający energię spoczynkową mc^2 otrzymując nową funkcję falową $\tilde{\psi}$. Kolejnym krokiem jest zastosowanie (7.21) w (7.19). Różniczkując (7.21) po czasie otrzymujemy:

$$\frac{\partial}{\partial t} \psi = \left(\frac{\partial}{\partial t} \tilde{\psi} - \frac{imc^2}{\hbar} \tilde{\psi} \right) e^{-imc^2 t/\hbar}, \quad (7.22)$$

a po kolejnym różniczkowaniu:

$$\frac{\partial^2}{\partial t^2} \psi = \left(\frac{\partial^2}{\partial t^2} \tilde{\psi} - 2\frac{imc^2}{\hbar} \frac{\partial}{\partial t} \tilde{\psi} - \frac{m^2 c^4}{\hbar^2} \tilde{\psi} \right) e^{-imc^2 t/\hbar}. \quad (7.23)$$

Stosując nierelatywistyczne ($n.r$) przybliżenie:

$$\frac{E_{n.r}}{mc^2} \ll 1, \quad (7.24)$$

gdzie $E_{n.r}$ jest określone poprzez poniższe zagadnienie własne:

$$-\hbar^2 \frac{\partial^2}{\partial t^2} \tilde{\psi} = E_{n.r}^2 \tilde{\psi}, \quad (7.25)$$

możemy pominąć pierwszy wyraz po prawej stronie (7.23).

Dla Odwołajmy się do przybliżenia stałego, słabego potencjału, dla którego zachodzi:

$$e\phi \ll mc^2, \quad (7.26)$$

oraz

$$\frac{\partial \phi}{\partial t} = 0. \quad (7.27)$$

Następnie wstawiając (7.22) oraz (7.23) do (7.19), i wykorzystując (7.24), (7.26) oraz (7.27), otrzymujemy po przemnożeniu przez $-1/2mc^2$ i oznaczeniu $\tilde{\psi}$ jako ψ , równanie:

$$i\hbar \frac{\partial}{\partial t} \psi - e\phi\psi = -\frac{\hbar^2}{2m} \nabla^2 \psi + \frac{ie\hbar}{mc} \vec{A} \cdot \nabla \psi + \frac{ie\hbar}{2mc} (\nabla \cdot \vec{A}) \psi + \frac{e^2 \vec{A}^2}{2mc^2} \psi. \quad (7.28)$$

Jest to równanie Schrödingera dla nierelatywistycznej funkcji falowej ψ z potencjałem elektromagnetycznym \vec{A} oraz ϕ .

W przypadku zerowania się części wektorowej potencjału \vec{A} i dla niezerowego potencjału skalar-
nego ϕ , równanie to przyjmuje znaną postać:

$$i\hbar \frac{\partial}{\partial t} \psi = -\frac{\hbar^2}{2m} \nabla^2 \psi + e\phi\psi. \quad (7.29)$$

Pytanie: *Powiedz dlaczego powyższe wyprowadzenie równania Schrödingera z równania Kleina-Gordona, świadczy o relatywistycznym charakterze spinu elektronu?*

7.3 Dodatek: Rezultaty EFI dla elektrodynamiki Maxwella oraz teorii grawitacji

7.3.1 Dodatek: Pole cechowania Maxwella

Poniżej zaprezentujemy rezultat metody EFI otrzymany w zapisie Friedena-Soffera dla wyprowadzenia równań Maxwella (M). Punktem wyjścia jest pojemność informacyjna (3.53):

$$I = \sum_{n=1}^N \int d^4\mathbf{x} \frac{1}{p_n(\mathbf{x})} \sum_{\nu=0}^3 \left(\frac{\partial p_n(\mathbf{x})}{\partial \mathbf{x}_\nu} \frac{\partial p_n(\mathbf{x})}{\partial \mathbf{x}^\nu} \right).$$

Rozważmy równania ruchu Maxwella dla pola z amplitudą rzeczywistą q_n , $n = 1, 2, 3, 4$, tzn. ranga pola wynosi $N = 4$. Zakładamy, że pola cechowania są proporcjonalne do tych rzeczywistych amplitud [7]:

$$q_\nu(\mathbf{x}) = a A_\nu(\mathbf{x}), \quad \text{gdzie } \nu \equiv n - 1 = 0, 1, 2, 3, \quad (7.30)$$

gdzie a jest pewną stałą.

Definiujemy amplitudy $q^\nu(\mathbf{x})$ dualne do $q_\nu(\mathbf{x})$:

$$q^\nu(\mathbf{x}) \equiv \sum_{\mu=0}^3 \eta^{\nu\mu} q_\mu(\mathbf{x}) = a \sum_{\mu=0}^3 \eta^{\nu\mu} A_\mu(\mathbf{x}), \quad \text{gdzie } \nu \equiv n - 1 = 0, 1, 2, 3, \quad (7.31)$$

tak, że amplitudy $q_\nu(\mathbf{x})$ oraz $q^\nu(\mathbf{x})$ są związane z punktowymi rozkładami prawdopodobieństwa $p_n(\mathbf{x})$ następująco:

$$p_n(\mathbf{x}) \equiv p_{q_\nu}(\mathbf{x}) = q_\nu(\mathbf{x}) q^\nu(\mathbf{x}) = a^2 A_\nu(\mathbf{x}) A^\nu(\mathbf{x}), \quad \text{gdzie } \nu \equiv n - 1 = 0, 1, 2, 3. \quad (7.32)$$

Dalej postępujemy jak w [7]. Stosujemy obie zasady informacyjne, strukturalną (z $\kappa = 1/2$), (3.31), oraz wariacyjną (3.32):

$$I + \kappa Q = 0 \quad \text{gdzie } \kappa = 1/2, \quad \text{oraz } \delta_{(q)} K = 0, \quad (7.33)$$

które rozwiązujemy samo-spójnie, wraz z nałożonym dodatkowo warunkiem Lorentza:

$$\partial_\mu A^\mu = 0. \quad (7.34)$$

Zgodnie z (3.53) oraz (7.32) pojemność informacyjna I jest następująca:

$$I = 4 a^2 \int d^4x \partial_\mu A_\nu \partial^\mu A^\nu. \quad (7.35)$$

Właściwa postać informacji strukturalnej $Q \equiv Q_M$ dla równań ruchu Maxwella, została wyprowadzona w [7].

Warunki fizyczne: W celu otrzymania zarówno Q_M jak i wartości κ , muszą być przyjęte pewne fizyczne założenia [7]. *Po pierwsze* jest to warunek Lorentza i *po drugie* pewna wstępna postać Q_M . *Po trzecie*, wymagane jest również założenie o braku dodatkowych źródeł pola elektromagnetycznego w przestrzeni wolnej od cztero-wektora prądu $j^\mu = (c\rho, \vec{j})$.

Drugi z tych warunków, zapisany zgodnie z notacją zawartą w (4.7), wyraża się żądaniem następującej faktoryzacji obserwowanej informacji strukturalnej:

$$a^2 A_\nu A^\nu \mathbf{q}_\nu(A_\nu(x)) = \frac{a}{\kappa} A^\nu F_\nu(j_\nu), \quad (7.36)$$

która jest następnie wykorzystana w (7.33). O ile funkcja $F_\nu(j_\nu)$ spełnia założenie o zależności jedynie od j_ν [7], wtedy z zasad informacyjnych metody EFI, (7.33), wynika zarówno wartość współczynnika efektywności $\kappa = 1/2$ jak i równanie ciągłości strumienia $\partial^\nu j_\nu = 0$ [35]. W efekcie informacja strukturalna dla równania Maxwella jest równa [7]:

$$Q = Q_M \equiv -\frac{64\pi}{c} a \int d^4x j_\mu A^\mu. \quad (7.37)$$

W końcu, dla I jak w (7.35) oraz Q jak w (7.37), i dla $a = 2$, metoda EFI daje wektorowe równanie falowe w cechowaniu Lorentza:

$$\square A^\mu = (4\pi/c)j^\mu, \quad (7.38)$$

co w konsekwencji prowadzi do znanej postaci równań Maxwella dla pola magnetycznego i elektrycznego [61, 7].

Podkreślmy, że proporcjonalność $q_\nu(\mathbf{x}) = aA_\nu(\mathbf{x})$ oraz warunek normalizacji:

$$(1/4) \sum_{\nu=0}^3 \int d^4x q_\nu^2(\mathbf{x}) = \sum_{\nu=0}^3 \int d^4x A_\nu^2(\mathbf{x}) = 1, \quad (7.39)$$

stawiają pytanie o znaczenie lokalizacji fotonu oraz istnienie jego funkcji falowej, które są ostatnio mocno dyskutowane w literaturze dotyczącej optyki. Dyskusja zawarta w [36] wspiera głównie pogląd, który stawia równania Maxwella na tych samych podstawach co równanie Diraca w sformułowaniu pierwszej kwantyzacji. Fakt ten był wcześniej zauważony w pracy Sakurai [50]. Przyjmując interpretację funkcji falowej fotonu jako mającą te same podstawy co występująca w (4.23) funkcja falowa cząstek materialnych z masą $m = 0$, z trzeciego z powyższych *fizycznych warunków* można by zrezygnować [7].

Zauważmy, że normalizacja¹ (7.39) cztero-potencjału A_ν uzgadnia wartość stałej proporcjonalności² $a = 2$ z wartością $N = 4$ dla pola świetlnego.

Normalizacja (7.39), nałożona jako warunek na rozwiązanie równania (7.38), jest spójna z narzuceniem warunku początkowego Cauchy'ego we współrzędnej czasowej (por. dyskusja w [7]). Gdy jednocześnie we współrzędnych przestrzennych narzucony jest warunek Dirichlet'a (lub Neumann'a), wtedy te mieszane czasowo-przestrzenne warunki brzegowe nie są współzmiennicze, co skutkuje tym, że rozwiązanie równania (7.38) nie jest współzmiennicze.

Jednakże tylko z mieszanymi warunkami brzegowymi rozwiązanie to jest jednoznaczne [60, 7], co dla metody EFI jest warunkiem koniecznym, gdyż jest ona metodą estymacji statystycznej. Fakt ten stoi w opozycji do przypadku, gdy warunek Dirichlet'a (lub Neumann'a) jest nałożony współzmienniczo zarówno we współrzędnych przestrzennych jak i współrzędnej czasowej, gdyż co prawda otrzymane rozwiązanie jest wtedy współzmiennicze, jednak nie jest ono jednoznaczne.

¹Normalizacja cztero-potencjału A_ν zadana przez (7.39) do jedności mogłaby nie zachodzić [59]. Warunkiem koniecznym dla $q_\nu(\mathbf{x})$ jest aby niezbędne średnie mogły być wyliczone. Porównaj tekst poniżej (3.7).

²Otrzymanej wcześniej jako wynik uzgodnienia rezultatu metody EFI z równaniami Maxwella.

7.3.2 Dodatek: Metoda EFI dla teorii grawitacji

Poniżej przedstawimy jedynie główne wyniki związane z konstrukcją EFI dla teorii grawitacji. Wychodząc z ogólnej postaci pojemności (3.53) i postępując analogicznie jak powyżej dla pola Maxwella przy definicji amplitud dualnych, otrzymujemy amplitudową postać dla pojemności informacyjnej metody EFI. Następnie postępujemy już jak w [7], gdzie zostało podane wyprowadzenie słabej (tzn. falowej) granicy równań ruchu Einsteina, dla przypadku pól z rangą $N = 10$, a amplitudy $q_n(\mathbf{x}) \equiv q_{\nu\mu}(\mathbf{x})$ w liczbie dziesięć, są rzeczywiste i symetryczne w indeksach $\nu, \mu = 0, 1, 2, 3$. Zatem, pojemności informacyjna jest następująca:

$$I = 4 \int d^4\mathbf{x} \sum_{\nu, \mu=0}^3 \sum_{\gamma=0}^3 \frac{\partial q_{\nu\mu}(\mathbf{x})}{\partial \mathbf{x}_\gamma} \frac{\partial q^{\nu\mu}(\mathbf{x})}{\partial \mathbf{x}^\gamma}, \quad (7.40)$$

gdzie amplitudy dualne mają postać $q^{\delta\tau}(\mathbf{x}) = \sum_{\nu, \mu=0}^3 \eta^{\delta\nu} \eta^{\tau\mu} q_{\nu\mu}(\mathbf{x})$. Rozwiązując samo-spójne równania różniczkowe obu zasad informacyjnych, strukturalnej i wariacyjnej, wraz z nałożonym warunkiem Lorentza, $\sum_{\nu=0}^3 \partial_\nu q^{\nu\mu}(\mathbf{x}) = 0$, który redukuje współczynnik efektywności do wartości $\kappa = 1/2$, otrzymujemy następującą postać informacji strukturalnej:

$$Q = -8 \int d^4\mathbf{x} \frac{1}{L^4} \sum_{\nu, \mu=0}^3 \bar{h}_{\nu\mu} \left[\frac{16\pi G}{c^4} T^{\nu\mu} - 2\Lambda \eta^{\nu\mu} \right], \quad (7.41)$$

gdzie

$$\bar{h}_{\nu\mu}(\mathbf{x}) \equiv L^2 q_{\nu\mu}(\mathbf{x}), \quad (7.42)$$

natomiast $\eta_{\nu\mu}$ jest metryką Minkowskiego, G jest stałą grawitacyjną, $T_{\nu\mu}$ jest tensorem energii-pędu, Λ jest tzw. stałą kosmologiczną, a stała L jest charakterystyczną skalą, na której amplitudy $\bar{h}_{\nu\mu}$ są uśrednione.

Rozwiązanie metody EFI pojawia się w postaci równania falowego dla amplitud $\bar{h}_{\nu\mu}$:

$$\square \bar{h}_{\nu\mu} = \frac{16\pi G}{c^4} T_{\nu\mu} - 2\Lambda \eta_{\nu\mu}, \quad (7.43)$$

gdzie $\square \equiv \sum_{\nu=0}^3 \partial^\nu \partial_\nu = \sum_{\mu, \nu=0}^3 \eta^{\mu\nu} \partial_\mu \partial_\nu$ jest operatorem d'Alemberta. Równania (7.43) mają postać właściwą dla równań ruchu w granicy słabego pola w ogólnej teorii względności. Rezultatem EFI jest również równanie ciągłości strumienia, $\sum_{\nu=0}^3 \partial^\nu T_{\nu\mu}(\mathbf{x}) = 0$, dla tensora $T_{\nu\mu}$.

Pozostaje pytanie o rozwiązanie dla dowolnie silnego pola. Jedną z odpowiedzi jest następująca. Ponieważ $\square \bar{h}_{\nu\mu}$ jest jednoznaczną liniową aproksymacją tensora rangi drugiej $2R_{\nu\mu} - g_{\mu\nu}R$, gdzie $R_{\nu\mu}$ jest tensorem Ricci'ego [62], $g_{\nu\mu}$ jest tensorem metrycznym, a $R = \sum_{\nu=0}^3 R^\nu_\nu$, zatem można by uznać, że równanie Einsteina wyłania się jako jedyne możliwe, w tym sensie, że jego linearyzacją jest równanie falowe słabego pola (7.43) otrzymane w EFI [7].

Jednakże postać rozwiązania (7.43) metody EFI mogłaby również sugerować inną selekcję przyszłego modelu grawitacji. Otóż w całym formalizmie EFI amplitudy są podstawą do definicji pola, a nie metryki czaso-przestrzeni. Zatem bardziej naturalnym wydaje się zinterpretowanie $\bar{h}_{\nu\mu}$ jako pola grawitacyjnego³. Zatem równanie (7.43) metody EFI dla grawitacji leży bliżej

³A nie tzw. metryki słabego pola pochodzącej od liniowej części zaburzenia metryki. Metryka słabego pola ma postać: $\bar{h}_{\nu\mu} = h_{\nu\mu} - \frac{1}{2}\eta_{\mu\nu}h$, gdzie $h = \sum_{\mu, \nu=0}^3 \eta^{\mu\nu} h_{\mu\nu}$, oraz $h_{\mu\nu} = g_{\mu\nu} - \eta_{\mu\nu}$, dla $|h_{\mu\nu}| \ll 1$.

innej teorii grawitacji, nazywanej “szczególną teorią względności grawitacji”, która prowadzi do efektywnej teorii grawitacji typu Logunov’a [34], i jest poprzez swój widoczny związek z teoriami cechowania bardziej “obfita” niż sama ogólna teoria względności. Problem statystycznego porównania obu teorii grawitacji jest kwestią przyszłych prac.

7.4 Dodatek: Informacyjny odpowiednik drugiej zasady termodynamiki: Twierdzenie I. Temperatura Fishera

Niech próba będzie $N = 1$ -wymiarowa. Podobnie jak przechodzi się z (2.100) do (2.101) w celu otrzymania dyskretnej postaci informacji Fishera⁴, tak można też pokazać, że można przejść do jej następującej dyskretnej postaci:

$$I_F = \Delta x \sum_k \frac{1}{p(x_k)} \left[\frac{p(x_k + \Delta x) - p(x_k)}{\Delta x_k} \right]^2. \quad (7.46)$$

Postępując w sposób analogiczny jak w przypadku wyprowadzenia (2.106) z (2.101), otrzymujemy:

$$I_F = -\frac{2}{(\Delta x)^2} S(p(x) | p(x + \Delta x)). \quad (7.47)$$

Niech zmiana rozkładu prawdopodobieństwa z $p \equiv p(x)$ w chwili t do $p_{\Delta x} \equiv p(x + \Delta x)$ następuje na skutek infinitezymalnej zmiany czasu o dt . Jeśli więc dla rozkładów p oraz $p_{\Delta x}$ entropia względna spełnia warunek:

$$\frac{dS(p | p_{\Delta x})(t)}{dt} \geq 0, \quad (7.48)$$

wtedy ze względu na (7.47) informacja Fishera spełnia warunek:

$$\frac{dI_F(t)}{dt} \leq 0 \quad (7.49)$$

nazywamy I -twierdzeniem. Oznacza ono, że informacja Fishera I_F dla parametru θ maleje monotonicznie z czasem.

Uwaga: Dowód (7.48) oraz faktu, że relacja ta istotnie prowadzi do drugiej zasady termodynamiki w rozumieniu twierdzenia H , czyli dowód który trzeba przeprowadzić dla entropii Shannona, można znaleźć w pracach na temat dynamiki układów otwartych [64].

Wprowadzenie twierdzenia I jako mającego związek z twierdzeniem H jest jak widać nieprzypadkowe. Ale jest (również nieprzypadkowe) podobieństwo działające w drugą stronę. A mianowicie, pojemność informacyjna jest związana z rangą pola N , co było widoczne w całej treści skryptu. W istniejącej literaturze można znaleźć wyprowadzenia pokazujące, że istnieje entropijny odpowiednik zasady nieoznaczoności Heisenberga, oraz związek informacji Shannona z wymiarem

⁴Pojemność informacyjna (3.53) dla $N = 1$ i skalarnego parametru θ wynosi:

$$I_F = \int dx \frac{1}{p_\theta(x)} \left(\frac{\partial p_\theta(x)}{\partial x} \right)^2, \quad (7.44)$$

i jest to informacja Fishera parametru θ , gdzie informację o parametrze θ pozostawiono w indeksie rozkładu. Interesujący teraz związek informacji Fishera z entropią Kullbacka-Leiblera pojawia się na skutek zmiany rozkładu zmiennej losowej spowodowanego (nie zmianą parametru rozkładu ale) zmianą wartości x na $x + \Delta x$. Zastępujemy więc (7.44) sumą Riemanna (7.46).

Wprowadźmy wielkość:

$$\delta_{\Delta x} \equiv \frac{p_\theta(x_k + \Delta x)}{p_\theta(x_k)} - 1. \quad (7.45)$$

Postępując dalej podobnie jak przy przejściu od (2.102) do (2.106) (tyle, że teraz rozkłady różnią się z powodu zmiany wartości fluktuacji x), otrzymujemy (7.47).

reprezentacji grupy obrotów dla pól fermionowych rangi N .

Uwaga: Jako kolejną ilustrację (nieprzypadkowego) podobieństwa rachunków informacyjnych i entopijnych rozważmy definicję temperatury Fishera (dla) parametru θ . Otóż, informacja Fishera parametru θ pozwala na zdefiniowanie *temperatury Fishera* T_θ związanej z estymacją tego parametru:

$$\frac{1}{T_\theta} \equiv -k_\theta \frac{\partial I}{\partial \theta}. \quad (7.50)$$

Tak określona temperatura T_θ jest miarą czułości informacji Fishera na zmianę parametru θ , podobnie jak temperatura Boltzmanna T (dla której zachodzi $\frac{1}{T} \equiv \frac{\partial H_B}{\partial E}$) jest miarą czułości entropii Boltzmanna H_B na zmianę energii E układu. Temperatura Fishera znalazła swoje zastosowanie między innymi w badaniu rynków finansowych [65, 66].

Bibliografia

- [1] R.A. Fisher, Phil. Trans. R. Soc. Lond.**222**, 309 (1922). R.A. Fisher, "Statistical methods and scientific inference", 2 nd edn. (London, Oliver and Boyd, 1959).
- [2] C.R. Rao, Bulletin of the Calcutta Mathem. Soc. **37**, pp.81-91 (1945). "Information and accuracy attainable in the estimation of statistical parameters".
- [3] B. Efron, The Annals of Statistics **3**, pp.1189-1242 (1975). "Defining the curvature of statistical problem (with application to second order efficiency) (with discussion)".
- [4] A.P. Dawid, The Annals of Statistics **3**, pp.1231-1234 (1975). "Discussion to Efron's paper". A.P. Dawid, The Annals of Statistics **5**, pp.1249 (1977). "Further comments on a paper by Bradley Efron".
- [5] S. Amari, POST RAAG Report **106** (1980). "Teory of information spaces-a geometrical foundation of statistics". S. Amari, The Annals of Statistics **10**, pp.357-385 (1982). "Differential geometry of curved exponential families-curvature and information loss."
- [6] A. Fujiwara and H. Nagaoka, Phys. Lett. A **201**, pp.119-124 (1995). "Quantum Fisher metric and estimation for pure state models". A. Fujiwara and H. Nagaoka, Journal of Mathem. Phys. **40**, pp. 4227-4239 (1999).
An estimation theoretical characterization of coherent states". K. Matsumoto, Technical Report METR 96-09, Dept. of Math. Eng. and Inform. Phys, Univ. of Tokyo, (1996). A new approach to the Cramer-Rao type bound of the pure state model". K. Matsumoto, PhD thesis, Univ. of Tokyo, (1998). A geometrical approach to quantum estimation theory".
- [7] B.R. Frieden, Found.Phys.**16**, 883 (1986). B.R. Frieden, Phys.Rev.A**41**, 4265 (1990). B.R. Frieden, B.H. Soffer, Phys.Rev.E**52**, 2247 (1995). B.R. Frieden, Phys.Rev.A**66**, (2002). B.R. Frieden, A. Plastino, A.R. Plastino and B.H. Soffer, Phys.Rev.E**66**, 046128 (2002); Phys.Lett.A**304**, 73 (2002). B.R. Frieden, Science from Fisher information: A unification, (Cambridge University Press, 2004).
- [8] B. Roy Frieden, „Science from Fisher Information. A Unification”, Cambridge University Press, 2004.
- [9] J. Syska, "The Fisher information and quantum models-classical field theory, classical statistics similarity", Phys. Stat. Sol.(b), **244**, No.7, 2531-2537 (2007)/DOI 10.1002/pssb.200674646.

- [10] E.W. Piotrowski, J. Śładkowski, J. Syska, S. Zajęc, "The method of the likelihood and the Fisher information in the construction of physical models", *Phys. Stat. Sol.(b)*, **246**, No.5, 1033-1037, (2009)/DOI 10.1002/pssb.200881566, arXiv:physics/0811.3554.
- [11] R. Nowak, „Statystyka dla fizyków”, PWN, Warszawa 2002.
- [12] S. Amari, H. Nagaoka, *Methods of information geometry, translations of Mathematical monographs*, Vol.191, Oxford Univ. Press, (2000).
A. Fujiwara and H. Nagaoka, *Quantum Fisher metric and estimation for pure state models*, *Phys. Lett. A* **201**, 119-124, (1995). A. Fujiwara and H. Nagaoka, *An estimation theoretical characterization of coherent states*, *J. Math. Phys.* **40**, 4227-4239, (1999). K. Matsumoto, *A geometrical approach to quantum estimation theory*, PhD thesis, Univ. of Tokyo, (1998).
- [13] Y. Pawitan, *In all likelihood: Statistical modelling and inference using likelihood*, (Oxford Univ. Press, 2001).
- [14] D.G. Kleinbaum, L.L. Kupper, K.E. Muller, A. Nizam, *Applied Regression Analysis and Other Multivariable Method*, Duxbury Press, (1998).
- [15] D. Mroziakiewicz, „Analiza regresji Poissona z estymatorami metody największej wiarygodności z wykorzystaniem programu statystycznego SAS”, Praca licencjacka, Inst. Fizyki, Univ. Śląski (Rybnik), Polska, (2006).
- [16] D. Mroziakiewicz, "Metoda największej wiarygodności i informacja Fishera w fizyce i statystyce", Praca mgr., (*The maximum likelihood method and Fisher information in physics and statistics*), unpublished, Uniwersytet Śląski, Inst. Fizyki, (2008);
- [17] M. Czerwik, "Wykorzystanie programu SAS jako narzędzia do analizy współzależności zmiennych metodą regresji", Praca licencjacka, Uniwersytet Śląski, Inst. Fizyki, Jastrzębie Zdrój, (2004).
- [18] S. Mania, „Informacja Fishera. Wyprowadzenie równania Kleina-Gordona”, Praca licencjacka, Uniwersytet Śląski, Inst. Fizyki, Jastrzębie Zdrój, (2007).
- [19] R.F. Streater, “Statistical dynamics. A stochastic approach to nonequilibrium thermodynamics”, 2nd ed., Imperial College Press, (2009).
- [20] M.K. Murray, J.W. Rice, "Differential geometry and statistics", Chapman and Hall/CRC, (1993).
- [21] I. Bengtsson, K. Życzkowski, *Geometry of quantum states*, (Cambridge Univ. Press, 2006).
- [22] A.J. Stam, *Some mathematical properties of quantities of information*, Ph.D. dissertation. Technical University of Delf, (1959). A. J. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon", *Inform. and Control* **2**, 101–112 (1959).
- [23] F. Pennini, A. Plastino, B.H. Soffer, C. Vignat, *Physical symmetries and Fisher's information measure*, *Phys.Lett. A* **373**, pp.817-820, (2009).
- [24] I.N. Bronsztejn, K.A. Siemiendajew, G. Musiol, H. Mühlig, "Nowoczesne kompendium matematyki", 305-306 Wydawnictwo PWN, 305-306, (2004).

- [25] P.J. Brockwell, R.A. Davis , "Time Series: Theory and Methods", 2nd edition, Springer, (2009).
- [26] S. Zajac, "Modelowanie szeregów czasowych za pomocą procesów ARMA i ARIMA, (aplikacje w systemie SAS)", Praca licencjacka, Inst. Fizyki, Univ. Śląski (Rybnik), Polska, (2005).
- [27] J. von Neumann, *Mathematische Grundlagen der Quantenmechanik*, (Springer, Berlin), 171 (1932); transl. by E.T. Beyer, *Mathematical foundations of quantum mechanics*, (Princeton Univ. Press, Princeton), 324 (1955).
- [28] A. Peres, *Quantum Theory: Concepts and Methods*, (Kluwer Academic Publishers, 1995).
- [29] M.R. Esfandiary, H.J. Fink, *Phys.Lett.A***54**, 383 (1975). V.A. Schweigert, F.M. Peeters et.al., *Phys.Rev.Lett.***79**, 4653 (1997); *Phys.Rev.Lett.***83**, 2409 (1999); *Phys.Rev.B***59**, 6039 (1999); *Phys.Rev.B***62**, 9663 (2000). J.J. Palacios, *Phys.Rev.B***57**, 10873 (1998); *Phys.Rev.Lett.***83**, 2409 (1999); *Phys.Rev.Lett.***84**, 1796 (2000). G.F. Zharkov, V.G. Zharkov and A.Yu. Zvetov, *Phys.Rev.B***61**, 12293 (2000). G.F. Zharkov, *Phys.Rev.B***63**, 214502 (2001); *Phys.Rev.B***63**, 224513 (2001).
- [30] A.O. Barut, J.F. Van Huele, *Phys.Rev.A***32**, 3187 (1985). A.O. Barut and N. Ünal, *J.Math.Physics***27**, 3055 (1986). A.O. Barut, J.P. Dowling, *Phys.Rev.A***36**, 649 (1987). A.O. Barut and N. Ünal, *Physica* **142 A**, 467, 488 (1987). A.O. Barut, *Phys.Scr.***T21**, 18 (1988). A.O. Barut, Y.I. Salamin, *Phys.Rev.A***37**, 2284 (1988).
J. Syska, *Self-consistent classical fields in gauge field theories*, PhD thesis, (University of Silesia, 1995/99), (unpublished); *Trends in Boson Research*, e.d. A.V. Ling, (Nova Science Publishers), 163-181 (2006); *Frontiers in field theory*, ed. O. Kovras, (Nova Science Publishers), 125-154 (2005).
- [31] I. Bednarek, R. Mańka, *J.Phys.G* **31**, 1, (2005).
- [32] J. Syska, *Boson ground state fields in the classical counterpart of the electroweak theory with non-zero charge densities*, in *Frontiers in field theory*, ed. O. Kovras, (Nova Science Publishers), (Nowy Jork 2005), chapter 6, 125-154 and in *Focus on Boson Research*, (Nova Publishers), 233-258, (2006).
- [33] J. Syska, *Remarks on self-consistent models of a particle*, in *Trends in Boson Research*, ed. A.V. Ling, (Nova Science Publishers), 163-181, (2006).
- [34] V.I. Denisov and A.A. Logunov, *The Theory of Space-Time and Gravitation*, in *Gravitation and Elementary Particle Physics*, Physics Series, ed. A.A. Logunov, (MIR Publishers, Moscow), 14-130, (1983). A.A. Logunov, *The Theory of Gravity*, arXiv:gr-qc/0210005v2, (2002).
- [35] J. Syska, R. Szafron, "Informational criteria for the field theory models building", in preparation.
- [36] Edited by Ch. Roychoudhuri, A.F. Kracklauer, K. Creath, *The nature of light. What is a photon?*, (CRC Press, Taylor&Francis Group), (2008).

- [37] F.del Aguila, J. Syska, R. Szafron, S. Zając, M. Zralek, "The dynamics of the neutrino flavour transition with the general effective interaction", in preparation.
- [38] R. Ticciati, "Quantum field theory for mathematicians", (Cambridge University Press), (1999).
- [39] M. Fecko, "Differential geometry and Lie groups for physicists", (Cambridge University Press), (2006).
- [40] K. Sobczyk, "Stochastyczne równania różniczkowe. (Teoria i zastosowania)", Wydawnictwo NT, (1996); J. Łuczka, "Procesy i zjawiska losowe", Skrypt dla studentów ekonofizyki, Uniwersytet Śląski, Instytut Fizyki, (2010).
- [41] M. Aoki, H. Yoshikawa, "Reconstructing macroeconomics. A perspective from statistical physics and combinatorial stochastic processes, (Cambridge University Press), (2007).
- [42] U. Garibaldi, E. Scalas, "Finitary Probabilistic Methods in Econophysics", Cambridge Univ. Press, 1st ed., (2010).
- [43] J. Flachsmeier, "Kombinatoryka", PWN, (1977).
- [44] M. Abramowitz, I. A. Stegun. „Handbook of Mathematical Functions” Washington, DC.: National Bureau of Standards, 1965.
- [45] E. Kamke, „Sprawocznik po obykowniennym differencjalnym urawnieniam”, red. C.W. Fomin, Izdatielstwo inostranoj litieratury, Moskwa, 1951.
- [46] I.B. Rumer and M.Sh. Ryvkin, "Thermodynamics, Statistical Physics and Kinetics", Moscow: Mir Publishers, p.526, (1980).
- [47] E.B. Manoukian, *Quantum Theory A Wide Spectrum*, Springer, (2006).
- [48] J. Syska, "Kaluza-Klein Type Model for the Structure of the Neutral Particle-like Solutions", International Journal of Theoretical Physics, Volume 49, Issue 9, p.2131-2157, DOI:10.1007/s10773-010-0400-8, Open Access, (2010).
- [49] J-P. Gazeau, "Coherent states in quantum physics", Wiley-VCH, (2009).
- [50] J.J. Sakurai, *Advanced quantum mechanics*, pp.79-80,169, Addison-Wesley Publishing Company, (1967).
- [51] J.J. Sakurai, *Modern quantum mechanics*, ed. San Fu Tuan, (Addison-Wesley Publishing Company), (1994).
- [52] S.S. Afshar, *Violation of the principle of complementarity and its implications, in The nature of light. What is a photon?*, Edited by Ch. Roychoudhuri, A.F. Kracklauer, K. Creath, CRC Press, Taylor&Francis Group, pp.175-195, (2008). S.S. Afshar, <http://irims.org/quant-ph/030503/>, M. Chown in New Scientist, July 24, (2004).

- [53] X.Y. Zou, L.J. Wang and L. Mandel, Phys.Rev.Lett. **67**, p.318, (1991). U. Eichmann et al., Phys.Rev.Lett. **70**, p.2359, (1993). T.J. Herzog, P.G. Kwiat, H. Weinfurter and A. Zeilinger, Phys.Rev.Lett. **75**, p.3034, (1995). S. Durr, T. Nonn and D. Rempe, Nature **395**, p.33, (1998).
- [54] J.I. Kapusta, Ch. Gale, "Finite-Temperature Field Theory, Principles and applications", 2nd ed., (Cambridge Monographs on Mathematical Physics), Cambridge University Press, (2006).
- [55] H. Georgi, *Effective Field Theory*, Annual Review of Nuclear and Particle Science, Vol. **43**, pp.209-252, doi:10.1146/annurev.ns.43.120193.001233, (1993).
E. Castellani, *Reductionism, emergence, and effective field theories*, Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics, Vol.**33**, Issue 2, pp.251-267, doi:10.1016/S1355-2198(02)00003-5, (2002), arXiv:physics/0101039v1 [physics.hist-ph].
C.P. Burgess, *An Introduction to Effective Field Theory*, Ann.Rev.Nucl.Part.Sci. **57**, pp.329-362, doi:10.1146/annurev.nucl.56.080805.140508, (2007), arXiv:hep-th/0701053v2.
Z. Han, *Effective Theories and Electroweak Precision Constraints*, Int.J.Mod.Phys.A **23**, pp.2653-2685, doi:10.1142/S0217751X08041384, (2008), arXiv:0807.0490v1 [hep-ph].
- [56] M.O. Scully and U.W. Rathe, Opt.Comm. **110**, p.373, (1994). L. de la Peña and A.M. Catto, *The quantum dice*, Kluwer Academic, Dordrecht, (1996). A.F. Kracklauer, *Pilot wave steering: a mechanism and test*, Found.Phys.Lett.**13**, pp.441-453, (1999).
- [57] E.T. Jaynes, *Coherence and Quantum Optics*, ed. by L. Mandel and E. Wolf, (Plenum, New York), 495-509, (1978).
P.W. Milonni, *Foundations of Radiation Theory and Quantum Electrodynamics*, ed. by A.O. Barut, (Plenum, New York), 15, (1980).
- [58] A.F. Kracklauer, "Quantum ? beats in classical physics", Journal of Russian Laser Research C, Vol **26**, Nu.6, 524-529, (2005).
- [59] U. Leonhardt, *Essential quantum optics. From quantum measurements to black holes*, (Cambridge Univ.Press.), (2010).
- [60] P.M. Morse and H. Feshbach, *Methods of Theoretical Physics*, (New York: McGraw-Hill), (1953).
- [61] J. D. Jackson, "Elektrodynamika klasyczna", PWN, Warszawa, (1982).
- [62] C.W. Misner, K.S. Thorne, J.A. Wheeler, *Gravitation*, New York: Freeman, (1973).
- [63] H. Risken, "The fokker-Planck Equation:Methods of solution and Applications", New York: Springer-Verlag, 1996.
- [64] A. Wehrl, Rev. Mod. Phys. **50**, 221, (1978); R. Alicki and K. Lendi, "Quantum Dynamical Semigroups and Applications", Lect. Notes Phys. **717**, (Springer, Berlin Heidelberg), (2007), DOI 10.1007/b11976790.

- [65] E.W. Piotrowski, J. Sładkowski, “The thermodynamics of portfolios”, *Acta Physica Polonica B*, Vol.**32**, No.2, (2001); H. Kleinert, X.J. Chen, “Boltzmann distribution and market temperature”, arXiv:physics/0609209v2, (2007).
- [66] René Brand, “An econophysical investigation: Using the Boltzmann-distribution to determine market temperature as applied to the JSE All Share Index”, Stellenbosch University, (16 March 2009).