

UNIwersytet Kardynała Stefana Wyszyńskiego
w Warszawie

Wydział Matematyczno-Przyrodniczy
Szkoła Nauk Ścisłych

Katarzyna Jarnutowska

Nr albumu: 105162

Kierunek studiów: fizyka

Informacja Fishera w modelach fizycznych.

Praca licencjacka

Promotor: dr Sebastian Zając

Warszawa, 2017

Oświadczenie autora pracy

Świadoma odpowiedzialności prawnej oświadczam, że niniejsza praca dyplomowa została napisana przeze mnie samodzielnie i nie zawiera treści uzyskanych w sposób niezgodny z obowiązującymi przepisami.

Oświadczam również, że przedstawiona praca nie była wcześniej przedmiotem procedur związanych z uzyskaniem tytułu zawodowego w wyższej uczelni.

Oświadczam ponadto, że niniejsza wersja pracy jest identyczna z załączoną wersją elektroniczną.

Oświadczam, że poinformowano mnie o zasadach dotyczących kontroli samodzielności prac dyplomowych i zaliczeniowych. W związku z powyższym oświadczam, że wyrażam zgodę na przetwarzanie moich prac pisemnych (w tym prac zaliczeniowych i pracy dyplomowej) powstałych w toku studiów i związanych z realizacją programu kształcenia w Uczelni, a także na przechowywanie pracy dyplomowej w celach realizowanej procedury antyplagiatowej w ogólnopolskim repozytorium pisemnych prac dyplomowych.

Data

Podpis autora pracy

Oświadczenie

Oświadczam, że niniejsza praca napisana przez Panią Katarzynę Jarnutowską, nr albumu 105162 została przygotowana pod moim kierunkiem i stwierdzam, że spełnia ona warunki do przedstawienia jej w postępowaniu o nadanie tytułu zawodowego.

Data

Podpis kierującego pracą

Spis treści

Wykaz stosowanych oznaczeń	5
Wykaz stosowanych akronimów	6
Wstęp	7
Uzasadnienie wyboru tematu	7
Zakres pracy	7
Rozdział 1. Narzędzia rachunku prawdopodobieństwa i statystyki matematycznej	8
1.1. Typy zbiorów danych	8
1.2. Przestrzeń probabilistyczna	9
1.3. Zmienna losowa i jej rozkład	10
1.4. Funkcje zmiennej losowej	12
1.5. Charakterystyki liczbowe zmiennej losowej	13
1.5.1. Momenty	13
1.5.2. Wartość oczekiwana	14
1.5.3. Wariancja	14
1.5.4. Inne charakterystyki	15
1.5.5. Statystyka	15
1.6. Kowariancja zmiennych losowych	16
1.7. Rozkład normalny i twierdzenia graniczne	17
1.7.1. Rozkład normalny (jednowymiarowy oraz wielowymiarowy)	17
1.7.2. Nierówność Czebyszewa	18
1.7.3. Centralne twierdzenia graniczne (CTG)	18
1.8. Wnioskowanie statystyczne	19
1.8.1. Budowa modelu statystycznego	19
1.8.2. Teoria estymacji parametrycznej	20
1.8.3. Własności estymatorów EMNW oraz wybór najlepszego.	23
1.8.4. Podsumowanie	29
Rozdział 2. Wprowadzenie do podstawowych równań mechaniki kwantowej	30
2.1. Podstawowe równania mechaniki nierelatywistycznej i relatywistycznej	30
2.2. Wyprowadzenie równań mechaniki kwantowej	31
2.2.1. Równanie mechaniki nierelatywistycznej - równanie Schrödingera	31
2.2.2. Równanie mechaniki relatywistycznej - Równanie Kleina-Gordona i równanie Diraca	32

2.3.	Konsekwencje wynikające z równania Diraca	36
2.3.1.	Równanie Diraca przewiduje istnienie antycząstek	36
2.3.2.	Interpretacja ujemnych rozwiązań równania Diraca	37
2.4.	Podsumowanie	38
Rozdział 3. Metodyka EFI w teorii pola		39
3.1.	Strukturalne zasady informacyjne	39
3.2.	Zmodyfikowane zasady strukturalne	42
3.3.	Reprezentacje kinematyczne pojemności informacyjnej I	42
3.4.	Znaczenie rangi N i pojemności informacyjnej I	45
3.5.	Rozwiązania metody EFI - równania różniczkowe, jako analogia zasady wariacyjnej oraz równań różniczkowych Eulera-Lagrange'a	46
3.6.	Postać formy Q w teorii pola	47
3.7.	Omówienie rozwiązania zasad informacyjnych metody EFI dla modelu teorii pola - równanie Diraca	51
Rozdział 4. Podsumowanie		53
Dodatek A – Z historii fizyki		54
4.1.	Sylwetki postaci	54
	Ronald A. Fisher (1890 - 1964)	54
	Paul Dirac (1902 - 1984)	54
	B. Roy Frieden (1936 -)	55
Dodatek B – Aparat matematyczny		56
4.2.	Algebry	56
4.2.1.	Algebry Grassmanna	56
4.2.2.	Algebry Clifforda	56
4.3.	Komutator i antykomutator	56
4.4.	Czasoprzestrzeń Minkowskiego	57
4.4.1.	Iloczyn skalarny	58
4.4.2.	Czterowektor położenia	58
4.4.3.	Czterowektor prędkości	58
4.4.4.	Czterowektor energii i pędu	58
4.5.	Mechanika w ujęciu Lagrange'a	59
4.5.1.	Rachunek wariacyjny	59
4.5.2.	Równania Lagrange'a	59
Bibliografia		61

Wykaz stosowanych oznaczeń

Ω	przestrzeń probabilistyczna (zbiór zdarzeń elementarnych ω)
ω	zdarzenie elementarne (wynik doświadczenia)
\mathcal{F}	σ -ciało zbiorów
$(\Omega, \mathcal{F}, \mathbb{P})$	przestrzeń probabilistyczna
$\#A$	liczebność zbioru A
δ_{ij}	delta Kroneckera
$\mathbb{P}, \mathbb{P}_\theta$	funkcja prawdopodobieństwa
f, f_θ	funkcja gęstości prawdopodobieństwa
F_X	dystrybuanta rozkładu zmiennej losowej X
Φ	dystrybuanta rozkładu normalnego
$P(\theta), P(\Theta)$	funkcja wiarygodności wielkości skalarnej/wektorowej
θ, Θ	parametr (wielkość skalarne)/wektor parametrów modelu
\mathfrak{B}	przestrzeń prób z miarą $dyP(\Theta)$
\mathcal{S}	przestrzeń statystyczna z bazą $(\theta_n)_{n=1}^N$
\mathcal{P}	przestrzeń energetyczno-pędowa czteropędów \mathbf{p}
χ	przestrzeń próby w \mathbf{R}^n (rdz.1)/przestrzeń przesunięć czasoprzestrzeni (rdz.3)
Ψ	funkcja falowa cząstki
\mathcal{H}	energia cząstki (hamiltonian)
p_n	prawdopodobieństwo zmiennej losowej
q_n	amplituda prawdopodobieństwa zmiennej losowej
i	jednostka urojona, $i^2 = -1$
\hbar	stała Plancka (h) podzielona przez 2π
c	prędkość światła
m_0	masa spoczynkowa cząstki
e	ładunek elektronu
$\eta_{\mu\nu}$	metryka czasoprzestrzeni Minkowskiego
$I(\theta), I, \mathbf{i}$	informacja Fishera, pojemność informacyjna, gęstość I
\mathbf{iF}	obserwowana informacja Fishera
I_F	oczekiwana informacja Fishera
Q, \mathbf{q}	strukturalna informacja/gęstość SI
\mathbf{qF}	obserwowana macierz struktury
K	całkowita fizyczna informacja
D_μ	pochodna kowariantna

Wykaz stosowanych akronimów

EFI/EPI	metoda ekstremalnej fizycznej informacji/extreme physical information
TFI/TPI	całkowita (totalna) fizyczna informacja/total physical information
QED	elektrodynamika kwantowa (z ang. quantum electrodynamics)
QFT	kwantowe teorie pola (z ang. quantum field theory)
STW	szczególna teoria względności Einsteina
NRC	nierówność Rao-Cramera
DORC	dolne ograniczenie (nierówności) Rao-Cramera
CTG	centralne twierdzenie graniczne
MNW	metoda największej wiarygodności
ENMW	estymator nieobciążony o minimalnej wariancji
ENW	estymacja metodą największej wiarygodności
MNK	metoda najmniejszych kwadratów
EMM	estymacja metodą momentów
EMK	estymacja metodą kwantyli
SI	strukturalna informacja
TI	transfer informacji, gdzie J jest wielkością transferu informacji

Wstęp

Uzasadnienie wyboru tematu

Równanie Diraca (1928) jest ważnym równaniem teorii pola. Od momentu pojawienia się jest intensywnie badane pod względem własności i rozwiązań. W fizyce łączy mechanikę kwantową ze szczególną teorią względności (STW). Kwantowa teoria pola (QFT) po części wyrosła z rozważania falowych równań relatywistycznych, takich jak równania Maxwella, równanie Kleina-Gordona, czy równanie Diraca [21]. Współcześnie teorie pól kwantowych traktowane są jako teorie fizyczne, które jak dotąd dobrze tłumaczą większość oddziaływań podstawowych¹ (poza grawitacyjnymi).

Paul Dirac uważał, że nauka rozwija się poprzez poszukiwanie teorii coraz ściślejszych i że musimy się godzić z przybliżeniami [1].

Podejście amerykańskiego fizyka B. Roy Friedena opiera się z kolei na idei, że obserwacja nigdy nie jest całkowicie precyzyjna, a cała fizyka zjawiska zawarta jest w jego fluktuacjach. Opracował on (wraz z B.H. Sofferem) metodę konstrukcji modeli fizycznych z wykorzystaniem wprowadzonego przez siebie formalizmu ekstremalnej fizycznej informacji (EFI).

Zakres pracy

Celem pracy jest przedstawienie konstrukcji relatywistycznego równania Diraca (dla przypadku cząstki swobodnej) bazujące na statystycznym opisie zjawiska losowego zaproponowanym przez B. Roya Friedena. Podejście to, w myśl konstrukcji metody EFI, opiera się na wykorzystaniu funkcji wiarygodności $P(\Theta)$ do zbudowania metody największej wiarygodności (MNW), oraz informacji Fishera (IF). W kolejnym kroku MNW i IF zawarte w układzie posłużą do wyprowadzenia równań teorii pola.

W pierwszej części pracy omówione zostały metody wykorzystywane do zbudowania pojęć informacji Fishera (obserwowanej i oczekiwanej), nierówności Rao-Fishera, czy metody największej wiarygodności.

Część druga przedstawia historyczną konstrukcję równania Diraca.

Część trzecia pokazuje alternatywne podejście do konstrukcji równania pola przy wykorzystaniu podejścia zaproponowanego przez Friedena.

¹ Oddziaływania podstawowe to oddziaływanie grawitacyjne (z nośnikiem oddziaływania: grawiton), słabe (bozon W_{\pm} , bozon Z), elektromagnetyczne (foton) oraz silne (gluon).

Rozdział 1

Narzędzia rachunku prawdopodobieństwa i statystyki matematycznej

Informacja Fishera (IF) oraz metoda największej wiarygodności (MNW), jako narzędzia statystyki matematycznej i teorii prawdopodobieństwa, mają zastosowanie w wielu dyscyplinach naukowych. W ramach fizyki w mechanice i elektrodynamice kwantowej, fizyce statystycznej czy ekonofizyce.

Obie wielkości zostały sformułowane przez Ronalda Fishera na początku XX wieku w ramach klasycznej teorii statystyki matematycznej. Kluczowa jest tu miara informacji zawarta w rozkładzie prawdopodobieństwa cech (obserwacji) zjawiska fizycznego. O mierze niepewności zawartej w rozkładzie mogą decydować wartości osiągane przez informację Fishera i nierówność Rao-Cramera (NRC).

Opracowana przez niego metoda największej wiarygodności (MNW) oraz informacja Fishera (IF) stanowią narzędzia, użyte w pracy do wyprowadzenia relatywistycznego równania Diraca.

W tej części pracy dokonany zostanie przegląd zagadnień prawdopodobieństwa i statystyki niezbędnych do zrozumienia fundamentów zarówno IF, MNW jak i NRC.

Zacznę od przedstawienia typów zbiorów danych oraz wielkości obserwowanych, którymi dysponuje badacz. W kolejnym kroku opiszę pojęcie zmiennej losowej oraz funkcji zmiennej losowej i jej rozkładów - dyskretnych i ciągłych. Następnie przejdę do definicji momentów rozkładu oraz estymacji jego parametrów - podziałowi, własnościom, modelowi statystycznemu oraz funkcji wiarygodności. Ostatnim krokiem będzie zbudowanie wskaźnika informacji Fishera oraz przedstawienie nierówności Rao-Cramera. Wszystko to zostanie zobrazowane przykładami.

1.1. Typy zbiorów danych

Przy analizie zbiorów danych mamy do czynienia z dwoma typami danych - próbą przekrojową oraz szeregiem czasowym.

Próba przekrojowa powstaje jako wynik doświadczenia przeprowadzonego w danym momencie czasu dla pewnej grupy obiektów lub czynników wchodzących do badania. Jako *próbę* rozumiemy tutaj część populacji¹ poddanej badaniu, czyli taki zbiór jednostek, które wchodzi do badania. Próba jest więc skończony zbiór wszystkich wyników pomiarów.

¹ *Populacją* jest zbiór wszystkich przedstawicieli naznaczonych wybraną cechą.

Szereg czasowy zawiera obserwacje dotyczące jednego obiektu bądź wyniku doświadczenia w kolejnych chwilach czasu.

Metody statystyczne i predykcyjne, mające zastosowanie w przypadku prób przekrojowych, różnią się dość znacznie od tych, które stosuje się przy analizie szeregów czasowych. Praca ta skupia się jedynie na części dotyczącej analizy danych z prób przekrojowych.

Przy określaniu typu badania wyróżniamy badanie **pełne** oraz badanie **reprezentacyjne**. Na każde z nich składa się odpowiednio cała bądź część populacji, zbiorowości poddawanej badaniu, charakteryzująca się wyróżnionymi *cechami*. Jako cechę rozumiemy wielkość losową charakteryzującą obiekt danej populacji lub interesującą nas zmienną losową. Cecha może być *ilościową* i *jakościową* a jednostka badania stanowi element populacji poddanej badaniu. Najczęściej w badaniu musimy zadowolić się pewnymi częściowymi informacjami o rozkładzie i o populacji. Informacje te wydobywamy z próbki (stosując odpowiednie narzędzia) i aby były one wiarygodne, próbka musi być **losowa**, czyli **reprezentacyjna** [3, 6].

Definicja 1.1.1. (Próbka reprezentacyjna) *Próbka reprezentacyjna stanowi zminiaturyzowaną formę całej populacji, odzwierciedlającą wszystkie cechy i relacje, jakie w niej występują.*

Często mamy jednak do czynienia z **próbami obciążonymi**.

Definicja 1.1.2. (Próba obciążona) *Próba obciążona zakłada niedomiar przypadków, których nie udało nam się dostrzec z uwagi na ograniczenia sprzętowe lub myślowe.*

Próbkę nazywamy **prostą**, gdy wszystkie występujące w niej zmienne losowe są niezależne. *Losowanie zależne* możemy sobie wyobrazić jako losowanie bez zwracania. Uzyskana w takim losowaniu próba nie jest prosta [3].

Z kolei wnioskowanie o całej populacji, na podstawie próby losowej, wymaga od nas znajomości rachunku prawdopodobieństwa oraz modeli i metod klasycznej statystyki matematycznej.

1.2. Przestrzeń probabilistyczna

Trójka probabilistyczna $(\Omega, \mathcal{F}, \mathbb{P})$ zwana też *przestrzenią probabilistyczną* lub *przestrzenią statystyczną* to struktura umożliwiająca opis procesu losowego (czyli zdarzenia, którego wynik jest losowy) poprzez określenie przestrzeni zdarzeń elementarnych Ω (zbioru wszystkich możliwych wyników doświadczenia losowego) i określenie na jej podzbiorach funkcji prawdopodobieństwa \mathbb{P} spełniającej odpowiednie aksjomaty.

Konstrukcja przestrzeni probabilistycznej następuje w trzech etapach:

1. ustalenie niepustego zbioru Ω zwanego przestrzenią zdarzeń elementarnych,
2. określenie na nim σ -ciała \mathcal{F} zwanego przestrzenią zdarzeń losowych (zdarzeniem losowym jest mierzalny podzbiór \mathbf{A} zbioru zdarzeń elementarnych ω danego doświadczenia losowego),

3. określenie na \mathcal{F} unormowanej miary probabilistycznej \mathbb{P} .

Definicja 1.2.1. (σ -ciało \mathcal{F}) σ -ciałem \mathcal{F} nazywamy zbiór zdarzeń losowych utworzonych na przestrzeni zdarzeń elementarnych Ω .

Opis doświadczenia losowego następuje poprzez zdefiniowanie przestrzeni probabilistycznej. Dowolność wyboru σ -ciała \mathcal{F} może wyglądać następująco:

Przykład 1.2.1. (Rzut kostką) Eksperyment losowy polega na rzuceniu sześcienną kostką do gry. Wtedy zbiór zdarzeń losowych ma postać:

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

Jednak σ -ciało \mathcal{F} nie jest z góry określone. Możemy wybrać różne σ -ciała zdarzeń losowych na przykład:

$$\mathcal{F}_1 = \{\emptyset, \Omega\},$$

$$\mathcal{F}_2 = \{\emptyset, \Omega, \{1\}, \{2, 3, 4, 5, 6\}\},$$

$\mathcal{F}_3 = 2^\Omega$ - to σ -ciało tworzące rodzinę wszystkich podzbiorów Ω , to znaczy dowolny podzbiór zbioru Ω należy do σ -ciała.

Wszystkie wybory są dopuszczalne i jednakowo uprawnione. Wybór podyktowany jest postawionym problemem (liczba oczek przy rzucie kostką, suma oczek wyrzuconych w kilku rzutach kostką lub inne), na który szukana jest odpowiedź.

1.3. Zmienna losowa i jej rozkład

Przy opisie doświadczenia można mówić o dyskretnym bądź ciągłym charakterze danego procesu losowego lub zbiorze jego możliwych rezultatów, czyli odpowiednio o **dyskretnej (ciągłej) zmiennej losowej** podlegającej **dyskretnemu² (ciągłemu³) rozkładowi prawdopodobieństwa**. Z wynikiem doświadczenia losowego wiąże się pewną liczbę. Takie przyporządkowanie to funkcja określona na zbiorze Ω i o wartościach w zbiorze liczb rzeczywistych.

Definicja 1.3.1. (Zmienna losowa) Zmienną losową nazywamy funkcję o wartościach rzeczywistych, określoną na przestrzeni próbkowej. Na funkcję

$$X : \Omega \rightarrow \chi \subset \mathbb{R} \tag{1.1}$$

patrzemy jak na "zmienną, której wartość zależy od wyniku doświadczenia losowego": wynikowi $\omega \in \Omega$ odpowiada wartość $X(\omega) \in \chi$. Zbiorem wartości zmiennej losowej może być na przykład: $\chi = \{0, 1, 2, \dots\}$ lub $\chi = [0, \infty]$ lub $\chi = [a, b]$ lub $\chi = \mathbb{R}$ [5].

Przykład 1.3.1. (Rzut kostką) Rzucamy raz kostką do gry. $\Omega = \{1, 2, 3, 4, 5, 6\}$. Jeśli X jest liczbą wyrzuconych oczek to: $X(i) = i$ dla $i = 1, 2, 3, 4, 5, 6$.

² Przykładem rozkładu dyskretnego są rozkład zero-jedynkowy, dwumianowy, jednostajny dyskretny, geometryczny, hipergeometryczny, Pascala (ujemny dwumianowy), Poissona i inne.

³ Przykładem rozkładu ciągłego są rozkład beta, Cauchy'ego, chi kwadrat, wykładniczy, F, gamma, Laplace'a, log normalny, normalny, Pareto, t-Studenta, jednostajny ciągły, Weibulla, Dirichleta i inne.

Przypisanie liczb poszczególnym zdarzeniom elementarnym jest dowolne, dlatego w jednym doświadczeniu mogą pojawić się różne odwzorowania $X : \Omega \rightarrow \mathbb{R}$, przypisujące każdemu wynikowi doświadczenia ω liczbę rzeczywistą. Jeżeli mamy wielu obserwatorów zjawiska fizycznego, którzy sczytują wartość obserwowanej wielkości, to wynik każdego z nich może być różny. Jest więc wiele różnych zmiennych losowych określonych na tej samej przestrzeni Ω , co prowadzi do odwzorowania $X : \Omega \rightarrow \mathbb{R}^n$, gdzie n jest liczbą obserwatorów.

Zmienne losowe są funkcjami, na których możemy dokonywać tych samych operacji co i na zwykłych funkcjach takich jak dodawanie, mnożenie i inne (na przykład $e^X, \ln X, |X|$) - co nie wprowadza nowo określonych wielkości z klasy zmiennych losowych.

Definicja 1.3.2. (Rozkład prawdopodobieństwa) *Rozkładem prawdopodobieństwa \mathbb{P} zmiennej losowej X jest funkcja, która zbiorowi $B \in \chi$ przyporządkowuje liczbę (za [5])*

$$\mathbb{P}(X \in B) = \mathbb{P}(\omega \in \Omega : X(\omega) \in B).$$

Aksjomaty, które spełnia funkcja prawdopodobieństwa \mathbb{P} to:

- $\mathbb{P}(A) \in [0, 1]$,
 - $\mathbb{P}(\Omega) = 1$,
 - jeżeli zdarzenia A_i wykluczają się parami, to: $\mathbb{P}(A_1 \cup A_2 \dots) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \dots$.
- Rozkłady mogą być dyskretne lub ciągłe.

Definicja 1.3.3. (Rozkład dyskretny) *Zmienna losowa X ma **rozkład dyskretny**, jeśli zbiór jej wartości jest skończony lub przeliczalny. Dyskretny rozkład prawdopodobieństwa można opisać podając, dla każdego $x \in \chi$, prawdopodobieństwo [5]*

$$f(x) = \mathbb{P}(X = x).$$

Przykład 1.3.2. (Rzut kostką) *X jest wynikiem rzutu symetryczną kostką do gry. Wtedy ⁴*

$$\mathbb{P}(X = i) = \frac{1}{6} = \sum_{i=1}^6 \frac{1}{6} \mathbb{I}_{A(i)}, \quad i = 1, 2, 3, 4, 5, 6.$$

Definicja 1.3.4. (Rozkład ciągły) *Jeśli istnieje funkcja $f : \mathbb{R} \rightarrow \mathbb{R}^+$ taka, że dla dowolnych $a < b$ mamy*

$$\mathbb{P}(a < X < b) = \mathbb{P}(a \leq X \leq b) = \int_a^b f(x) dx,$$

*to zmienna losowa X ma **rozkład ciągły** z gęstością prawdopodobieństwa f [5].*

Jeśli $X : \Omega \rightarrow \chi \subset \mathbb{R}$, to przyjmujemy, że $f(x) = 0$ dla $x \notin \chi$. Dla zmiennych losowych o wartościach w \mathbb{R} można zdefiniować funkcję rzeczywistą, która jednoznacznie wyznacza

⁴ Jeśli $A \in \Omega$ jest zdarzeniem, to odpowiada mu zmienna losowa $\mathbb{I}_A : \Omega \rightarrow \mathbb{R}$ zwana indykátorem zdarzenia A określona przez

$$\mathbb{I}_A(\omega) = \begin{cases} 1, & A \in \Omega, \\ 0, & \text{w pozostałych przypadkach.} \end{cases}$$

rozkład zmiennej losowej - *dystrybuantę*, zawierającą pełną informację o rozkładzie. Dystrybuanta dyskretnego rozkładu zmiennej losowej X jest to funkcja $F : \mathbb{R} \rightarrow \mathbb{R}$ określona wzorem

$$F(a) = \mathbb{P}(X \leq a) = \mathbb{P}(\omega : X(\omega) \leq a).$$

Dystrybuanta rozkładu absolutnie ciągłego jest funkcją ciągłą i jest dana wzorem

$$F(a) = \mathbb{P}(X \leq a) = \int_{-\infty}^a f(x)dx.$$

W punktach, w których gęstość f jest ciągła, mamy $f(x) = \frac{d}{dx}F(x)$ [5].

Przykład 1.3.3. (Pomiar z błędem losowym) *Dokonujemy pomiaru pewnej wielkości fizycznej. Prawdziwą wartość mierzonej wielkości jest liczba μ . Niedoskonałość przyrządu pomiarowego powoduje, że wynik pomiaru podlega pewnym losowym wahaniom i nie jest dokładnie równy μ , choć leży "na ogół blisko" tej liczby. Przyjmuje się, że wynik pomiaru ma normalny rozkład prawdopodobieństwa (zwany też rozkładem Gaussa). Niech $\Omega = \mathbb{R}$. Określamy wówczas prawdopodobieństwo wzorem [6]*

$$\mathbb{P}([a, b]) = \int_a^b \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx.$$

Przykład 1.3.4. (Rozkład normalny) *Rozkład normalny z parametrami μ i σ oznaczony jako $N(\mu, \sigma^2)$ ma gęstość*

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \quad (1.2)$$

Wynik pomiaru jest wartością zmiennej losowej $X \sim N(\mu, \sigma^2)$. W szczególnym przypadku $N(0, 1)$ to standardowy rozkład normalny. Dystrybuantę tego rozkładu oznaczamy jako Φ i przyjmuje ona wartość:

$$\Phi(z) = \int_{-\infty}^z \varphi(x)dx, \quad \varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

Rozkład $N(\mu, \sigma^2)$ ma dystrybuantę $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$ [5].

Definicja 1.3.5. (Rozkłady mieszane) *Istnieją także rozkłady, które nie są ani dyskretne, ani ciągłe a mogą stanowić mieszaną ich obu [5].*

Rozkład prawdopodobieństwa leżący u podstaw badanej cechy nie jest z reguły znany. Zadaniem analiz statystycznych jest uzyskanie postaci rozkładu poprzez wypracowane metody osiągnięcia tego celu.

1.4. Funkcje zmiennej losowej

Przekształcenia zmiennych losowych

Rozważaniom poddana jest zmienną losową X i funkcja $h : \mathbb{R} \rightarrow \mathbb{R}$. Można określić nową zmienną losową $Y = h(X)$ (będącą funkcją zmiennej losowej X), kładąc $Y(\omega) = h(X(\omega))$. Dystrybuantę zmiennej X oznacza się standardowo jako F_X , gęstość zaś jako f_X [5].

Zobrazuję to przykładem.

Przykład 1.4.1. (Kwadrat zmiennej normalnej) Chcąc wyznaczyć rozkład funkcji $Y = X^2$, mając dane $X \sim N(0, 1)$ w pierwszym kroku wyznaczam dystrybuantę, następnie gęstość nowej funkcji Y . Proces przebiega następująco:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(X^2 \leq y) = \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) = 2\Phi(\sqrt{y}) - 1$$

$$f_Y(y) = \frac{d}{dy}(2\Phi(\sqrt{y}) - 1) = 2\varphi(\sqrt{y}) \frac{1}{2\sqrt{y}} = \frac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2}}, \text{ dla } y > 0.$$

Otrzymaliśmy rozkład chi-kwadrat z jednym stopniem swobody: $X^2 \sim \chi^2(1)$.

1.5. Charakterystyki liczbowe zmiennej losowej

Rachunek prawdopodobieństwa definiuje dwa podstawowe typy momentów zmiennej losowej X - zwykły oraz centralny (jest też moment absolutny, którym rozpocznę definiowanie momentów). Wartość oczekiwana i wariancja są szczególnymi przypadkami momentów, o czym jeszcze wspomnę w dalszej części opracowania.

1.5.1. Momenty

Definicja 1.5.1. (Moment absolutny) Momentem absolutnym rzędu $n > 0$ nazywamy wartość oczekiwaną funkcji liczbę $h(k) = |k|^n$ dla zmiennej losowej dyskretnej (dla której $\mathbb{P}_k = \frac{\#A_k}{\#\Omega}$ a $\#$ oznacza liczebność zbioru):

$$\langle |k|^n \rangle = \sum_k |k|^n \mathbb{P}_k \quad (1.3)$$

i funkcji $h(x) = |x|^n$ dla zmiennej losowej ciągłej:

$$\langle |x|^n \rangle = \int_{-\infty}^{+\infty} |x|^n f(x) dx. \quad (1.4)$$

Definicja 1.5.2. (Momenty zwykłe) Momentem m_n rzędu n nazywamy wartość oczekiwaną funkcji $h(k) = k^n$ dla zmiennej losowej dyskretnej:

$$m_n \equiv \langle k^n \rangle = \sum_k k^n \mathbb{P}_k \quad (1.5)$$

i funkcji $h(x) = x^n$ dla zmiennej losowej ciągłej:

$$m_n \equiv \langle x^n \rangle = \int_{-\infty}^{+\infty} x^n f(x) dx. \quad (1.6)$$

Definicja 1.5.3. (Momenty centralne) Momentem centralnym μ_n rzędu n nazywamy wielkość

$$\mu_n \equiv \langle (k - \langle k \rangle)^n \rangle = \sum_k (k - \mu_k)^n \mathbb{P}_k \quad (1.7)$$

oraz

$$\mu_n \equiv \langle (x - \langle x \rangle)^n \rangle = \int_{-\infty}^{+\infty} (x - \mu_x)^n f(x) dx. \quad (1.8)$$

odpowiednio dla zmiennej dyskretnej i ciągłej.

Wartość oczekiwana to pierwszy moment zwykły m_1 . Wariancja jest z kolei drugim momentem centralnym μ_2 . Wartość trzeciego momentu centralnego pozwala wnioskować o asymetrii rozkładu empirycznego, a czwarty wykorzystany jest do wyznaczenia kurtozy (miary spłaszczenia rozkładu wartości cechy).

1.5.2. Wartość oczekiwana

Definicja 1.5.4. (Wartość oczekiwana dyskretnej zmiennej losowej) *Jeśli $h(k)$ jest losową funkcją dyskretnej zmiennej losowej k o rozkładzie prawdopodobieństwa \mathbb{P}_k , to wartość oczekiwana funkcji $h(k)$ przyjmuje postać*

$$\mathbb{E}[h(k)] \equiv \langle h(k) \rangle \equiv \sum_k h(k) \mathbb{P}_k, \quad (1.9)$$

sumowanie przebiega po całym zakresie zmiennej losowej.

Definicja 1.5.5. (Wartość oczekiwana ciągłej zmiennej losowej) *Dla funkcji $h(x)$ ciągłej zmiennej losowej x , podlegającej rozkładowi $f(x)$, wartością oczekiwaną funkcji $h(x)$ nazywamy wielkość*

$$\mathbb{E}[h(x)] \equiv \langle h(x) \rangle \equiv \int_{-\infty}^{+\infty} h(x) f(x) dx, \quad (1.10)$$

gdzie przez nieskończone granice całkowania umownie zaznaczono cały zakres zmienności zmiennej losowej x .

Wartość oczekiwana charakteryzuje bezpośrednio zmienną losową, a tylko pośrednio rozkład.

Twierdzenie 1.5.1. (Własności wartości oczekiwanej) *Przy założeniu, że istnieją wartości średnie $\mathbb{E}X$ i $\mathbb{E}Y$ mamy:*

1. $\mathbb{E}X \geq 0$ jeśli $X \geq 0$,
2. $\mathbb{E}X \leq \mathbb{E}Y$ jeśli $X \leq Y$,
3. $|\mathbb{E}X| \leq \mathbb{E}|X|$,
4. $\mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y$,
5. $\mathbb{E}X = \mathbb{P}(A)$ jeśli $X = \mathbb{I}_A$.

Przykład 1.5.1. (Rozkład normalny) *Jeśli $X \sim N(\mu, \sigma^2)$, to $\mathbb{E}(X) = \mu$. Wynika to chociażby z symetrii gęstości: $f(\mu + x) = f(\mu - x)$.*

1.5.3. Wariancja

Definicja 1.5.6. *Dla wariancji (momentu centralnego rzędu drugiego) stosowana jest seria oznaczeń:*

$$\text{Var}[k] \equiv \mathbb{D}^2[k] \equiv \sigma_k^2 \equiv \langle (k - \langle k \rangle)^2 \rangle = \sum_k (k - m_1)^2 \mathbb{P}_k = \langle k^2 \rangle - \langle k \rangle^2, \quad (1.11)$$

$$\text{Var}[x] \equiv \mathbb{D}^2[x] \equiv \sigma_x^2 \equiv \langle (x - \langle x \rangle)^2 \rangle = \int_{-\infty}^{+\infty} (x - m_1)^2 f(x) dx = \langle x^2 \rangle - \langle x \rangle^2. \quad (1.12)$$

Co możemy zapisać jako:

$$\text{Var}X = \mathbb{E}X^2 - \mathbb{E}^2X. \quad (1.13)$$

Termin wariancji σ_x^2 został wprowadzony przez R.A. Fishera [6]. Wariancja ma wymiar kwadratu zmiennej losowej. Wymiar zmiennej losowej daje pierwiastek kwadratowy z wariancji zwane odchyleniem standardowym⁵.

Wariancja jest klasyczną miarą zmienności, czyli rozproszenia wartości cechy wokół wartości centralnych (jej wartości oczekiwanej lub średniej). Aby istniała wariancja, trzeba założyć, że $\mathbb{E}X^2 < \infty$.

Twierdzenie 1.5.2. (Własności wariancji) *Jeśli X jest zmienną losową, dla której $\mathbb{E}X^2 < \infty$, to istnieje $\mathbb{D}^2X (\equiv \text{Var}X)$ oraz (za [5]):*

1. $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2$,
2. $\text{Var}(cX) = c^2\text{Var}(X)$ dla dowolnej liczby c ,
3. $\text{Var}(a + X) = \text{Var}(X)$ dla dowolnej liczby a ,
4. $\text{Var}X = 0$ wtedy i tylko wtedy, gdy $X = \text{const}$.

Przykład 1.5.2. (Rozkład normalny) *Jeśli $Z \sim N(0, 1)$, to $\mathbb{E}Z = 0$. Stąd*

$$\text{Var}Z = \mathbb{E}Z^2 - (\mathbb{E}Z)^2 = \mathbb{E}Z^2 = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} z^2 \exp\left(-\frac{z^2}{2}\right) dz = 1$$

Powyższą całkę oblicza się przez części.

Można też sprawdzić, że zmienna $X = \mu + \sigma Z$ ma rozkład $N(\mu, \sigma^2)$. Z własności wariancji wynika, że $\text{Var}X = \sigma^2$, a z własności wartości oczekiwanej $\mathbb{E}X = \mu$.

1.5.4. Inne charakterystyki

Wyższe momenty (rzędu $n > 2$) wykorzystuje się przy budowaniu współczynników do oceny charakterystyk rozkładu zmiennych losowych. Za pomocą skośności, kurtozy i innych badamy koncentrację rozkładu wokół wartości średniej, spłaszczenie czy asymetrię rozkładu.

1.5.5. Statystyka

Statystyka jest funkcją zmiennych losowych. Jest to funkcja mierzalna służąca do wyodrębnienia pewnych istotnych cech danych doświadczalnych. Statystyka jest miarą rozkładu, czyli liczbową charakterystyką definiującą rozkład. Statystyka może być *klasyczna* (której wartości zależą od wszystkich zmiennych losowych wchodzących w skład próby na przykład odchylenie standardowe i wariancja) lub *pozycyjna* (której wartości zależą tylko od niektórych zmiennych losowych wchodzących w skład próby, głównie od tych, które zajmują odpowiednią pozycję w próbie na przykład: mediana, kwantyl, dominanta, rozstęp, decyl, centyl i inne). Statystyka jest też estymatorem parametru rozkładu zmiennej losowej.

Twierdzenie 1.5.3. (Statystyka) *Niech $(\Omega, \mathcal{F}, \mathbb{P})$ będzie przestrzenią statystyczną, gdzie*

$$\mathbb{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$$

⁵ Określenie odchylenia standardowego zostało wprowadzone za K. Pearsonem. W literaturze funkcjonuje termin *dyspersja*, dla określenia *fluktuacji*, *rozrzutu* i *nieokreśloności* [6].

jest rodziną miar probabilistycznych określonych na σ -ciele \mathcal{F} podzbiorów zbioru Ω , indeksowaną parametrem θ . (χ, C) jest przestrzenią mierzalną. Funkcję mierzalną $T : \Omega \rightarrow \chi$ nazywamy statystyką.

Jeśli $\chi = \mathbb{R}$ to statystykę T nazywamy statystyką o wartościach rzeczywistych. Jeśli $\chi = \mathbb{R}^n$ to statystykę T nazywamy statystyką o wartościach wektorowych [3].

Przykład 1.5.3. (Najważniejsze statystyki) Załóżmy, że X_1, \dots, X_n jest próbką z rozkładu $N(\mu, \sigma^2)$. Ważną rolę w rachunkach statystycznych odgrywają statystyki (za [3]):

$$\text{średnia z próby: } \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad (1.14)$$

$$\text{wariancja z próby (nieobciążona): } S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \quad S = \sqrt{S^2}, \quad (1.15)$$

$$\text{wariancja z próby (obciążona): } \tilde{S}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2, \quad (1.16)$$

$$n\tilde{S}^2 = (n-1)S^2. \quad (1.17)$$

Przykład 1.5.4. (Rozkłady statystyk rozkładu normalnego) W modelu normalnym, \bar{X} i S^2 są niezależnymi zmiennymi losowymi (tak jest tylko, gdy obserwacje pochodzą z rozkładu normalnego),

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right), \quad (1.18)$$

$$\frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1). \quad (1.19)$$

Wynika stąd, że $\mathbb{E}_{\mu, \sigma} S^2 = \sigma^2$ i $\text{Var}_{\mu, \sigma} S^2 = \frac{2\sigma^4}{(n-1)}$.

1.6. Kowariancja zmiennych losowych

Kowariancja charakteryzuje związek między zmiennymi losowymi.

Definicja 1.6.1. (Kowariancja zmiennych losowych) Kowariancję całkowalnych zmiennych losowych X_1 i X_2 , spełniających warunek $\mathbb{E}|XY| < \infty$, nazywamy wielkość:

$$\text{cov}(X, Y) = \mathbb{E}[(X_1 - \mathbb{E}X_1)(X_2 - \mathbb{E}X_2)], \quad (1.20)$$

$$\text{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y. \quad (1.21)$$

Zmienne X i Y są nieskorelowane, jeśli $\text{cov}(X, Y) = 0$.

Twierdzenie 1.6.1. (Wariancja sumy zmiennych losowych) Jeśli zmienne losowe X_1, \dots, X_n mają wariancję, to istnieje wariancja ich sumy:

$$\text{Var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{Var}X_i + 2 \sum_{1 \leq i < j \leq n} \text{cov}(X_i, X_j). \quad (1.22)$$

Jeśli zmienne losowe X_1, \dots, X_n mają wariancję i są parami nieskorelowane, to

$$\text{Var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{Var}X_i. \quad (1.23)$$

1.7. Rozkład normalny i twierdzenia graniczne

Rozkład normalny jest najważniejszym rozkładem statystyki rozkładów prawdopodobieństwa. Wynika to ze stosowności centralnego twierdzenia granicznego, które mówi, że przy pewnych warunkach rozkład normalny jest granicznym rozkładem ciągu zmiennych losowych. Na ogół, dla rozkładów ciągłych, mamy do czynienia z rodzinami rozkładów indeksowanych przez parametry θ_i .

1.7.1. Rozkład normalny (jednowymiarowy oraz wielowymiarowy)

Jednowymiarowy standardowy rozkład normalny $N(0, 1)$ ma gęstość:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \quad (1.24)$$

Dla rozkładu normalnego (gdzie wektor parametrów $\Theta = (\theta_1, \theta_2) = (\mu, \sigma)$) przy założeniu że $X \sim N(0, 1)$ mamy dla $\sigma X + \mu \sim N(\mu, \sigma^2)$ gęstość:

$$\varphi_{\Theta}(x) = \varphi_{\mu, \sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \quad (1.25)$$

Momenty rozkładu normalnego to: $\mathbb{E}X = \mu$, $VarX = \sigma^2$.

Dla nielosowych a i b oraz zmiennej losowej $X \sim N(\mu, \sigma^2)$, zmienna losowa $Y = a + bX \sim N(a + b\mu, b^2\sigma^2)$.

Parametr $\mu \in \mathbb{R}$ jest parametrem położenia, a parametr $\sigma \in (0, \infty)$ to parametr skali.

W przypadku pozostałych rozkładów ciągłych (innych niż rozkład normalny), jeśli nie widać prostej zależności liniowej, to indeksujące rozkład parametry θ_i możemy traktować jako parametry jego kształtu.

Wielowymiarowy standardowy rozkład normalny dla zmiennej losowej $\mathbf{X} = (X_1, \dots, X_n) \in \mathbb{R}^n$, gdzie $X_i \sim N(0, 1)$ ma gęstość:

$$g(X_1, \dots, X_n) = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2} \sum_{i=1}^n X_i^2}. \quad (1.26)$$

Wektor wielowymiarowy $\mathbf{X} = (X_1, \dots, X_n)$, gdzie $X_i \sim N(\mu_i, \sigma_i^2)$, gdzie $\mathbb{E}(\mathbf{X}) = \mu$ oraz macierzy wariancji i kowariancji $Var\mathbf{X} = \Sigma$, ma gęstość:

$$g(X_1, \dots, X_n) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2} (\mathbf{X}-\mu)' \Sigma^{-1} (\mathbf{X}-\mu)}. \quad (1.27)$$

Używa się tu zapisu, że: $\mathbf{X} \sim N(\mu, \Sigma)$.

Kombinacje liniowe elementów wektora losowego o wielowymiarowym rozkładzie normalnym mają także wielowymiarowy rozkład normalny. Dla nielosowego \mathbf{a} i \mathbf{B} oraz $\mathbf{X} \sim N(\mu, \Sigma)$ zachodzi:

$$\mathbf{a} + \mathbf{B}\mathbf{X} \sim (\mathbf{a} + \mathbf{B}\mu, \mathbf{B}\Sigma\mathbf{B}'). \quad (1.28)$$

Dla wielowymiarowego rozkładu normalnego, jeśli składowe wektora losowego \mathbf{X} są niezależne wówczas funkcja gęstości wektora \mathbf{X} jest iloczynem funkcji gęstości każdej ze zmiennych:

$$f_{\Theta}(X) = \prod_{i=1}^n f_{\mu_i, \sigma_i}(x_i). \quad (1.29)$$

1.7.2. Nierówność Czebyszewa

Podstawowym narzędziem matematycznym służącym do uzyskania oszacowań różnego rodzaju nierówności jest nierówność Czebyszewa postaci:

Twierdzenie 1.7.1. (Nierówność Czebyszewa) *Dla dowolnej nieujemnej zmiennej losowej X oraz dla każdego $\varepsilon > 0$,*

$$\mathbb{P}(X \geq \varepsilon) \leq \frac{\mathbb{E}X}{\varepsilon}. \quad (1.30)$$

Nierówność Czebyszewa ma kilka rozszerzeń, z uwagi na to, że można ją stosować nie tylko do zmiennych losowych, ale też do funkcji od zmiennej losowej $f(X)$.

Stosując nierówność Czebyszewa dla funkcji postaci $(X - \mathbb{E}X)^2$ otrzymuje się nierówność Czebyszewa-Bienayme (Definicja 1.7.1), dla $|X|^p$ nierówność Markowa, a dla $e^{\lambda X}$ wykładniczą nierówność Czebyszewa - można o nich doczytać w [5].

Definicja 1.7.1. (Nierówność Czebyszewa-Bienayme) *Niezależnie od charakteru rozkładu zmiennej losowej (o ile istnieje dla niej wartość oczekiwana $\mathbb{E}X$ i wariancja $\text{Var}X = \sigma_X^2$) zachodzi:*

$$\mathbb{P}(|X - \mathbb{E}X| \leq \varepsilon) \leq \frac{\text{Var}X}{\varepsilon^2}, \quad (1.31)$$

gdzie ε jest dowolną liczbą dodatnią. Dla $\varepsilon = k\sigma_X$ otrzymujemy ograniczenie na prawdopodobieństwo uzyskania wartości zmiennej losowej odległej od wartości oczekiwanej o więcej niż k dyspersji σ_X :

$$\mathbb{P}(|X - \mathbb{E}X| \leq k\sigma_X) \leq \frac{1}{k^2}.$$

Dla **dowolnego rozkładu**, gdy rozkład cech jest nieznan, prawdziwa jest **nierówność Czebyszewa-Bienayme**, która mówi, że prawdopodobieństwo, że wybrana losowo cecha różni się od wartości oczekiwanej o więcej niż $\pm k\sigma$ wynosi co najwyżej $\frac{1}{k^2}$. Na przykład poza przedziałem $\pm 2\sigma$ leży 25% wartości cech.

Powyższy wynik jest jednocześnie ilustracją dobrze znanej reguły pomiarowej [6]:

Jeśli chcesz zmniejszyć błąd dwakroć, próbkę musisz powiększyć czterokroć.

1.7.3. Centralne twierdzenia graniczne (CTG)

W teorii prawdopodobieństwa zakłada się, że rozkłady zmiennych losowych są znane. W zagadnieniach rzeczywistych tak zazwyczaj nie jest. Rozkład prawdopodobieństwa leżący u podstaw badanej cechy nie jest znany, a celem analiz jest uzyskanie jego postaci. Informacje o rozkładzie wyodrębnia się z próbki dbając o to by była ona losowa, tak by informacje były jak najbardziej wiarygodne.

Jeśli X_1, X_2, \dots, X_n są niezależnymi zmiennymi losowymi o jednakowym rozkładzie, to można na te zmienne patrzeć jak na wyniki powtarzanego wielokrotnie takiego samego doświadczenia losowego. Często analizowane twierdzenia rachunku prawdopodobieństwa dotyczą ciągu sum

$$S_n = X_1 + \dots + X_n$$

dla rosnącej liczby powtórzeń (przy $n \rightarrow \infty$) i to na nich opiera się centralne twierdzenie graniczne.

Twierdzenie 1.7.2. (Centralne twierdzenie graniczne (CTG)) *Jeżeli*

X_1, X_2, \dots, X_n są niezależnymi zmiennymi losowymi o jednakowym rozkładzie prawdopodobieństwa o wartości oczekiwanej $\mathbb{E}X_i = \mu$ i wariancji $\text{Var}X_i = \sigma^2$ oraz $S_n = X_1 + X_2 + \dots + X_n$, to dla każdej liczby a zachodzi:

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{S_n - n\mu}{\sqrt{n}\sigma}\right| \leq a\right) = \Phi(a),$$

gdzie

$$\Phi(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a e^{-\frac{x^2}{2}} dx.$$

Wniosek CTG: dla dostatecznie dużych n , suma S_n ma w przybliżeniu rozkład normalny $N(n\mu, n\sigma^2)$.

1.8. Wnioskowanie statystyczne

Estymacja stanowiąca część wnioskowania statystycznego opiera się na założeniu, że dane doświadczalne z populacji (a ściślej jego próba losowa) są wynikiem działania pewnego *mechanizmu losowego*.

Jeśli nie znane są *tylko* wartości parametrów rozkładu prawdopodobieństwa \mathbb{P} na przestrzeni Ω , który rządzi zachowaniem interesujących nas zmiennych losowych, to zagadnienie wnioskowania jest określane jako **wnioskowanie parametryczne**.

Natomiast gdy nie znany jest sam rozkład populacji badanej cechy i chcemy o nim cokolwiek powiedzieć, to wiąże się to z **wnioskowaniem nieparametrycznym**.

Oba podejścia cechuje inny aparat matematyczny wykorzystywany do wnioskowania o cenie. W metodzie estymacji nieparametrycznej w znajdowaniu postaci rozkładu cechy populacji często przechodzi się do metod opartych na weryfikacji hipotez statystycznych (jako prostszego wariantu wnioskowania).

Wnioskowanie statystyczne i późniejsze modelowanie danych zawiera takie elementy jak:

- budowę modelu statystycznego,
- estymację parametryczną (nieparametryczną) nieznanymi parametrów rozkładu (lub odpowiednio samego rozkładu),
- testowanie hipotez statystycznych,
- predykcję oraz jej ocenę.

W dalszej części pracy przedstawię budowę modelu, estymację parametryczną (nie analizuję tu wnioskowania nieparametrycznego zakładając, że znamy postać rozkładu badanej cechy) oraz narzędzia szacowania błędów wynikających z przyjętych uogólnień.

1.8.1. Budowa modelu statystycznego

Podstawowym narzędziem używanym do analizy danych z próby jest model statystyczny. Celem jego jest, za pomocą niewielkiej liczby szacowanych parametrów, uchwycić

najważniejsze zależności zachodzące między obserwowanymi zmiennymi, a pominać te, które są nieistotne w ramach stawianego pytania badawczego.

Model statystyczny określa rodzinę wszystkich branych pod uwagę rozkładów prawdopodobieństwa.

Definicja 1.8.1. (Model statystyczny) *Model statystyczny będziemy określać przez podanie przestrzeni zdarzeń elementarnych Ω (przestrzeni próbkowej), rodziny rozkładów prawdopodobieństwa $\{\mathbb{P}_\theta : \theta \in \Theta\}$ ⁶ określonych na Ω oraz zmiennych losowych X_1, \dots, X_n , które nazywamy obserwacjami. W praktyce przestrzeń próbkowa będzie pewnym podzbiorem \mathbb{R}^n , natomiast rozkład prawdopodobieństwa na przestrzeni próbkowej jest łącznym rozkładem prawdopodobieństwa wektora losowego (X_1, \dots, X_n) .*

Przykład 1.8.1. (Pomiar z błędem losowym) *Powtarzamy niezależnie n razy pomiar pewnej wielkości fizycznej μ . Wyniki poszczególnych pomiarów X_1, \dots, X_n są zmiennymi losowymi, bo przyrząd pomiarowy jest niedoskonały. Najczęściej zakłada się, że każdy z pomiarów ma jednakowy rozkład normalny $N(\mu, \sigma^2)$. Uzyskuje się zatem:*

$$\mathbb{P}_{\mu, \sigma}(X_1 \leq x_1, \dots, X_n \leq x_n) = \prod_{i=1}^n \Phi\left(\frac{x_i - \mu}{\sigma}\right);$$

$$f_{\mu, \sigma}(x_1, \dots, x_n) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left[-\frac{1}{2\sigma^2} \sum (x_i - \mu)^2\right].$$

Tutaj rolę parametru θ gra para liczb (μ, σ) gdzie $-\infty < \mu < \infty$ i $\sigma \geq 0$. Przestrzenią parametrów jest $\Theta = \mathbb{R} \times [0, \infty[$. Przestrzenią próbkową jest $\Omega = \mathbb{R}^n$. Wiemy, że $\mathbb{E}_{\mu, \sigma} X_i = \mu$ i $\text{Var}_{\mu, \sigma} X_i = \sigma^2$.

1.8.2. Teoria estymacji parametrycznej

Jak już zostało wspomniane, zagadnienie estymacji parametrycznej polega na oszacowaniu nieznanego parametru θ_i rozkładu prawdopodobieństwa populacji na podstawie wyników obserwacji uzyskanej z próby.

W teorii estymacji parametrycznej mamy do czynienia z dwoma rodzajami estymacji. Pierwsza z nich to **estymacja punktowa**, gdzie wynikiem estymacji parametru θ jest estymator $\hat{\theta}$ w postaci liczbowej. Drugim rodzajem estymacji jest **estymacja przedziałowa** dająca w wyniku oszacowania przedział liczbowy nazwany przedziałem ufności, do którego z pewnym prawdopodobieństwem należy szukana wartość. W dalszej części pracy skoncentruję się na parametrycznej estymacji punktowej.

Parametryczna estymacja punktowa

Istnieje kilka metod punkowego wyznaczania estymatorów $\hat{\theta}$ dla parametrów θ . Najpopularniejsze to:

1. **Metoda podstawiania częstości.** Za oszacowanie nieznanego prawdopodobieństwa pojawiania się zdarzeń przyjmujemy częstości ich wystąpienia w próbie losowej. Podstawianie częstości jest jedną z łatwiejszych metod estymacji dopuszczającą dużą dowolność w konstrukcji estymatorów.

⁶ Parametr Θ jest przestrzenią parametrów.

Przykład 1.8.2. (Wadliwy element) Zastępujemy tu nieznane prawdopodobieństwo θ napotkania wadliwego wyrobu przez prawdopodobieństwo empiryczne. Estymatorem jest $\hat{\theta} = \bar{X} = \frac{k}{n}$, czyli frakcją wadliwych sztuk w próbie losowej.

2. **Estymacja metodą największej wiarygodności (ENW).** Metoda największej wiarygodności (MNW) to metoda wyznaczania estymatorów parametru θ , polegająca na tym, że estymator przyjmuje taką wartość parametru θ , dla której funkcja wiarygodności (łączne prawdopodobieństwo lub łączna gęstość prawdopodobieństwa) danej próby losowej jest największa [4]. O MNW powiem więcej w dalszej części opracowania.
3. **Metoda najmniejszych kwadratów (MNK).** Metoda najmniejszych kwadratów (MNK) polega na znajdowaniu estymatorów modelu regresji ($y_i = \mathbf{x}_i\beta + \varepsilon_i$, dla $i = 1, \dots, N$) minimalizując sumę kwadratów reszt (która to suma jest miarą jakości dopasowania). W przypadku tej metody możliwa jest precyzyjna dekompozycja (podział) wariancji empirycznej zmiennej zależnej na część wyjaśnioną przez model i część niewyjaśnioną przez model. Dla pomiarów obarczonych błędem gaussowskim, o wartości oczekiwanej zero ($\mathbb{E}(\varepsilon_i) = 0$ dla $i = 1, \dots, N$) i stałej wariancji (założenie o homoskedastyczności błędu losowego, czyli $\text{Var}(\varepsilon_i) = \sigma^2$ dla $i = 1, \dots, N$) oraz zakładając niezależność kolejnych pomiarów, kryterium metody MNK jest równoważne kryterium metody MNW (maksymalizacja logarytmu funkcji wiarygodności prowadzi do metody najmniejszych kwadratów). Wówczas estymatory wielkości badanej wyznaczone obema metodami są identyczne i obarczone są takim samym błędem [2].
4. **Estymacja metodą momentów (EMM).** Metoda momentów, opracowana przez angielskiego statystyka K. Pearsona pod koniec XIX wieku, polega na przyrównaniu momentów rozkładu teoretycznego (najczęściej momentów centralnych), zależnych od nieznanego parametru $\theta_i \in \Theta$, do odpowiednich momentów empirycznych. Z powstałych równań wyznacza się parametry, układając tyle równań, ile jest niewiadomych parametrów (współrzędnych wektora Θ).

Przykład 1.8.3. (Rozkład normalny) Rozpatrujemy próbę losową (X_1, \dots, X_n) z rozkładu normalnego $N(\mu, \sigma^2)$ o nieznannej wartości oczekiwanej μ i nieznannej wariancji σ^2 . Szukamy estymatora wektora parametrów modelu $\Theta = (\mu, \sigma^2)$.

Korzystając z równości $\text{Var}X = \mathbb{E}X^2 - (\mathbb{E}X)^2$, którą można zapisać jako⁷ $m_2 = \mu_2 - \mu_1^2$, można otrzymać układ równań, gdzie podstawiając w miejsce momentów zwykłych ich estymatory uzyskuje się:

$$\begin{aligned}\mu_1 &= \mu = \frac{1}{n} \sum_{i=1}^n X_i, \\ \mu_2 &= \mu^2 + \sigma^2 = \frac{1}{n} \sum_{i=1}^n X_i^2.\end{aligned}$$

Rozwiązując układ ze względu na μ i σ^2 otrzymuje się szukane wartości estymatorów postaci:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}, \quad (1.32)$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \left[\frac{1}{n} \sum_{i=1}^n X_i \right]^2. \quad (1.33)$$

⁷ $m_i \equiv i$ -ty moment centralny, $\mu_i \equiv i$ -ty moment zwykły

5. **Estymacja metodą kwantyli (EMK).** W tej metodzie zamiast momentów używa się kwantyli przyrównując ze sobą wartości teoretyczne i empiryczne. Kwantyli dobiera się tyle ile jest niewiadomych $\theta_i \in \Theta$.

Metoda największej wiarygodności (MNW)

Metodę największej wiarygodności po raz pierwszy zaproponował Ronald Fisher (1922). Daje ona zazwyczaj najlepsze rezultaty w ocenie jakości uzyskanego estymatora niż inne metody estymacji.

Metodę największej wiarygodności stosuje się przede wszystkim w sytuacjach, w których nie widać od razu (bądź nie da się odgadnąć) rozsądnie dobrego estymatora.

Kluczową rolę w MNW odgrywa funkcja wiarygodności $P(\theta)$.

Definicja 1.8.2. (Funkcja wiarygodności $P(\theta)$) Funkcję wiarygodności nazywamy funkcję $P : \Theta \rightarrow \mathbb{R}$ daną wzorem:

$$P(\theta, x) = f_\theta(x_1, x_2, \dots, x_n) = f_\theta(x_1)f_\theta(x_2)\dots f_\theta(x_n), \quad (1.34)$$

gdzie $x = (x_1, x_2, \dots, x_n)$ jest próbka zaobserwowanych wartości zmiennych X_1, X_2, \dots, X_n będących próbą prostą⁸ z rozkładu o gęstości $f_\theta(x)$, gdzie θ jest nieznanym parametrem a Θ - przestrzenią parametrów.

W celu ułatwienia obliczeń wyznaczenia estymatora θ korzysta się z funkcji $\ln P(x, \theta)$ wykorzystując wiedzę, że funkcja wiarygodności osiąga maksimum w tym samym punkcie co jej logarytm.

Definicja 1.8.3. (Estymator największej wiarygodności $ENW(\theta)$) Estymatorem największej wiarygodności parametru θ ($ENW(\theta)$) nazywamy maksimum funkcji $P(\theta)$ lub funkcji $\ln P(\theta)$ co zapisujemy jako:

$$ENW(\theta) = \hat{\theta} = \operatorname{argmax}_{\theta \in \Theta} P(\theta, x) = \operatorname{argmax}_{\theta \in \Theta} \ln P(\theta, x).$$

Zachodzi też zależność:

$$ENW(g(\theta)) = g(ENW(\theta)).$$

Jeżeli $\Theta = (\theta_1, \theta_2, \dots, \theta_k)$ jest parametrem ciągłym i $P(\Theta)$ jest funkcją różniczkowalną, to ENW wyznaczamy rozwiązując układ równań:

$$\frac{\partial P(\Theta, x)}{\partial \theta_j} = 0, \quad j = 1, 2, \dots, k,$$

lub równoważny układ:

$$\frac{\partial \ln P(\Theta, x)}{\partial \theta_j} = 0, \quad j = 1, 2, \dots, k.$$

⁸ Próba prosta określana też jako iid. (z ang. independent and identically distributed random variables) czyli niezależne zmienne losowe o jednakowym rozkładzie \mathbb{P} . Niezależność zmiennych losowych oznacza tutaj brak korelacji między nimi (kowariancje są zerowe).

Jest to warunek na maksimum funkcji odpowiednio $P(\Theta)$ lub $\ln P(\Theta)$. Poza tym potrzeba i wystarcza, aby forma kwadratowa

$$\sum_{i=1}^k \sum_{j=1}^k \left(\frac{\partial^2 \ln P(\Theta)}{\partial \theta_i \partial \theta_j} \right) \Big|_{\Theta=\hat{\Theta}} \Delta \theta_i \Delta \theta_j \quad (1.35)$$

była określona ujemnie. Przyrosty $\Delta \theta_i \Delta \theta_j$ są zmiennymi rzeczywistymi i nie zerują się jednocześnie.

Interpretacja jest następująca: większa wartość formy kwadratowej (wzięta ze znakiem przeciwnym) oznacza węższe maksimum $\ln P(\theta)$ w punkcie $\theta = \hat{\theta}$, tzn. większą krzywiznę funkcji wiarygodności i mniejszą niepewność określenia parametru θ .

Przykład 1.8.4. (Rozkład normalny - Metoda Największej Wiarygodności)

Rozważmy próbkę X_1, \dots, X_n z rozkładu $N(\mu, \sigma^2)$ z nieznanymi parametrami μ i $\sigma > 0$. Z definicji funkcji wiarygodności i niezależności zmiennych losowych widać, że:

$$P(\mu, \sigma) = f_{\mu, \sigma}(x_1, \dots, x_n) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right), \quad (1.36)$$

$$\ln P(\mu, \sigma) = -\frac{n}{2} \ln 2\pi - n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2. \quad (1.37)$$

Układ równań przybiera postać:

$$\begin{aligned} \frac{\partial \ln P}{\partial \sigma} &= \frac{n}{\sigma} + \frac{1}{\sigma^3} (\sum_{i=1}^n x_i^2 - 2\mu \sum_{i=1}^n x_i + n\mu^2) = 0, \\ \frac{\partial \ln P}{\partial \mu} &= \frac{1}{\sigma} \sum_{i=1}^n x_i - \frac{n\mu}{\sigma^2} = 0. \end{aligned}$$

Otrzymujemy ENW postaci:

$$\hat{\mu} = \bar{X}; \quad (1.38)$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum x_i^2 - \bar{x}^2 = \frac{1}{n} \sum (x_i - \bar{x})^2. \quad (1.39)$$

Jak się okazuje są to znane estymatory \bar{X} oraz \tilde{S}^2 ($ENW(\mu)$ i $ENW(\sigma^2)$).

1.8.3. Własności estymatorów EMNW oraz wybór najlepszego.

Dlaczego estymator MNK jest najlepszy? Estymatory MNW posiadają pewne uniwersalne własności co sprawia, że są one na ogół *najlepszymi* estymatorami. Są to nieobciążoność, zgodność i efektywność (lub asymptotyczna efektywność).

Obciążenie estymatora

Interpretacja: Jeśli estymator parametru jest nieobciążony, to otrzymujemy oceny bez błędu systematycznego. Wprowadzamy oznaczenie, że \hat{g} jest estymatorem parametru g . Jeśli byłoby $\hat{g} < g$, to otrzymywalibyśmy oceny średnio zaniżone. Natomiast, gdyby $\hat{g} > g$, to otrzymywalibyśmy oceny średnio zawyżone. Estymator asymptotycznie nieobciążony jest praktycznie estymatorem nieobciążonym, gdy próba jest liczna.

Definicja 1.8.4. (Obciążenie estymatora) Obciążeniem estymatora parametru θ jest funkcja [3]:

$$B_\theta(\hat{g}(X)) = \mathbb{E}_\theta \hat{g}(X) - g(\theta).$$

Estymator $\hat{g}(X)$ wielkości $g(\theta)$ jest **nieobciążony**, jeśli dla każdego $\theta \in \Theta$ funkcja obciążenia $B_\theta(\hat{g}(X)) = 0$, czyli gdy

$$\mathbb{E}_\theta \hat{g}(X) = g(\theta).$$

Przykład 1.8.5. (Czym różnią się S i \hat{S} ?) Załóżmy, że X_1, \dots, X_n jest próbką z rozkładu $N(\mu, \sigma^2)$ i

- $\mathbb{E}(\hat{S}^2) = \mathbb{E}(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2) = \frac{n-1}{n} \sigma^2$ jest estymatorem obciążonym,
- $\mathbb{E}(S^2) = \mathbb{E}(\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2) = \sigma^2$ jest estymatorem nieobciążonym wariancji i dlatego S^2 stosuje się częściej.

Definicja 1.8.5. (Asymptotyczna nieobciążoność) Mówimy, że estymator parametru jest asymptotycznie nieobciążony, gdy dąży on do granicy wartości oczekiwanej równej parametrowi [3]

$$\forall \theta \in \Theta \quad \lim_{n \rightarrow +\infty} B_\theta(\hat{g}(X)) = 0.$$

Ryzyko estymatora (błąd średniokwadratowy)

Definicja 1.8.6. (Ryzyko estymatora) Funkcję

$$R(\theta, \hat{g}) = \mathbb{E}_\theta(\hat{g}(X) - g(X))^2 = B_\theta^2(\hat{g}) + \text{Var}_\theta(\hat{g}) \quad (1.40)$$

nazywamy ryzykiem estymatora \hat{g} przy kwadratowej funkcji straty lub błędem średniokwadratowym⁹ [3].

Jeżeli \hat{g} jest estymatorem nieobciążonym funkcji $g(\theta)$, tzn. $\mathbb{E}_\theta(\hat{g}) = g(\theta)$ dla każdego $\theta \in \Theta$, to

$$R(\theta, \hat{g}) = \text{Var}_\theta(\hat{g}). \quad (1.41)$$

WNIOSEK: Rozkład ryzyka ma dwa odmienne składniki: wariancję i kwadrat obciążenia wskazując na dwa odmienne źródła błędu estymacji. Przy estymatorach nieobciążonych miernikiem jakości estymatora jest jego wariancja.

Porównanie dwóch estymatorów ma sens, gdy są one estymatorami tej samej wielkości $g(\theta)$. Lepszym estymatorem jest ten o mniejszym ryzyku. Definicja poniżej wymaga by cały wykres funkcji $R(\theta, \hat{g}_1)$ leżał poniżej funkcji $R(\theta, \hat{g}_2)$. Jeśli oba wykresy się krzyżują, nie możemy na bazie tej definicji powiedzieć o tym, który z nich jest lepszy.

Definicja 1.8.7. (Porównanie dwóch estymatorów) Mówimy, że estymator \hat{g}_1 jest lepszy niż $\hat{g}_2 \Leftrightarrow \forall \theta R(\theta, \hat{g}_1) \leq R(\theta, \hat{g}_2)$ i $\exists \theta_0 R(\theta_0, \hat{g}_1) < R(\theta_0, \hat{g}_2)$ [3].

Porównywanie estymatorów zaczyna być interesujące dopiero wtedy, gdy ograniczymy się do estymatorów, które spełniają dodatkowe warunki. Jeśli bowiem ograniczymy się tylko do estymatorów nieobciążonych, które stanowią ważną klasę estymatorów, to możemy poszukiwać estymatorów najlepszych.

⁹ UWAGA: Można użyć innej funkcji ryzyka np. $R(\theta, \hat{g}) = \mathbb{E}_\theta |\hat{g}(X) - g(X)|$. Funkcja ta jest jednak mniej korzystna bo znacznie mniej wygodna rachunkowo.

Zgodność estymatora

Interpretacja: Jeśli estymator jest zgodny, to dla dużej próby z prawdopodobieństwem bliskim 1 ocena parametru i parametr mało różnią się.

Definicja 1.8.8. (Zgodność estymatora) Zgodność estymatora $\hat{g}(X_1, X_2, \dots, X_n) = \hat{g}_n$, funkcji $g(\Theta)$ dla każdego $\varepsilon > 0$ i $\theta \in \Theta$ za [3]

$$\lim_{n \rightarrow \infty} \mathbb{P}_\theta(|\hat{g}_n - g(\theta)| < \varepsilon) = 1.$$

Podczas estymacji parametru g statystyką \hat{g} chcemy by błąd przybliżenie był minimalny, czyli by różnica $\hat{g}(\theta) - g(\theta)$ miała średnio najmniejszą wartość (dlatego też posługujemy się średnim kwadratem błędu).

Informacja Fishera

Niech f_θ będzie rodziną gęstości zmiennej losowej X . Zakładamy, że istnieje pochodna $\frac{d}{d\theta} f_\theta$ a Θ jest przedziałem otwartym (tak by pochodna na brzegu była określona) wówczas, jeśli zbiór $x \in \chi : f_\theta(x) > 0$ nie zależy od θ to możemy mówić o wartości informacji Fishera określonej na próbie.

Definicja 1.8.9. (Informacja Fishera $I(\theta)$) Niech X będzie zmienną losową o łącznej gęstości f_θ wektora losowego zależnego od jednowymiarowego parametru $\theta \in \Theta \in \mathbb{R}$. Funkcję

$$I_1(\theta) = E_\theta \left(\frac{d \ln f_\theta(X)}{d\theta} \right)^2 \quad (1.42)$$

nazywamy **informacją Fishera** zawartą w pojedynczej obserwacji X . Informację zawartą w ciągu obserwacji X_1, \dots, X_n określamy wzorem:

$$I_n(\theta) = E_\theta \left(\frac{\partial \ln f_\theta(X_1, \dots, X_n)}{\partial \theta} \right)^2. \quad (1.43)$$

Dla zmiennej ciągłej i dyskretniej mamy więc odpowiednio:

$$I_1(\theta) = \begin{cases} \int \left(\frac{d}{d\theta} \ln f_\theta(x) \right)^2 f_\theta(x) dx & \text{dla zmiennej ciągłej,} \\ \sum_x \left(\frac{d}{d\theta} \ln P_\theta(x) \right)^2 P_\theta(x) & \text{dla zmiennej dyskretniej.} \end{cases} \quad (1.44)$$

Jeżeli X_1, \dots, X_n jest próbka i.i.d. z rozkładu o gęstości $f_\theta(x)$ to

$$I_n(\theta) = n I_1(\theta). \quad (1.45)$$

Definicja 1.8.10. (Informacja Fishera $I(\theta)$) Jeżeli funkcja gęstości jest dwukrotnie różniczkowalną funkcją zmiennej θ to informację Fishera można obliczyć korzystając ze wzoru

$$I_n(\theta) = -\mathbb{E}_\theta \left(\frac{\partial^2 \ln f_\theta(X_1, \dots, X_n)}{\partial^2 \theta} \right). \quad (1.46)$$

Dla parametru wektorowego $\Theta = (\theta_1, \dots, \theta_d)$ możemy zdefiniować postać **obserwowanej informacji Fishera** zdefiniowanej w punkcie $\Theta = \hat{\Theta}$ jako $iF = iF(\hat{\Theta}) = iF(\Theta)|_{d \times d}$:

$$iF(\Theta) = -\frac{\partial^2}{\partial \Theta \partial \Theta^T} \ln P(x|\Theta) = -\left(\frac{\partial^2 \ln P(\Theta)}{\partial \theta_i \partial \theta_j} \right)_{d \times d}. \quad (1.47)$$

Jest ona dodatnio określona oraz symetryczna. Ma ona sens tylko w pobliżu $\hat{\Theta}$. Jej wartość odnosi się do pojedynczego zestawu danych, zmienia się od próbki do próbki i jest związana z obserwowaną wartością wiarygodności. **Oczekiwana informację Fishera** I_F definiujemy natomiast jako:

$$I_F(\Theta) \equiv \mathbb{E}_{\Theta}(\mathbf{i}F(\Theta)) = \int_{\mathfrak{B}} dy P(y|\Theta) \mathbf{i}F(\Theta), \quad (1.48)$$

gdzie \mathfrak{B} jest przestrzenią próby. Oczekiwana I_F ma sens jako funkcja dopuszczalnych wartości Θ .

Przykład 1.8.6. (Informacja Fishera I dla wariancji $N(\theta, 1)$) Wyznaczyć wartość informacji Fishera dla wartości średniej θ , gdzie $X_1, \dots, X_n \sim N(\theta, 1)$.

Rozwiązanie: Wiadomo już, że estymator największej wiarygodności przyjmuje wartość: $EMNW(\theta) = \frac{1}{n} \sum X_i = \hat{\theta}$. Wartość oczekiwana estymatora $\hat{\theta}$ wynosi $\mathbb{E}\hat{\theta} = \frac{1}{n} \mathbb{E} \sum x_i = \theta$ i jest on estymatorem nieobciążony. Celem jest wyznaczenie informacji Fishera $I_n(\theta) = nI_n(\theta)$. Definicja wymaga posłużenie się funkcją wiarygodności i jej logarytmem:

$$\ln P(\theta) = \frac{1}{2} \ln 2\pi - \frac{1}{2} (X - \theta)^2,$$

$$\frac{d^2 \ln P(\theta)}{d\theta^2} = -1,$$

$$I_n(\theta) = n.$$

Informacja Fishera przyjmuje wartość n . Widać też, że wariancja obserwowana jest odwrotnością informacji Fishera I , czyli ma minimalną wariancję $\text{Var}\hat{\theta} = \frac{1}{n}$, co zostanie wspomniane poniżej.

Nierówność Rao-Cramera (NRC)

Nierówność Rao-Cramera daje najważniejsze ograniczenie statystyki informacyjnej na jakość estymacji.

Definicja 1.8.11. (Nierówność Rao-Cramera) Nierówność Rao-Cramera (zwana nierównością informacyjną) przy pewnych warunkach regularności, tzn. jeżeli \hat{g} jest estymatorem nieobciążonym funkcji różniczkowalnej $g(\theta)$, to za [3]

$$\forall \theta \in \Theta \quad \text{Var}_{\theta} \hat{g}(X) \geq \frac{(g'(\theta))^2}{I_n(\theta)}. \quad (1.49)$$

W szczególności, jeżeli $\hat{\theta}(X)$ jest nieobciążonym estymatorem parametru θ , to

$$\text{Var}_{\theta} \hat{\theta}(X) \geq \frac{1}{I_n(\theta)}. \quad (1.50)$$

Widać stąd, że funkcja ryzyka $R(\theta)$ każdego nieobciążonego estymatora (wzór (1.41)) nie może nigdzie być mniejsza od pewnej ustalonej funkcji. Istnieje granica jakości nieprzeracalna dla wszystkich nieobciążonych estymatorów. Jeśli ryzyko jakiegoś estymatora nieobciążonego jest równe dolnemu oszacowaniu Rao-Cramera (DORC), to jest to estymator najlepszy wśród nieobciążonych, a więc jest on ENMW (estymatorem nieobciążonym o minimalnej wariancji).

Przykład 1.8.7. (Nierówność Rao-Cramera i jej dolne ograniczenie (DORC))

Na podstawie nierówności Rao-Cramera wyznaczyć dolne ograniczenie dla wariancji nieobciążonego estymatora wariancji σ^2 w rozkładzie $N(0, \sigma^2)$.

Rozwiązanie: Celem jest uzyskanie oszacowania nierówności informacyjnej Rao-Cramera: $\text{Var}_\theta T(x) \geq \frac{[g'(\theta)]^2}{I_n(\theta)}$. W tym przypadku: $g(\theta = \sigma) = \sigma^2 \Rightarrow [g'(\theta)]^2 = (2\sigma)^2 = 4\sigma^2$. Chcąc wyznaczyć wartość $I_n(\theta = \sigma)$ można posłużyć się funkcją wiarygodności, i tak:

$$P(\sigma) = (1/(\sqrt{2\pi}\sigma))e^{-\sum x^2/2\sigma^2},$$

$$\ln P(\sigma) = -\frac{1}{2}\ln 2\pi - n\ln \sigma - \frac{1}{2\sigma^2} \sum x^2,$$

$$(\ln P(\sigma))' = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum x^2,$$

$$(\ln P(\sigma))'' = \frac{n}{\sigma^2} - \frac{3}{\sigma^4} \sum x^2,$$

$$I_n(\sigma) = -\mathbb{E}\left(\frac{d^2}{d\theta^2} \ln P(\sigma)\right) = \mathbb{E}\left(-\frac{n}{\sigma^2} + \frac{3}{\sigma^4} \sum x_i^2\right) = -\frac{n}{\sigma^2} + \frac{3}{\sigma^4} n \mathbb{E}x_1^2 = \frac{n}{\sigma^2} + \frac{3}{\sigma^4} n \sigma^2 = \frac{4n}{\sigma^2}.$$

Odpowiedź: DORC przyjmuje wartość: $\frac{[g'(\theta)]^2}{I_n} = \frac{4\sigma^2}{\frac{4n}{\sigma^2}} = \frac{\sigma^2}{n}$.

Definicja 1.8.12. (Efektywność estymatora nieobciążonego) Efektywność estymatora nieobciążonego $\hat{g}(X_1, \dots, X_n)$ funkcji różniczkowalnej $g(\theta)$ definiujemy jako:

$$eff_\theta(\hat{g}(X_1, \dots, X_n)) = \frac{(g'(\theta))^2}{I_n(\theta) \text{Var}_\theta(\hat{g})}. \quad (1.51)$$

Estymator jest asymptotycznie najefektywniejszy, gdy jest nieobciążony (lub asymptotycznie nieobciążony) i jego efektywność zbiega do jedynki.

Własności estymatora największej wiarygodności

Estymator największej wiarygodności posiadają pewne uniwersalne własności, co sprawia, że jest on na ogół *dobrym* estymatorem. Są to:

1. Zgodność estymatorów największej wiarygodności.
2. Asymptotyczna nieobciążoność estymatora największej wiarygodności.
3. Asymptotycznie najefektywniejszy.
4. Asymptotyczna normalność rozkładu estymatora największej wiarygodności parametru $\theta \sim N(0, \frac{1}{I_1(\theta)})$, $I_n(\theta)$ jest informacją Fishera.

Definicja 1.8.13. (Asymptotyczna normalność) Jeżeli dodatkowo istnieje $\frac{\partial^2 \ln P(\theta, x_1, \dots, x_n)}{\partial \theta^2}$ i spełnione są założenia umożliwiające zamianę kolejności operacji różniczkowania po $\frac{\partial}{\partial \theta}$ lub $\frac{\partial^2}{\partial \theta^2}$ i całkowania $\int \dots dx$ i informacja Fishera $I(\theta) > 0$ jest określona, to $\hat{\theta}_n = \hat{\theta}(X_1, \dots, X_n) = ENW(\theta)$ jest **asymptotycznie normalny** i

$$(\hat{\theta}_n - \theta)\sqrt{n} \rightarrow N(0, \frac{1}{I_1(\theta)})$$

przy $n \rightarrow +\infty$.

Jeżeli g jest różniczkowalna i $g'(\theta) \neq 0$ i $\hat{\theta}_n = ENW(\theta)$, to

$$(g(\hat{\theta}_n) - g(\theta))\sqrt{n} \rightarrow N(0, \frac{[g'(\theta)]^2}{I(\theta)}).$$

Przykład 1.8.8. (Własności estymatora wartości oczekiwanej rozkładu normalnego)

Dana jest cecha $X \sim N(m, \sigma)$, gdzie σ jest znane. Estymator wartości oczekiwanej jest średnią z próby:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Celem przykładu jest tu zbadanie własności $EMNW(m)$, czyli zgodności, nieobciążoności, funkcji ryzyka oraz efektywności.

Zgodność

Średnia z próby \bar{X} jest estymatorem zgodnym wartości oczekiwanej m zgodnie z Definicja 1.8.8.

Nieobciążoność

Ponieważ $\mathbb{E}\bar{X}_n = \mathbb{E}[\frac{1}{n} \sum_{i=1}^n X_i] = \frac{1}{n} \sum_{k=1}^n \mathbb{E}X_k = \frac{1}{n} \sum_{k=1}^n m = m$ więc średnia z próby jest estymatorem nieobciążonym wartości oczekiwanej.

Funkcja ryzyka

Funkcja ryzyka tego estymatora ma postać

$$R(m, \sigma) = \mathbb{E}_{m, \sigma}(\bar{X} - m)^2 = \frac{\sigma^2}{n}.$$

Efektywność

Rozwiązanie:

$$g'(m) = 1,$$

$$\text{Var} \bar{X} = \text{Var} \left[\frac{1}{n} \sum_{i=1}^n X_i \right] = \frac{1}{n^2} n \text{Var}[X_i] = \frac{\sigma^2}{n},$$

wiedząc, że: $I_n(\theta = m) = nI_1(\theta = m)$ wyznacza się wartość informacji Fishera dla pojedynczej próby:

$$\ln f(X_i) = -\frac{1}{2} \ln 2\pi - \ln \sigma - \frac{(X_i - m)^2}{2\sigma^2},$$

$$\frac{\partial^2 \ln f(X_i)}{\partial m^2} = \frac{\partial}{\partial m} \left(\frac{\partial \ln f(X_i)}{\partial m} \right) = \frac{\partial}{\partial m} \left(\frac{x - m}{\sigma^2} \right) = -\frac{1}{\sigma^2} \text{ otrzymując:}$$

$$I_1(m) = \frac{1}{\sigma^2} \Rightarrow I_n(m) = \frac{n}{\sigma^2}.$$

Efektywność (za Definicja 1.8.12) wynosi jeden, więc średnia z próby jest estymatorem najefektywniejszym wartości oczekiwanej rozkładu normalnego.

Odpowiedz: Średnia z próby X jest estymatorem zgodnym, nieobciążonym i najefektywniejszym wartości oczekiwanej m rozkładu normalnego.

Estymacja przedziałowa

Estymacja punktowa pozwala uzyskiwać oceny wartości nieznanymi parametrów rozkładu, lecz nie daje odpowiedzi na pytanie, jaka jest dokładność uzyskanej oceny. Sposobem estymacji udzielającym takiej odpowiedzi jest metoda przedziałowa. Polega ona na podaniu przedziałów ufności dla nieznanymi parametrów danego rozkładu. Temat estymacji przedziałowej nie będzie poruszany w pracy.

1.8.4. Podsumowanie

Celem tego rozdziału było wprowadzenie podstawowych pojęć statystyki matematycznej takich jak funkcja wiarygodności $P(\Theta)$, metoda największej wiarygodności (MNV) oraz informacja Fishera ($I(\Theta)$), obserwowana ($iF(\Theta)$) oraz oczekiwana wartość ($I_F(\Theta)$). W trzecim rozdziale stanowią one podstawę używanych przez nas pojęć metody ekstremalnej fizycznej informacji (EFI), która to z informacji Fishera wyprowadza człon kinematyczny konieczny do konstrukcji modeli fizycznych.

Rozdział 2

Wprowadzenie do podstawowych równań mechaniki kwantowej

Równanie Schrödingera jest tym w nierelatywistycznej mechanice kwantowej, czym równania dynamiki Newtona w mechanice klasycznej. Równanie to nie spełnia jednak warunków transformacyjnych wymaganych przez szczególną teorię względności Einsteina (STW)¹. Doświadczalnie i obliczeniowo przestaje je spełniać w momencie, gdy prędkości cząstek nie są już wystarczająco małe w porównaniu z prędkością światła, a dzieje się tak na przykład dla elektronów ciężkich pierwiastków (największą prędkość elektrony mają w pobliżu jąder, tym większą im większa jest jego liczba atomowa Z). Dla atomu wodoru istnieją również mierzalne poprawki (struktura subtelna) rzędu $(v/c)^4$, które trzeba dodatkowo uwzględnić w rozwiązaniu równania nierelatywistycznego. Jednak dla pierwiastków ciężkich poprawki relatywistyczne okazały się niezbędne do tego stopnia, że wymagały skonstruowania nowego równania, które od samego początku byłoby już relatywistyczne. Uogólnienia równania falowego Schrödingera na przypadek relatywistyczny dla elektronów dokonał Paul Dirac.

2.1. Podstawowe równania mechaniki nierelatywistycznej i relatywistycznej

W 1928 roku znane było już nierelatywistyczne równanie mechaniki kwantowej zwane równaniem Schrödingera (1926) oraz jego relatywistyczny odpowiednik dla cząstek o spinie zero - równanie Kleina-Gordona (1926)². Równanie Diraca zostało opublikowane w roku 1928 jako relatywistyczne uogólnienie równania Schrödingera dla cząstek o spinie $\frac{1}{2}$ (fermionów). Równania te przyjmują postać³:

Równanie Schrödingera:

$$\left(i\hbar\frac{\partial}{\partial t} + \frac{\hbar^2}{2m_0}\nabla^2\right)\Psi = 0 \quad (2.1)$$

Równanie Kleina-Gordona⁴:

$$(\square - m_0^2)\Psi = 0 \quad (2.2)$$

¹ Równanie Schrödingera nie jest niezmiennicze względem transformacji Lorentza. Jest niezmiennicze względem transformacji Galileusza.

² Równanie Kleina-Gordona jest niezmiennicze względem transformacji Lorentza.

³ Przyjmujemy tu: $\hbar, c = 1$ oraz dla równania Kleina-Gordona operator d'Alemberta $\square = \nabla^2 - \frac{\partial^2}{\partial t^2}$, wielkość γ^μ dla równania Diraca zostanie zdefiniowana w dalszej części rozdziału.

⁴ Dla $\hbar, c \neq 1$ mamy $\left(c^2\hbar^2\nabla^2 - \hbar^2\frac{\partial^2}{\partial t^2}\right)\Psi = m_0^2c^4\Psi$, gdzie $\square = \nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}$.

Równanie Diraca:

$$\left(i\gamma^\mu \frac{\partial}{\partial x^\mu} - m_0\right)\Psi = 0 \quad (2.3)$$

2.2. Wyprowadzenie równań mechaniki kwantowej**2.2.1. Równanie mechaniki nierelatywistycznej - równanie Schrödingera**

Równanie Schrödingera można wyprowadzić przy wykorzystaniu klasycznego wyrażenia na energię całkowitą cząstki:

$$\mathcal{H} = \frac{\vec{p}^2}{2m_0} + V(\vec{r}).$$

Ograniczę się tutaj do cząstki swobodnej - do członu jej energii kinetycznej. Gdy cząstka nie jest ograniczona przestrzennie i znajduje się poza działaniem jakiegokolwiek oddziaływania - nie działa na nią żadna siła zewnętrzna i znika człon potencjału oddziaływań. Korzystając z zależności energii i pędu postaci:

$$\mathcal{H} = \frac{\vec{p}^2}{2m_0}, \quad (2.4)$$

zastępując wartość energii \mathcal{H} jego operatorem oraz składowe pędu \vec{p} operatorami \hat{p} (zgodnie z regułą Jordana):

$$\left. \begin{aligned} \mathcal{H} = h\nu &\rightarrow i\hbar \frac{\partial}{\partial t} \\ p_x &\rightarrow \frac{\hbar}{i} \frac{\partial}{\partial x} \\ p_y &\rightarrow \frac{\hbar}{i} \frac{\partial}{\partial y} \\ p_z &\rightarrow \frac{\hbar}{i} \frac{\partial}{\partial z} \end{aligned} \right\} \quad \vec{p} = \hbar \vec{k} \rightarrow \frac{\hbar}{i} \nabla \quad (2.5)$$

otrzyma się *równanie Schrödingera* (2.1).

Schrödinger formułując postać równania falowego nie uwzględnił dwuskładnikowej postaci funkcji falowej. Postać ta doczekała się modyfikacji na postać dwuskładnikową chociażby za sprawą doświadczenia Sterna-Gerlacha. W doświadczeniu tym obserwuje się (z uwagi na występujące tu pole magnetyczne) rozszczepienie wiązki na dwie składowe. Obserwujemy dubletową strukturę widma co może świadczyć o tym, że fale prawdopodobieństwa występują w dwóch różnych stanach polaryzacji opisanych funkcjami Ψ_+ , Ψ_- . Obydwie składowe są w przypadku cząstki swobodnej nierozróżnialne. Różnice pojawiają się dopiero, gdy elektron porusza się w polu elektromagnetycznym⁵ [14].

Schrödinger sądził, że wielkość $e\Psi^*(\vec{r}, t)\Psi(\vec{r}, t)$ opisuje gęstość ładunku elektronu. Rozumiał on, że ładunek ulega rozmyciu w przestrzeni. Interpretację probabilistyczną funkcji falowej Ψ podał w 1926 roku Max Born. Według niego funkcja falowa $\Psi(\vec{r}, t)$ opisuje falę prawdopodobieństwa. Wielkość $\Psi^*(\vec{r}, t)\Psi(\vec{r}, t)$ jest gęstością prawdopodobieństwa znalezienia cząstki w punkcie \vec{r} w chwili czasu t [14].

⁵ Zewnętrzne pole elektryczne również powoduje rozszczepienie wiązki (tak jak robi to pole magnetyczne w doświadczeniu Sterna-Gerlacha), ale jest to niewielki efekt relatywistyczny (za [14]).

2.2.2. Równanie mechaniki relatywistycznej - Równanie Kleina-Gordona i równanie Diraca

W celu uzyskania równań opisujących zachowanie się cząstek w świecie relatywistycznym wychodzi się od relatywistycznego związku Einsteina dla energii całkowitej cząstki swobodnej postaci:

$$\mathcal{H} = \sqrt{\vec{p}^2 c^2 + m_0^2 c^4}. \quad (2.6)$$

Zastępując energię \mathcal{H} oraz składowe pędu \vec{p} operatorami zgodnie ze wzorem (2.5) otrzymuje się:

$$i\hbar \frac{\partial}{\partial t} \Psi = \sqrt{-c^2 \hbar^2 \nabla^2 + m_0^2 c^4} \Psi. \quad (2.7)$$

Podejście to zawiodło, nie dając poprawnych rezultatów. Zmienne przestrzenne i czas nie są tu traktowane jednakowo.

Symetria zmiennych przestrzennych i czasu została uzyskana dwoma sposobami. Pierwszy sposób doprowadził do równania Kleina-Gordona, drugi do równania Diraca.

Wychodząc raz jeszcze od związku na energię i pęd, ale uwzględniając kwadrat wartości energii \mathcal{H}^2 , a nie samą wartość energii \mathcal{H} , uzyskuje się równanie uwzględniające poprawki relatywistyczne - równanie Kleina-Gordona dla cząstek o spinie zero.

Równanie Kleina-Gordona

Relatywistyczny związek między kwadratem energii i pędem (dla cząstki swobodnej) ma postać:

$$\mathcal{H}^2 = \vec{p}^2 c^2 + m_0^2 c^4. \quad (2.8)$$

Zastępując energię \mathcal{H} oraz składowe pędu \vec{p} operatorami zgodnie z (2.5) otrzymuje się równanie *Kleina-Gordona* opisujące cząstki zwane bozonami (spin zero) w postaci (2.2). Równanie to posiada symetrię między zmiennymi przestrzennymi i czasem. Najprostszym rozwiązaniem tego równania jest fala płaska $\Psi = \exp i(\vec{k} \cdot \vec{r} - \omega t)$ dająca relatywistyczną zależność na energię:

$$E_{1,2} = \pm \sqrt{\vec{p}^2 c^2 + m_0^2 c^4}. \quad (2.9)$$

Rozwiązanie z ujemną energią dla równania Kleina-Gordona nie posiada interpretacji fizycznej. Funkcja falowa Ψ jest tu funkcją skalarną (jednoskładnikową).

Równanie Diraca

Równanie Kleina-Gordona prowadziło do ujemnych wartości prawdopodobieństw. Było to konsekwencją tego, że funkcja falowa spełniała równanie różniczkowe drugiego rzędu względem czasu. Rozwiązanie polegało na zastąpieniu równania falowego drugiego rzędu równaniem z pierwszą pochodną względem czasu (podobnie jak w nierelatywistycznym równaniu Schroödingera).

Wychodząc od relatywistycznego związku (2.6) między energią i pędem Paul Dirac zastanawiał się, w jaki sposób usunąć w tym równaniu pierwiastek kwadratowy, tak by zachować symetrię między zmiennymi przestrzennymi i czasem. Zażądał by wielkość pod pierwiastkiem równania (2.6) można było zapisać jako kwadrat pewnej wielkości liniowej

względem pędu. Możliwe było wówczas wykonanie pierwiastkowania w celu otrzymania równania pierwszego rzędu względem czasu i współrzędnych przestrzennych. Zakładając, że pęd może przyjmować wszystkie wartości od 0 do ∞ w granicy $p = 0$ otrzymano:

$$\sqrt{\vec{p}^2 c^2 + m_0^2 c^4} \rightarrow m_0 c^2, \quad (2.10)$$

natomiast w przypadku, gdy $m_0 = 0$ uzyskano:

$$\sqrt{\vec{p}^2 c^2 + m_0^2 c^4} \rightarrow \vec{p} c. \quad (2.11)$$

Dla przestrzeni jednowymiarowej otrzymał on wzór na energię w postaci:

$$E = \sqrt{p_x^2 c^2 + m_0^2 c^4} = \alpha c p_x + \beta m_0 c^2. \quad (2.12)$$

W przypadku, gdy p_x i m_0 są różne od zera, związek ten jest spełniony, gdy α i β są macierzami. Podnosząc obie strony równania (2.12) do kwadratu (i pamiętając, że macierze na ogół nie komutują, czyli uważając by zachować odpowiednią kolejność czynników przy wykonywaniu mnożenia) z porównania uzyskuje się warunki na α i β . I tak:

$$p_x^2 c^2 + m_0^2 c^4 = \alpha^2 c^2 p_x^2 + (\alpha\beta + \beta\alpha) m_0 c^3 p_x + \beta^2 m_0^2 c^4 \rightarrow \left\{ \begin{array}{l} \alpha^2 = 1 \\ \beta^2 = 1 \\ \alpha\beta + \beta\alpha = 0 \end{array} \right\}. \quad (2.13)$$

Dla trzy wymiarowej przestrzeni stosowany jest zapis:

$$E = \sqrt{(p_x^2 + p_y^2 + p_z^2) c^2 + m_0^2 c^4} = \alpha_1 c p_x + \alpha_2 c p_y + \alpha_3 c p_z + \beta m_0 c^2, \quad (2.14)$$

który prowadzi do rozwiązań postaci:

$$\alpha_j^2 = 1 \quad (2.15)$$

$$\beta^2 = 1 \quad (2.16)$$

$$\alpha_j \beta + \beta \alpha_j = 0 \quad (2.17)$$

$$\alpha_j \alpha_k + \alpha_k \alpha_j = 0 \quad j \neq k, \quad j, k = 1, 2, 3. \quad (2.18)$$

Można to zapisać wykorzystując zapis z algebry Grassmanna (por. Dodatek 4.2.1):

$$\{\alpha_i, \alpha_j\} = 2\mathbb{I}\delta_{ij} \quad (2.19)$$

$$\beta^2 = \mathbb{I} \quad (2.20)$$

$$\{\alpha_i, \beta\} = 0, \quad (2.21)$$

gdzie δ_{ij} jest deltą Kroneckera (o wartości równej jeden dla $i = j$; zero dla $i \neq j$), \mathbb{I} jest macierzą jednostkową a nawiasy $\{, \}$ są nawiasami Poissona odpowiadającymi za operację $\{a, b\} = ab - ba$.

W podejściu Diraca postaci α nie są określone jednoznacznie. Są to macierze o wymiarze

4x4 wybrane w taki sposób, żeby zapewnić poprawność wykonywania działań w równaniu. Związki te są spełnione przez następujące macierze zaproponowane przez Diraca (reprezentacja Pauliego-Diraca):

$$\alpha_j = \begin{pmatrix} 0 & \sigma_j \\ \sigma_j & 0 \end{pmatrix} \quad \beta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \quad (2.22)$$

gdzie wielkość σ_j to macierze spinowe Pauliego⁶ (brane bez czynnika $\frac{\hbar}{2}$).

Zastępując energię \mathcal{H} oraz składowe pędu \vec{p} w (2.14) operatorami zgodnie ze wzorem (2.5) otrzymujemy **równanie Diraca** opisujące cząstki o spinie $\frac{1}{2}$ (fermiony) postaci:

$$i\hbar \frac{\partial}{\partial t} \Psi = (\alpha_1 c p_x + \alpha_2 c p_y + \alpha_3 c p_z + \beta m_0 c^2) \Psi, \quad (2.24)$$

co można przepisać jako:

$$i\hbar \frac{\partial}{\partial t} \Psi = c(\vec{\alpha} \mathbf{p} + \beta m_0 c) \Psi. \quad (2.25)$$

Równanie Diraca można przedstawić w prostszej formule. Zapisując czas i współrzędne przestrzenne w postaci czterowektora czasoprzestrzeni, gdzie $x^0 = ct$ oraz mnożąc obie strony równania (2.24) lewostronnie przez $\gamma_0 = \beta$ można wprowadzić nowe macierze:

$$\gamma_j = \beta \alpha_j \quad \text{dla } j = 1, 2, 3. \quad (2.26)$$

Tym sposobem otrzymuje się nową postać **równania Diraca**:

$$i\hbar \left(\gamma^0 \frac{\partial}{\partial x^0} + \gamma^1 \frac{\partial}{\partial x^1} + \gamma^2 \frac{\partial}{\partial x^2} + \gamma^3 \frac{\partial}{\partial x^3} \right) \Psi = m_0 c \Psi, \quad (2.27)$$

gdzie $x_0 = ct, x_1 = x, x_2 = y, x_3 = z$. Macierze γ^0 i γ^j to odpowiednio:

$$\gamma^0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \gamma^j = \begin{pmatrix} 0 & \sigma_j \\ -\sigma_j & 0 \end{pmatrix}, \quad (2.28)$$

gdzie σ^j oznaczają, tak jak i wcześniej, macierze spinowe Pauliego. Macierze γ^μ spełniają związki antyprzemienności:

$$\frac{1}{2}(\gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu) = \eta^{\mu\nu} = \begin{cases} +1 & \mu = \nu = 1, 2, 3 \\ -1 & \mu = \nu = 0 \\ 0 & \mu \neq \nu \end{cases}. \quad (2.29)$$

Dirac zauważył, że powyższe związki antyprzemienności są lorentzowsko niezmiennicze i spełniają je także macierze $\Lambda_\nu^\mu \gamma^\nu$, gdzie Λ jest dowolnym przekształceniem Lorentza.

⁶ Macierze spinowe zaproponowane przez Pauliego to:

$$\hat{s}_x = \frac{\hbar}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \hat{s}_y = \frac{\hbar}{2} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \hat{s}_z = \frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (2.23)$$

Ostatecznie otrzymano **relatywistyczne niezmiennicze równanie Diraca dla cząstki swobodnej** postać:

$$\left(\gamma^\mu \frac{\partial}{\partial x^\mu} + m_0\right)\Psi = 0, \quad (2.30)$$

przy oznaczeniu $\gamma^\mu = (\gamma^0, \gamma^j)$.

Aby zbadać zachowanie elektronów w dowolnym zewnętrznym polu elektromagnetycznym Dirac rozszerzył równanie o wpływ pola elektromagnetycznego o potencjały (\mathbf{A}, ϕ) , posługując się procedurą polegającą na wykonaniu podstawień:

$$i\hbar \frac{\partial}{\partial t} \rightarrow i\hbar \frac{\partial}{\partial t} + e\phi \quad -i\hbar \nabla \rightarrow -i\hbar \nabla + \frac{e}{c} \mathbf{A}. \quad (2.31)$$

Równanie falowe $i\hbar \frac{\partial \psi}{\partial t} = \mathcal{H}\psi$ przyjmuje wówczas postać:

$$(i\hbar \frac{\partial}{\partial t} + e\phi)\psi = (-i\hbar c \nabla + e\mathbf{A}) \cdot \alpha \psi + mc^2 \beta \psi. \quad (2.32)$$

Wnioski:

Istnienie spinu i momentu magnetycznego elektronu: Z równania (2.32) przy dokładności rozwinięcia równania Diraca rzędu $(v/c)^2$ oraz przy założeniu $\phi = 0$ opisuje ono cząstkę poprzez funkcję falową o dwóch składowych i dla tej cząstki $g = 2$ ⁷. Funkcja falowa ma wówczas postać:

$$\Psi = \begin{bmatrix} \chi \\ \Phi \end{bmatrix}, \quad (2.33)$$

gdzie χ i Φ są spinorami o dwóch składowych.

Struktura subtelna poziomów wodoru: Z kolei przy dokładności rozwinięcia w równaniu Diraca rzędu $(v/c)^4$ równanie to daje poprawny opis struktury subtelnej. Zakłada się tu postać potencjału dla elektronu w atomie wodoru jako:

$$V = e\phi = -e^2/4\pi\epsilon_0 r.$$

Jeden ze składników rozwinięcia równania interpretuje się jako *poprawkę relatywistyczną do energii kinetycznej*, kolejny składnik opisuje oddziaływanie spinowo-orbitalne otrzymując *czynnik Thomasa*. Dalszy składnik rozwinięcia (przy uwzględnieniu kilku dodatkowych poprawek) to *składnik Darwina*, który ma znaczenie tylko dla stanów s . Rozwijając energię poziomów w kolejnych rzędach przybliżenia dla potencjału kulombowskiego względem α , dostaje się w kolejnych rzędach energię spoczynkową, energię schrödingerską oraz poprawkę związaną ze strukturą subtelną dające bardzo dobrą zgodność z doświadczeniem. Poprawki te, tłumaczą również wynik różnicy pomiędzy poziomem $2S_{1/2}$, który leży nieco wyżej niż $2P_{1/2}$. Jest to tzw. przesunięcie Lambda.

Rozwiązanie równania Diraca

Równanie Diraca najłatwiej rozwiązuje się dla jednej cząstki. Dla układów złożonych (atomy i molekuly) stosuje się różne metody przybliżeń na przykład przybliżenie pola

⁷ Wielkość fizyczna określająca stosunek orbitalnego lub spinowego momentu pędu elektronu do jego momentu magnetycznego. Postać doświadczalna g jest bardzo bliska 2.

średniego oraz zakłada się, że całkowity hamiltonian jest sumą hamiltonianów Diraca dla poszczególnych cząstek plus columbowskie oddziaływanie cząstek. Są też inne metody szerzej opisane w [15].

Równanie to daje się analitycznie rozwiązać tylko w kilku bardzo prostych przypadkach. Pierwszym jest cząstka swobodna, dla której rozwiązanie przedstawił sam Paul Dirac, a drugim atom wodoropodobny - rozwiązanie przedstawione przez Charles Galton Darwin. Ponieważ macierze α (i dalej γ) mają wymiar 4×4 to rozwiązania równania Diraca stanowią układ 4 równań różniczkowych cząstkowych pierwszego rzędu. Aby taki układ rozwiązać trzeba go rozprzęgnąć. Rozwiązaniem równania jest wektor czterokomponentowy bi-spinor (podwójny spinor), czyli w ogólności zespolona funkcja falowa o czterech składowych $\Psi(\vec{r}, t)$. Nosi ona nazwę spinora Lorentza [11]. Znaczenie czterech składowych Ψ dla cząstki spoczywającej jest związane z dwiema możliwymi orientacjami spinu i dwiema wartościami.

Równanie Diraca dopuszcza dla cząstki swobodnej rozwiązania o energiach zarówno dodatnich, jak i o ujemnych o energiach odpowiednio (2.9), podobnie jak w rozwiązaniu równania Kleina-Gordona (widma energetyczne rozwiązań obu równań są takie same).

W fizyce klasycznej zakłada się, że tylko rozwiązanie o dodatniej energii występuje w naturze, natomiast rozwiązanie o ujemnej energii nie występuje. Jednak równanie Diraca postuluje istnienie procesów, w których ujemne rozwiązanie może powstać z *normalnych* dodatnich cząstek, wobec czego nie można go ignorować.

2.3. Konsekwencje wynikające z równania Diraca

2.3.1. Równanie Diraca przewiduje istnienie antycząstek

Konsekwencją równania Diraca jest informacja, że istnieją inne cząstki o właściwościach identycznych z elektronem, różniące się od niej tylko ładunkiem elektrycznym, czyli antycząstki. Dało to podejrzenie o istnienie antymaterii, co też zostało potwierdzone eksperymentalnie w późniejszych latach. Po raz pierwszy pozyton został zaobserwowany w 1932 roku⁸ przez Carla Andersona w komorze mgłowej.

Carl Anderson opublikował pracę mówiącą o odkryciu pozytonu w czasopiśmie "Science". W swojej naukowej pracy zajmował się promieniowaniem kosmicznym. Zbudowana przez niego komora Wilsona znajdowała się w najsilniejszym polu magnetycznym, możliwym wówczas dla niego do uzyskania (do 2,5 tesli). W swoim doświadczeniu umieścił on wewnątrz komory płytkę metalową, efektem czego cząstki po przejściu przez nią wytrącały energię (prędkość) zakrzywiając tor ruchu (po wyjściu z płytki ich krzywizna była większa niż przed wejściem - na podstawie tych informacji można dowodzić kierunku biegu cząstki oraz jej ładunku). Na podstawie interpretacji jednego z uzyskanych zdjęć zaobserwował

⁸ Rok 1932 nazwany został za rokiem 1913 (ogromny postęp w zrozumieniu budowy atomu) oraz rokiem 1920 (odkrycia Oersteda i Ampera prowadzące do powstania elektrodynamiki) CUDOWNYM rokiem 1932, w którym dokonano nagłego postępu w bardzo krótkim czasie. W roku tym dokonano kilku przełomowych odkryć o fizyce wnętrza atomu (odkryto neutron, pozyton oraz ciężki wodór tzw. deuter, przeprowadzono pierwszą reakcję jądrową oraz uzyskano energię przekraczającą miliony eV) [12].

śląd cząstki o masie zbliżonej do masy elektronu oraz o dodatnim ładunku co pozwoliło mu przypuszczać, że zaobserwowana cząstka jest pozytonem (antycząstką dla elektronu)⁹. Odkrycie to jest potwierdzeniem istnienia antycząstek wynikającego z równania Diraca.

2.3.2. Interpretacja ujemnych rozwiązań równania Diraca

Jak interpretować rozwiązania o energii ujemnej? Przy istnieniu oddziaływań zewnętrznych w otoczeniu elektronu o energiach dodatnich powinny przejść do stanów o energiach ujemnych. Istnieją jednak elektrony o energiach dodatnich i nad pochodzeniem ich trwałości zastanawiali się zarówno Dirac i Feynman, którzy zapostulowali swoje eksperymenty myślowe jako wyjaśnienia obserwowanego zjawiska. Dirac interpretował pozytron jako dziurę w tzw. morzu Diraca. Richard Feynman rozważał go natomiast jako cząstkę poruszającą się do tyłu w czasie.

Morze Diraca - teoria dziur

Dirac zakładał, że wszystkie stany o energiach ujemnych są już całkowicie obsadzone przez elektrony o ujemnych energiach. Obszar ten stanowi tzw. morze Diraca, które jest niewidoczne. Z uwagi na fakt, że wszystkie stany w morzu są już obsadzone oraz na stosowalność zakazu Pauliego, niemożliwe jest przejście elektronu o dodatniej energii do obsadzonych już stanów o energii ujemnej. Jeśli elektronowi o ujemnej energii dostarczy się wystarczająco dużo energii by mógł przejść do stanu o energii dodatniej wówczas cząstką ma energię dodatnią i ładunek $-e$. Powstaje też 'dziura' w morzu Diraca, która jest już widoczna. Luka ta ma dodatnią energię oraz ładunek $+e$. Luka ta została utożsamiona z pozytonem.

Gdy elektron napotyka pozytron (analog dziury w morzu Diraca, wskakuje do niej) ulega on anihilacji uwalniając energię w postaci fotonów.

Równanie Diraca nie stosuje się do cząstek o spinie zero (gdzie nie obowiązuje zakaz Pauliego). Koncepcja Feynmana stosuje się do fermionów i bozonów.

Koncepcja Feynmana - ruch wstecz w czasie

Formalizm ten umożliwia opis tworzenia się cząstek (lub par cząstka-antycząstka). Koncepcja Feynmana traktuje cząstki o energii ujemnej jako cząstki poruszające się wstecz w czasie. Opisuje to operator ewolucji Feynmana. Elektron może oddziaływać z polem dowolną ilość razy a stany pośrednie mogą zawierać dowolną liczbę par e^+e^- , zwiększając tym liczbę stopni swobody układu. Zaczynając od równania dla jednej cząstki otrzymuje się rozwiązania o ujemnej energii, które opisują tworzenie się nowych cząstek. Interpretacja wygląda tak, że z istnienia rozwiązań o ujemnej energii wynika istnienie nieskończonego morza cząstek 'uśpionych', dzięki którym pojedynczy elektron może poruszać się na przemian w przód i w tył w czasie, stając się w pewnych przedziałach czasu wieloma cząstkami.

⁹ Doświadczenie, w którym zaobserwował 13 pozytronów liczyło 1300 zdjęć cząstek kosmicznych [12].

2.4. Podsumowanie

W rozdziale została przedstawiona historyczna konstrukcja relatywistycznego równania Diraca. Opisano tu również wnioski płynące z rozwiązań tego równania dla współczesnej fizyki kwantowej i atomowej.

Konsekwencje sformułowania równania Diraca sięgają powstania koncepcji kwantowej teorii pola (QFT).

Rozdział 3

Metodyka EFI w teorii pola

W tym miejscu przechodzę do zasadniczej części opracowania, którym jest wyprowadzenie relatywistycznego równania Diraca dla cząstki swobodnej, wychodząc od statystycznego opisu zjawiska fizycznego. B. Roy Frieden i B.H. Soffer wyprowadzili modele falowe posługując się pojęciem *pojemności informacyjnej* I oraz *zasadami informacyjnymi* estymacyjnej metody ekstremalnej fizycznej informacji EFI [28].

Według koncepcji Friedena i Soffera granica między teoriami przebiega nie ze względu na podział świata na mikro i makro, ale na pochodzenie statystyczne (inaczej określone jako pochodzenie falowe (bądź polowe)) i klasyczno-mechaniczne (pochodzenie ściśle punktowe).

Aby otrzymać teorię pola metoda EFI zasadniczo używa trzech zasad. Podejście tu przedstawione, bazujące na wynikach pracy [26], korzysta z wariacyjnej zasady informacyjnej (3.17), która minimalizuje całkowitą fizyczną informację układu oraz obserwowanej zasady strukturalnej (3.16) i oczekiwanej zasady strukturalna (tak zwanej całkowitej zasady strukturalnej) (3.18), które tą informację zerują.

Praca ta opisuje (za [26]) aparat pojęciowy informacji kinetycznej i strukturalnej wykorzystywany w metodzie EFI, które to posłużą wyprowadzeniu równania Diraca.

3.1. Strukturalne zasady informacyjne

Estymacja fizyczna, będąca tematem pracy, wychodzi od zastosowania zasad informacyjnych na funkcję wiarygodności $P(\Theta)$, gdzie parametr $\Theta \equiv (\theta_n)_{n=1}^N \in V_\Theta$ jest zbiorem wartości oczekiwanych zmiennej losowej położenia układu w N punktach pomiarowych. Wychodząc od $\ln \frac{P(\tilde{\Theta})}{P(\Theta)}$ i rozwijając w punkcie $\tilde{\Theta}$ funkcję $\ln P(\tilde{\Theta})$ w szereg Taylora wokół prawdziwej wartości Θ można (za [26]) zdefiniować *obserwowaną strukturę* układu \mathbf{tF} oraz zauważyć, że pojawia się tu *obserwowana macierz informacyjna Fishera* \mathbf{iF} , która jest symetryczna i dodatnio określona. $\tilde{\Theta} \equiv (\tilde{\theta}_n)_{n=1}^N \in V_\Theta$ jest wartością estymatora $\hat{\Theta}$ parametru Θ .

Wyłania się tu postać **strukturalnego równania macierzowego** jako sumy:

$$\mathbf{qF} + \mathbf{iF} = 0, \quad (3.1)$$

gdzie **obserwowana macierz informacyjna Fishera** to:

$$\mathbf{iF} = \left(- \frac{\partial^2 \ln P(\Theta)}{\partial \theta_{n'} \partial \theta_n} \right), \quad (3.2)$$

a \mathbf{qF} jest, jak już było nazwane, **obserwowaną macierzą struktury**.

Wprowadzono tu dwie zasady informacyjne: *strukturalną zasadę informacyjną* (*obserwowaną i oczekiwaną* (zw też. całkową)) oraz *zasadę wariacyjną*, które wchodzą w skład metody estymacji statystycznej EFI.

Twierdzenie 3.1.1. (Obserwowana strukturalna zasada informacyjna) *Sumując wszystkie elementy obserwowanej macierzy informacyjnej Fishera \mathbf{iF} oraz obserwowanej macierzy struktury \mathbf{qF} w równaniu (3.1) otrzymuje się postać obserwowaną strukturalnej zasady informacyjnej Frieden'a:*

$$\sum_{n,n'=1}^N (\mathbf{iF})_{nn'} + \kappa \sum_{n,n'=1}^N (\mathbf{qF})_{nn'} = 0. \quad (3.3)$$

Postać oczekiwana (całkowa strukturalna zasada informacyjna) ma postać:

$$I + \kappa Q = 0, \quad (3.4)$$

gdzie I to uogólniona pojemność informacyjna Fishera (3.5), Q to strukturalna informacja (SI) (3.6), κ jest współczynnikiem efektywności [28], do którego jeszcze powrócę przy (3.13).

Dla uogólnionej postaci **pojemności informacyjnej I Fishera**, gdzie całkowanie odbywa się po całej przestrzeni prób \mathfrak{B} lub na jej podprzestrzeni z miarą $dyP(\Theta)$, jej wartość to:

$$I = \int_{\mathfrak{B}} dy P(\Theta) \sum_{n,n'=1}^N (\mathbf{iF})_{nn'}. \quad (3.5)$$

Dla zmiennych Y_n parami niezależnych \mathbf{iF} jest diagonalna tzn. $(\mathbf{iF})_{nn'} = \delta_{nn'} \mathbf{iF}_{nn} \equiv \mathbf{iF}_n$, gdzie $\delta_{nn'}$ jest deltą Kroneckera. Możemy więc wprowadzić definicję gęstości pojemności informacyjnej \mathbf{i} jako:

$$I = \int_{\mathfrak{B}} dy \mathbf{i} = \int_{\mathfrak{B}} dy P(\Theta) \sum_{n=1}^N \mathbf{iF}_n, \quad \mathbf{i} = P(\Theta) \sum_{n=1}^N \mathbf{iF}_n. \quad (3.6)$$

Q to z kolei **informacja strukturalna (SI)**, którą można powiązać z gęstością informacji strukturalnej \mathbf{q} :

$$Q = \int_{\mathfrak{B}} dy P(\Theta) \sum_{n,n'=1}^N (\mathbf{qF})_{nn'}, \quad (3.7)$$

Postać SI wyrażona w amplitudach q_n dla zmiennych Y_n parami niezależnych ma postać:

$$Q = \int_{\mathfrak{B}} dy \mathbf{q} = \int_{\mathfrak{B}} dy P(\Theta) \sum_{n=1}^N \mathbf{qF}_n(q_n(\mathbf{y}_n)), \quad \mathbf{q} := P(\Theta) \sum_{n=1}^N \mathbf{qF}_n(q_n(\mathbf{y}_n)). \quad (3.8)$$

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

$$Q = \sum_{n=1}^N \int d^4 \mathbf{y}_n p_n(\mathbf{y}_n | \theta_n) \mathbf{qF}_n(q_n(\mathbf{y}_n)), \quad (3.9)$$

lub przepisać w przesunięciach \mathbf{x}_n (przejście do przestrzeni przesunięć zostanie przedstawione na stronie 43).

$$Q = \sum_{n=1}^N \int_{\chi} d^4 \mathbf{x} q_n^2(\mathbf{x}) \mathbf{q} F_n(q_n(\mathbf{x})), \quad (3.10)$$

bądź w przypadku pola ¹ $\psi(\mathbf{x})$ (przejście do przestrzeni funkcji falowej opisano na stronie 44):

$$Q \equiv Q_{\psi} = \int_{\chi} d^4 \mathbf{x} \sum_{n,n'=1}^{N/2} \psi_n^*(\mathbf{x}) \psi_{n'}(\mathbf{x}) \mathbf{q} F_{nn'}^{\psi}(\psi(\mathbf{x}), \psi^*(\mathbf{x}), \psi^{(l)}(\mathbf{x}), \psi^{*(l)}(\mathbf{x})). \quad (3.11)$$

Tym samym można zapisać *obserwowaną informacyjną zasadę strukturalną zapisaną w gęstościach* jako:

$$\mathbf{i} + \kappa \mathbf{q} = 0. \quad (3.12)$$

Współczynnik κ określamy jako **współczynnik efektywności**, który zasadniczo przyjmuje dwie wartości (za [25, 28]):

$$\kappa = 1 \quad \vee \quad \frac{1}{2}. \quad (3.13)$$

Obserwowana strukturalna zasada informacyjna jest równaniem strukturalnym współczesnych modeli fizycznych wyprowadzanych metodą EFI.

Pojemność informacyjna Fishera I odpowiada za *postać kinematyczną*, natomiast informacja *postać strukturalna* Q będzie zależna od fizycznych więzów nałożonych na układ.

Twierdzenie 3.1.2. (Całkowita (totalna) fizyczna informacja (TFI/TPI))

Całkowita fizyczna informacja K , która jest nieujemna i addytywna, przyjmuje postać:

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

gdzie Q jest informacją strukturalną SI zawartą w strukturalnych stopniach swobody, a I jest pojemnością informacyjną układu zawartą w kinematycznych stopniach swobody.

UWAGA: Podczas pomiaru (którego dokonuje sam układ) ma miejsce transfer informacji (TI) z $J \geq 0$, gdzie J jest dokonany transfer informacji i $\delta Q = Q' - Q \leq 0$ oraz $\delta I = I' - I \geq 0$, gdzie Q' oraz I' są wielkościami po pomiarze. W pomiarze idealnym $\delta I = -\delta Q$, skąd $K' = K$. Informacja fizyczna TFI pozostaje w tym przypadku niezmienną. Na poziomie próbkowania czasoprzestrzeni przez sam układ otrzymuje się *wariacyjną zasadę informacji*.

Twierdzenie 3.1.3. (Wariacyjna zasada informacyjna) Skalarna postać zasady wariacyjnej ma postać:

$$\delta K = \delta(I + Q) = 0 \Rightarrow K = I + Q \text{ jest ekstremalne.} \quad (3.15)$$

Zasada wariacyjna, wykorzystana w metodzie EFI, prowadzi do równań Eulera-Lagrange'a. Z kolei postać obserwowana, zasada wariacyjna oraz postać oczekiwana są podstawą estymacji EFI równań ruchu teorii pola.

¹ Całą funkcja podcałkowa jest wielomianem pól $\psi(\mathbf{x})$ oraz $\psi^*(\mathbf{x})$, stopnia nie mniejszego niż 2, oraz ich pochodnych rzędu $l = 1, 2, \dots$, natomiast $\mathbf{q} F_{nn'}^{\psi}$ jest pewną obserwowaną, w ogólności zespoloną, informacją strukturalną układu.

3.2. Zmodyfikowane zasady strukturalne

Obie zasady, obserwowana zmodyfikowana zasada strukturalna oraz zmodyfikowana zasada wariacyjna są podstawą metody estymacyjnej EFI, dając rozwiązanie dla amplitud i prowadząc do dobrze znanych modeli np. teorii pola.

Definicja 3.2.1. (Zmodyfikowana obserwowana zasada strukturalna)

Zmodyfikowana obserwowana zasada strukturalna zapisana w gęstościach jest następująca:

$$\tilde{i}' + \tilde{\mathbf{C}} + \kappa q = 0, \text{ przy czym } I = \int_{\mathfrak{B}} dy \tilde{i} = \int_{\mathfrak{B}} dy (\tilde{i}' + \tilde{\mathbf{C}}), \quad (3.16)$$

gdzie $\tilde{\mathbf{C}}$ jest pochodną zupełną, która wynika z całkowania przez części całki $I = \int_{\mathfrak{B}} dy \tilde{i}$.

Zmodyfikowana informacyjna zasada wariacyjna ma postać:

Definicja 3.2.2. (Zmodyfikowana zasada wariacyjna)

$$\delta(I + Q) = \delta \int_{\mathfrak{B}} dy (\tilde{i} + q) = 0 \quad (3.17)$$

Definicja 3.2.3. (Zmodyfikowana oczekiwana zasada strukturalna)

$$\int_{\mathfrak{B}} dy (\tilde{i}' + \tilde{\mathbf{C}} + \kappa q) = 0 \Leftrightarrow \int_{\mathfrak{B}} dy (i + \kappa q) = 0 \quad (3.18)$$

bo na poziomie oczekiwanym pojemność informacyjna $\tilde{\mathbf{i}} := P(\Theta) \sum_{n,n'=1}^N \tilde{\mathbf{i}} \tilde{\mathbf{F}}_{n,n'}$ oraz $\mathbf{i} := P(\Theta) \sum_{n=1}^N \mathbf{i} \mathbf{F}_n$ spełniają zależność: $I = \int_{\mathfrak{B}} dy \tilde{\mathbf{i}} = \int_{\mathfrak{B}} dy \mathbf{i}$.

3.3. Reprezentacje kinematyczne pojemności informacyjnej I

W celu estymacji parametrów modelu teorii pola, zgodnie ze strukturalnymi zasadami informacyjnymi nałożonymi na fizyczną informację układu, wprowadzono pojemność informacyjną I w różnych reprezentacjach (w prawdopodobieństwach $p_n(\mathbf{y}_n|\theta_n)$, w prawdopodobieństwach przesunięć $p_n(\mathbf{x})$), w amplitudach $q_n(\mathbf{y}_n|\theta_n)$ oraz $q_n(\mathbf{x})$ oraz w funkcji falowej $\psi_n(\mathbf{x})$).

Zacznę od postaci amplitudowej $p_n(\mathbf{y}_n|\theta_n)$ pojemności informacyjnej I układu. Można tu rozważyć zmienną losową \mathbf{Y} położenia układu przyjmującą wartość \mathbf{y} będące punktem zbioru \mathfrak{Y} . W rozważaniach mechaniki falowej jest to punkt czasoprzestrzenny przestrzeni Minkowskiego. Wówczas $\mathbf{y} \equiv (\mathbf{y}^\nu)_{\nu=0}^3 \in \mathfrak{Y} \equiv \mathbb{R}^4$ są realizacjami rozkładu $p(\mathbf{y})$.

Dane $y \equiv (\mathbf{y}_n)_{n=1}^N = (\mathbf{y}_1, \dots, \mathbf{y}_N)$ są realizacjami próby dla położenia układu, gdzie zapisano $\mathbf{y}_n \equiv (\mathbf{y}_n^\nu)_{\nu=0}^3$. Frieden i Soffer [28] założyli, że zbieranie danych y (z dostępnej czasoprzestrzeni) następuje przez sam układ, w zgodzie z rozkładami gęstości prawdopodobieństwa $p_n(\mathbf{y}_n|\theta_n)$, gdzie $n = 1, \dots, N$.

Bazując na funkcji wiarygodności $P(y|\Theta)$ postaci:

$$P(\Theta) \equiv P(y|\Theta) = \prod_{n=1}^N p_n(\mathbf{y}_n|\Theta) = \prod_{n=1}^N p_n(y_n|\theta_n), \quad (3.19)$$

gdzie założono, że dane zbierane przez układ $y = (\mathbf{y}_1, \dots, \mathbf{y}_N)$ są uzyskane niezależnie i łączny rozkład prawdopodobieństwa dla próby faktoryzuje się na rozkłady brzegowe.

Przejście od pojemności informacyjnej I do tzw. postaci kinematycznej wykorzystywanej w teorii pola oraz w fizyce statystycznej polega na wyrażeniu funkcji prawdopodobieństw poprzez jej amplitudy $p_n(\mathbf{y}_n|\theta_n)$. Pojemność informacyjną I postaci:

$$I = \sum_{n=1}^N I_n = \sum_{n=1}^N \int_{\mathfrak{B}} d^4 \mathbf{y}_n P(y|\Theta) \sum_{\nu=0}^3 \left(\frac{\partial \ln P(y|\Theta)}{\partial \theta_{n\nu}} \frac{\partial \ln P(y|\Theta)}{\partial \theta_n^\nu} \right) \quad (3.20)$$

wykorzystując przekształcenie pochodnej $\ln P$:

$$\frac{\partial \ln P(y|\Theta)}{\partial \theta_{n\nu}} = \frac{\partial}{\partial \theta_{n\nu}} \sum_{n=1}^N \ln p_n(\mathbf{y}_n|\theta_n) = \sum_{n=1}^N \frac{1}{p_n(\mathbf{y}_n|\theta_n)} \left(\frac{\partial p_n(\mathbf{y}_n|\theta_n)}{\partial \theta_{n\nu}} \frac{\partial p_n(\mathbf{y}_n|\theta_n)}{\partial \theta_n^\nu} \right) \quad (3.21)$$

można sprowadzić rachunkiem do postaci pojemności informacyjnej wyrażonej przez $p_n(\mathbf{y}_n|\theta_n)$.

Definicja 3.3.1. (I wyrażona w prawdopodobieństwach $p_n(\mathbf{y}_n|\theta_n)$) Postać wyrażona w (punktowych) prawdopodobieństwach $p_n(\mathbf{y}_n|\theta_n)$:

$$I = \sum_{n=1}^N \int_{\mathfrak{Y}} d^4 \mathbf{y}_n \frac{1}{p_n(\mathbf{y}_n|\theta_n)} \sum_{\nu=0}^3 \left(\frac{\partial p_n(\mathbf{y}_n|\theta_n)}{\partial \theta_{n\nu}} \frac{\partial p_n(\mathbf{y}_n|\theta_n)}{\partial \theta_n^\nu} \right), \quad (3.22)$$

Przedstawię teraz przejście od przestrzeni statystycznej \mathcal{S} z bazą $(\theta_n)_{n=1}^N$ dla reprezentacji danych pomiarowych $\mathbf{y}_n \in \mathfrak{Y}$, do przestrzeni przesunięć $\mathbf{x}_n := \mathbf{y}_n - \theta_n$ określonych na przestrzeni bazowej χ .

Skorzystano tu z przejścia do **addytywnych przesunięć**² $\mathbf{x}_n \equiv (\mathbf{x}_n^\nu)$ oraz reguły łańcuchowej dla pochodnej³ opisaney w [26] oraz zakładając, że zakres zmienności wszystkich \mathbf{x}_n^u jest dla każdego n taki sam.

Postać pojemności informacyjnej wyrażonej w prawdopodobieństwach $p_n(\mathbf{x})$ ma postać:

Definicja 3.3.2. (I wyrażona w prawdopodobieństwach $p_n(\mathbf{x})$)

$$I = \sum_{n=1}^N \int_{\chi} d^4 \mathbf{x} \frac{1}{p_n(\mathbf{x})} \sum_{\nu=0}^3 \left(\frac{\partial p_n(\mathbf{x})}{\partial \theta_{n\nu}} \frac{\partial p_n(\mathbf{x})}{\partial \theta_n^\nu} \right). \quad (3.25)$$

Korzystając dodatkowo z zależności amplitudowej $q_n(\mathbf{y}_n|\theta_n)$ określonej jako:

$$p_n(\mathbf{y}_n|\theta_n) = q_n^2(\mathbf{y}_n|\theta_n), \quad (3.26)$$

² Niech $\mathbf{x}_n \equiv (\mathbf{x}_n^\nu)$ są przesunięciami (addytywne fluktuacje) danych $\mathbf{y}_n \equiv (\mathbf{y}_n^\nu)$ od ich wartości oczekiwanej θ_n^ν postaci:

$$\mathbf{y}_n^\nu = \theta_n^\nu + \mathbf{x}_n^\nu. \quad (3.23)$$

Przesunięcia \mathbf{x}_n^ν są zmiennymi Fisher'owskimi, spełniającymi warunek $\frac{\partial \mathbf{x}_n^\nu}{\partial \mathbf{x}^\mu} = \delta_\mu^\nu$. \mathbf{x}_n^ν są przesunięciami wartości pomiarowych położenia, zebranymi przez układ, od ich wartości oczekiwanych.

³ Reguła łańcuchowa dla pochodnej:

$$\frac{\partial}{\partial \theta_n^\nu} = \frac{\partial(\mathbf{y}_n^\nu - \theta_n^\nu)}{\partial \theta_n^\nu} \frac{\partial}{\partial(\mathbf{y}_n^\nu - \theta_n^\nu)} = - \frac{\partial}{\partial(\mathbf{y}_n^\nu - \theta_n^\nu)} = - \frac{\partial}{\partial \mathbf{x}^\nu} \quad (3.24)$$

lub

$$q_n(\mathbf{y}_n|\theta_n) \equiv \sqrt{p_n(\mathbf{y}_n|\theta_n)} \quad (3.27)$$

uzyskuje się amplitudową postać pojemności informacyjnej.

Definicja 3.3.3. (I wyrażona w amplitudach $q_n(\mathbf{y}_n|\theta_n)$) N -wymiarowa próbna $\mathbf{y}_n \equiv (\mathbf{y}_n^\nu)$ jest pobierana przez układ posiadający rozkład $p_n(\mathbf{y}_n)$, gdzie $n = 1, \dots, N$ to indeks próby oraz $\nu = 0, 1, 2, 3$ - indeks czasoprzestrzenny

$$I \equiv I(\Theta) = 4 \sum_{n=1}^N \int_{\mathfrak{Y}} d^4 \mathbf{y}_n \sum_{\nu=0}^3 \left(\frac{\partial q_n(\mathbf{y}_n|\theta_n)}{\partial \theta_{n\nu}} \frac{\partial q_n(\mathbf{y}_n|\theta_n)}{\partial \theta_n^\nu} \right) \quad (3.28)$$

Przejście do przestrzeni amplitud przesunięć $\mathbf{x}_n := \mathbf{y}_n - \theta_n$ (gdzie wprowadzono oznaczenie $q_n(\mathbf{x}) \equiv q_n(\mathbf{x}_n + \theta_n|\theta_n) = q_n(\mathbf{y}_n|\theta_n)$) określonych na przestrzeni bazowej χ daje:

Definicja 3.3.4. (I wyrażona w amplitudach $q_n(\mathbf{x})$)

$$I = 4 \sum_{n=1}^N \int_{\chi} d^4 \mathbf{x} \sum_{\nu} \left(\frac{\partial q_n(\mathbf{x})}{\partial \mathbf{x}_\nu} \frac{\partial q_n(\mathbf{x})}{\partial \mathbf{x}^\nu} \right) \quad (3.29)$$

Dla mechaniki falowej, część kinematyczną postaci informacji Fishera wyrażamy przez funkcję falową ψ . Dla kolejnej kinematycznej definicji pojemności informacyjnej I , przedstawimy tu konstrukcję funkcji falowej, wykorzystując do tego wprowadzone już amplitudy q_n (powiązane z gęstościami prawdopodobieństw p_n). Postać ta posłuży do wyprowadzenia za [28] równania Diraca.

Pojemność informacyjna I dla ψ będzie miała więc postać zapisaną w definicji poniżej.

Definicja 3.3.5. (I wyrażona poprzez funkcję falową $\psi_n(\mathbf{x})$)

$$I = 4N \sum_{n=1}^{N/2} \int_{\chi} d^4 \mathbf{x} \sum_{\nu} \left(\frac{\partial \psi_n^*(\mathbf{x})}{\partial \mathbf{x}_\nu} \frac{\partial \psi_n(\mathbf{x})}{\partial \mathbf{x}^\nu} \right). \quad (3.30)$$

Wykorzystuję tu za [26] postać funkcji falowej (układu rangi N) zaproponowanej przez Friedena:

$$\psi(\mathbf{x}) \equiv (\psi_n(\mathbf{x}))_{n=1}^{N/2}, \quad (3.31)$$

gdzie założyliśmy, że w poniższym wzorze zachodzi równoważność zmiennych według $\mathbf{x}_n \equiv \mathbf{x}$ dla wszystkich $n = 1, 2, \dots, N$ tak, że:

$$\psi_n(\mathbf{x}_{2n-1}, \mathbf{x}_{2n}) \equiv \frac{1}{\sqrt{N}} (q_{2n-1}(\mathbf{x}_{2n-1}) + i q_{2n}(\mathbf{x}_{2n})), \quad n = 1, 2, \dots, N/2 \quad (3.32)$$

dając nam:

$$\psi_n(\mathbf{x}) \equiv \frac{1}{\sqrt{N}} (q_{2n-1}(\mathbf{x}) + i q_{2n}(\mathbf{x})), \quad n = 1, 2, \dots, N/2. \quad (3.33)$$

Wymiar próby N jest rangą pola układu zdefiniowanego jako zbiór amplitud $(q_n(\mathbf{x}_n))_{n=1}^N$. Z kolei gęstość rozkładu prawdopodobieństwa przesunięcia (lub fluktuacji) w układzie można (za [28]) zapisać jako:

$$p(\mathbf{x}) = \frac{1}{N} \sum_{n=1}^N q_n^2(\mathbf{x}). \quad (3.34)$$

Gęstość rozkładu prawdopodobieństwa w układzie opisanym funkcją falową ψ możemy zapisać jako:

$$p(\mathbf{x}) = \sum_{n=1}^{N/2} \psi_n^*(\mathbf{x}) \psi_n(\mathbf{x}). \quad (3.35)$$

UWAGA: Skorzystano tu z postaci amplitudy $q_n^2 = p_n$. Prawdopodobieństwo p_n jest prawdopodobieństwem warunkowym $p_n(\mathbf{x}_n|\theta_n)$. Postać funkcji $r(\theta_n) = \frac{1}{N}$ jest odzwierciedleniem całkowitego braku wiedzy odnośnie tego, która z N możliwych wartości θ_n pojawi się w konkretnym n -tym z N eksperymentów próby.

Metoda EFI stanowi szczególny typ estymacyjnej procedury statystycznej w ramach teorii pomiaru. Pokazuje ona w jaki sposób, wychodząc od funkcji wiarygodności $P(\Theta)$ oraz pojemności informacyjnej I , posilując się dodatkowo analityczną informacyjną zasadą strukturalną oraz wariacyjną zasadą informacyjną, otrzymać więzy strukturalne prowadzące do równań różniczkowych teorii EFI. Po drodze otrzymując informację strukturalną Q , która wraz z pojemnością informacyjną I stanowi informację fizyczną K układu.

Wszystkie warunki nałożone na układ należy traktować jako związki na odchylenie (fluktuacje) wartości pomiarowych od wartości oczekiwanych i na bazie tego podejścia wyprowadzamy równania ruchu lub równania generujące rozkłady.

Zostało powiedziane (na przykład w [26]), że metoda Friedena-Soffera jest pewną modyfikacją MNW, dzięki której możemy dokonywać nieparametrycznej estymacji równań ruchu teorii pola. Metoda rozwiązywania układu (różniczkowych) zasad informacyjnych EFI poza warunkami brzegowymi (i ewentualnymi równaniami ciągłości) nie jest ograniczona przez żadną konkretną postać rozkładu.

3.4. Znaczenie rangi N i pojemności informacyjnej I

Wielkość próby N pomaga w EFI w klasyfikowaniu modeli fizycznych. Modele możemy dzielić ze względu na rangę N , czyli wymiar (inaczej wielkość) próby oraz kategorię strukturalnej pojemności informacyjnej I (skończoność bądź nieskończoność pojemności informacyjnej I). Ze względu na wartość N dzieli się modele na dwie kategorie: ze skończoną wartością N (związaną ze skończoną I) oraz z nieskończoną wartością N .

Mechanika klasyczna posiada nieskończoną pojemność informacyjną I oraz nieskończoną wartość N . Teoria ta nie posiada struktury statystycznej. Modele mechanik falowej i klasycznych teorii pola należą do kategorii ze skończonym N i skończoną pojemnością I . Zestawiłam to w tabeli poniżej:

ranga N	$I < \infty$	$I \rightarrow \infty$
2	równanie Kleina-Gordona i Schrödingera równania Maxwella równanie Diraca teoria grawitacji	
4		
8		
10		
∞		mechanika klasyczna

Tabela 3.1. Podział modeli ze względu na wielkość próby N i na pojemność informacyjną I

3.5. Rozwiązania metody EFI - równania różniczkowe, jako analogia zasady wariacyjnej oraz równań różniczkowych Eulera-Lagrange'a

Zastosowane zasady informacyjne, strukturalna oraz wariacyjna oraz pojemność informacyjna, będą odwoływać się do miary $d\mathbf{x}p_n(\mathbf{x})$ określonej na przestrzeni przesunięć $\mathbf{x} \in \chi$ jako przestrzeni bazowej, gdzie χ jest czasoprzestrzenią Minkowskiego.

Dla modeli mechaniki falowej i teorii pola korzysta się z kinematycznej postaci strukturalnej pojemności informacyjnej I (3.29) oraz (3.30) (odpowiednio w amplitudach q_n oraz funkcji pola ψ). Postacie te, oparte o informację Fishera, zostaną wykorzystane do konstrukcji równań ruchu (lub równań generujących rozkład) modeli fizycznych.

Całkowita fizyczna informacja TFI (równanie (3.14) z $\kappa = 1$) ma postać $K = Q + I$, pojemność informacyjna ma postać (3.29) i (3.30) a informacja strukturalna Q (3.10) i (3.11).

Definicja 3.5.1. (Gęstość k całkowitej fizycznej informacji TFI) *Całkowitą (totalną) informację fizyczną można zapisać w gęstościach ⁴ jako:*

$$K = \int_{\chi} d^4\mathbf{x}k, \quad (3.36)$$

gdzie dla pola opisanego amplitudami q_n gęstość k to suma pojemności informacyjnej I (3.29) oraz strukturalnej Q (3.10)

$$k = 4 \sum_{n=1}^N \left[\sum_{\nu=0}^3 \frac{\partial q_n(\mathbf{x})}{\partial \mathbf{x}_{\nu}} \frac{\partial q_n(\mathbf{x})}{\partial \mathbf{x}^{\nu}} + \frac{1}{4} q_n^2(\mathbf{x}) q F_n(q_n(\mathbf{x})) \right], \quad (3.37)$$

a dla pola opisanego amplitudami ψ_n jest za (3.30) oraz (3.11) postać:

$$k = 4N \sum_{n,n'=1}^{N/2} \left[\sum_{\nu=0}^3 \delta_{nn'} \frac{\partial \psi_n^*(\mathbf{x})}{\partial \mathbf{x}_{\nu}} \frac{\partial \psi_{n'}(\mathbf{x})}{\partial \mathbf{x}^{\nu}} + \frac{1}{4} \psi_n^*(\mathbf{x}) \psi_{n'}(\mathbf{x}) q F_{nn'}^{\psi}(\psi(\mathbf{x}), \psi^*(\mathbf{x}), \psi^{(l)}(\mathbf{x}), \psi^{*(l)}(\mathbf{x})) \right]. \quad (3.38)$$

UWAGA: W nawiązaniu do sformułowania Lagrange'a współzrędnymi uogólnionymi są amplitudy a prędkościami uogólnionymi - pochodne amplitud układu. [Dodatek 4.5].

⁴ Za [26] całkowita informacja fizyczna K pełni tu funkcję statystycznego poprzednika (całki) działania, natomiast k , jako funkcja gęstości TFI, jest statystycznym poprzednikiem gęstości lagranżjanu.

W tym kroku należy nałożyć na układ informacyjną obserwowaną zasadę strukturalną (3.16) oraz informacyjną zasadę wariacyjną (3.17). W kolejnych krokach otrzymuje się warunek zerowy (wynikający z zastosowania zmodyfikowanej obserwowanej zasady strukturalnej) oraz układ równań Eulera-Lagrange'a (wynikający z zastosowania zasady wariacyjnej).

Definicja 3.5.2. (Równania różniczkowe - przypadek z amplitudą q) *Warunek zerowy wygląda następująco:*

$$\sum_{\nu=0}^3 \frac{\partial q_n(\mathbf{x})}{\partial \mathbf{x}_\nu} \frac{\partial q_n(\mathbf{x})}{\partial \mathbf{x}^\nu} + \frac{\kappa}{4} q_n^2(\mathbf{x}) qF_n(q_n(\mathbf{x})) = 0, \quad n = 1, \dots, N. \quad (3.39)$$

Układ równań Eulera-Lagrange'a przyjmuje postać:

$$\sum_{\nu=0}^3 \frac{\partial}{\partial \mathbf{x}_\nu} \left(\frac{\partial k}{\partial \left(\frac{\partial q_n(\mathbf{x})}{\partial \mathbf{x}^\nu} \right)} \right) = \frac{\partial k}{\partial q_n(\mathbf{x})}, \quad n = 1, \dots, N. \quad (3.40)$$

Powyższe równania są zapisane w postaci bliskiej ich bezpośredniego użycia i otrzymania jawnej postaci qF_n oraz uzyskania równań ruchu.

Definicja 3.5.3. (Równania różniczkowe - przypadek z amplitudą ψ) *Warunek zerowy:*

$$\sum_{\nu=0}^3 \frac{\partial \psi_n^*(\mathbf{x})}{\partial \mathbf{x}_\nu} \frac{\partial \psi_{n'}(\mathbf{x})}{\partial \mathbf{x}^\nu} + \frac{\kappa}{4} \sum_{n'=1}^{N/2} \psi_n^*(\mathbf{x}) \psi_{n'}(\mathbf{x}) qF_{nn'}^\psi \left(\psi(\mathbf{x}), \psi^*(\mathbf{x}), \frac{\partial \psi(\mathbf{x})}{\partial(\mathbf{x})}, \frac{\partial \psi^*(\mathbf{x})}{\partial(\mathbf{x})} \right) = 0, \quad n = 1, \dots, N/2. \quad (3.41)$$

Układ równań Eulera-Lagrange'a:

$$\sum_{\nu=0}^3 \frac{\partial}{\partial \mathbf{x}_\nu} \left(\frac{\partial k}{\partial \left(\frac{\partial \psi_n^*(\mathbf{x})}{\partial \mathbf{x}^\nu} \right)} \right) = \frac{\partial k}{\partial \psi_n^*(\mathbf{x})}, \quad n = 1, \dots, N/2. \quad (3.42)$$

Aby zobaczyć działanie wynikające z powyższych równań TFI, biorąc pod uwagę dość dużą ogólność postaci funkcji gęstości k całkowitej fizycznej informacji (TFI) (3.38), należy podać konkretną postać k , dla każdego zagadnienia z polem ψ , tak jak to podano dla teorii pola (i równania Diraca).

W dalszym kroku pokazano, że wszystkie modelowe różnice po stronie qF_n zależą od formy konkretnego fizycznego zagadnienia.

3.6. Postać formy Q w teorii pola

Za [26] zostaną tu przedstawione wyniki metody EFI otrzymane w [28]. Zostanie pokazane jak możliwe rozwiązanie zasad informacyjnych, strukturalnej i wariacyjnej, przewiduje pojawienie się pola fermionowego.

Przy próbie uzyskania równań teorii pola i równania Diraca procedura EFI musi od początku wybrać typ amplitudy (q_n lub ψ_n), określić zasady informacyjne i warunki

brzegowe oraz musi być wykonywana we właściwym układzie współrzędnym, który uwzględnia istnienie strukturalnej grupy symetrii oraz związanych z nią pól cechowania.

Transformacja Fouriera, która pojawi się poniżej, tworzy rodzaj samosplątania pomiędzy reprezentacją położeniową a pędową realizowanych wartości zmiennych układu występujących w pojemności informacyjnej I . Wyprowadzenie oczekiwanej zasady strukturalnej, wyjaśnia je jako splątanie pędowych stopni swobody układu spowodowane jego masą [26].

Przykład 3.6.1. (Informacja Fouriera dla pola ψ) Rozważamy cząstkę opisaną poprzez zbiór zespolonych funkcji falowych $\psi_n(\mathbf{x})$

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

w czasoprzestrzeni χ . Ich transformaty Fouriera $\phi_n(\mathbf{p})$ w sprzężonej do przestrzeni przesunięć χ energetyczno-pędowej przestrzeni \mathcal{P} czteropędów $\mathbf{p} \equiv (p^\mu)_{\mu=0}^3 = (\frac{E}{c}, p^1, p^2, p^3)$ mają postać:

$$\phi_n(\mathbf{p}) = \frac{1}{(2\pi\hbar)^2} \int_{\chi} d^4\mathbf{x} \psi_n(\mathbf{x}) e^{i(\sum_{\nu=0}^3 \mathbf{x}^\nu p_\nu)/\hbar}, \quad (3.43)$$

gdzie $\sum_{\nu=0}^3 \mathbf{x}^\nu p_\nu = Et - \sum_{l=1}^3 \mathbf{x}^l p^l$. Za [26] można przepisać (3.30) jako:

$$I[\psi(\mathbf{x})] = I[\psi(\mathbf{p})] = \frac{4N}{\hbar^2} \int_{\mathcal{P}} d^4\mathbf{p} \sum_{n=1}^{N/2} |\phi(\mathbf{p})|^2 \left(\frac{E^2}{c^2} - \mathbf{p}^2 \right), \quad (3.44)$$

gdzie $\mathbf{p}^2 = \sum_{k=1}^3 p_k p^k$. Kwadrat masy cząstki jest stałą niezależnie od statystycznych fluktuacji energii E oraz pędu \mathbf{p} i jest zdefiniowano jako:

$$m^2 := \frac{1}{c^2} \int_{\mathcal{P}} d^4\mathbf{p} \sum_{n=1}^{N/2} |\phi_n(\mathbf{p})|^2 \left(\frac{E^2}{c^2} - \mathbf{p}^2 \right). \quad (3.45)$$

Dla cząstki swobodnej można zapisać:

$$I[\psi(\mathbf{x})] = I[\psi(\mathbf{p})] = 4N \left(\frac{mc}{\hbar} \right)^2 = \text{const.} \quad (3.46)$$

Definicja 3.6.1. (Informacja Fouriera z ψ)

$$K_F = \int_{\chi} d^4\mathbf{x} k_F = 4N \int_{\chi} d^4\mathbf{x} \sum_{n=1}^{N/2} \left[\sum_{\nu=0}^3 \left(\frac{\partial \psi_n}{\partial \mathbf{x}_\nu} \right)^* \frac{\partial \psi_n}{\partial \mathbf{x}^\nu} - \left(\frac{mc}{\hbar} \right)^2 \psi_n^* \psi_n \right] = 0. \quad (3.47)$$

Jeśli $4N \left(\frac{mc}{\hbar} \right)^2$ jest zadana jako informacja strukturalna Q układu to wówczas równanie $K_F = 0$ staje się informacyjną zasadą strukturalną dla układu definiującego szczególny typ skalarnego pola Kleina-Gordona. k_F określa gęstość informacji Fouriera (F).

Przykład 3.6.2. (Informacja Fouriera dla amplitudy rzeczywistej q_n) Można rozważyć cząstkę opisaną poprzez zbiór rzeczywistych amplitud $q_n(\mathbf{x})$, $n = 1, 2, \dots, N$ na

czasoprzestrzeni przesunąć χ . Ich zespolone transformaty Fouriera $\tilde{q}_n(\mathbf{p})$ rzeczywistych funkcji $q_n(\mathbf{x})$, gdzie $\mathbf{p} \equiv (p^\mu)_{\mu=0}^3 = (\frac{E}{c}, p^1, p^2, p^3) \in \mathcal{P}$ jest czteropędem mają postać:

$$\tilde{q}_n(\mathbf{p}) = \frac{1}{(2\pi\hbar)^2} \int_{\chi} d^4\mathbf{x} q_n(\mathbf{x}) e^{i(\sum_{\nu=0}^3 \mathbf{x}^\nu p_\nu)/\hbar}, \quad (3.48)$$

gdzie $\sum_{\nu=0}^3 \mathbf{x}^\nu p_\nu = Et - \sum_{l=1}^3 \mathbf{x}^l p^l$. Dalej można to równanie przepisać jako:

$$I[q(\mathbf{x})] = I[\tilde{q}(\mathbf{p})] = \frac{4}{\hbar^2} \int_{\mathcal{P}} d^4\mathbf{p} \sum_{n=1}^N |\tilde{q}_n(\mathbf{p})|^2 \left(\frac{E^2}{c^2} - \mathbf{p}^2 \right), \quad (3.49)$$

gdzie $\mathbf{p}^2 = \sum_{k=1}^3 p_k p^k$. Kwadrat masy cząstki jest stałą niezależnie od statystycznych fluktuacji energii E oraz pędu \mathbf{p} również i dla naszego pola rzeczywistego (tak jak to miało miejsce dla pola zespolonego w poprzednim przypadku) i jest zdefiniowany jako:

$$m^2 := \frac{1}{Nc^2} \int_{\mathcal{P}} d^4\mathbf{p} \sum_{n=1}^N |\tilde{q}_n(\mathbf{p})|^2 \left(\frac{E^2}{c^2} - \mathbf{p}^2 \right). \quad (3.50)$$

Dla cząstki swobodnej opisanej polem rzeczywistym rangi N można zapisać:

$$I[q(\mathbf{x})] = I[\tilde{q}(\mathbf{p})] = 4N \left(\frac{mc}{\hbar} \right)^2 = \text{const.} \quad (3.51)$$

Definicja 3.6.2. (Informacja Fouriera dla pola q)

$$K_F = \int_{\chi} d^4\mathbf{x} k_F = 4 \int_{\chi} d^4\mathbf{x} \sum_{n=1}^N \left[\sum_{\nu=0}^3 \frac{\partial q_n}{\partial \mathbf{x}_\nu} \frac{\partial q_n}{\partial \mathbf{x}^\nu} - \left(\frac{mc}{\hbar} \right)^2 q_n^2 \right] = 0. \quad (3.52)$$

k_F określa gęstość informacji Fouriera (F).

Jako przestrzeń bazową członów kinetycznych, w celu zbudowania modelu teorii pola, wybierano zbiór przesunień χ , którą u nas jest czasoprzestrzeń Minkowskiego. Pojawiła się ona jako konsekwencja transformacji modelu statystycznego do przestrzeni przesunień. W celu uczynienia kinematycznej postaci pojemności informacyjnej I niezmienniczą ze względu na lokalne transformacje cechowania zapisano I w postaci współzmienniczej wykorzystując do tego pochodną kowariantną na przestrzeni bazowej χ . Z pochodną kowariantną związany jest układ współrzędnych na przestrzeni statystycznej co wiąże się z tym, że pola cechowania są amplitudami q_n [29].

Wprowadzam teraz postać pochodnej kowariantnej, która będzie wykorzystana przy konstrukcji równań pola.

Definicja 3.6.3. (Pochodna kowariantna)

$$(\partial_\mu) \equiv \left(\frac{\partial}{\partial \mathbf{x}^\mu} \right) \equiv \left(\frac{\partial}{\partial(ct)}, \frac{\partial}{\partial \mathbf{x}^1}, \frac{\partial}{\partial \mathbf{x}^2}, \frac{\partial}{\partial \mathbf{x}^3} \right) \equiv \left(\frac{\partial}{\partial(ct)}, \vec{\nabla} \right) \quad (3.53)$$

W przypadku pól cechowania A_μ pochodna kowariantna ma postać:

$$D_\mu \equiv (D_0, D_l) = \partial_\mu - i \frac{e}{c\hbar} A_\mu, \quad (3.54)$$

gdzie e jest ładunkiem elektronu.

Wprowadzono więc pochodną kowariantną do zasad informacyjnych i zastosowano metodę EFI w celu otrzymamy równań ruchu równania Diraca. Zgodnie z powyższymi rozważaniami, współmiennicza forma (3.30) kinematycznej postaci pojemności informacyjnej dla ψ ma postać:

Definicja 3.6.4. (I wyrażona w funkcji falowej $\psi_n(\mathbf{x})$ z pochodną kowariantną D_μ)

$$I = 4N \int_{\chi} d^4\mathbf{x} \sum_{n=1}^{N/2} \sum_{\mu=0}^3 (D_\mu \psi_n)^* D^\mu \psi_n. \quad (3.55)$$

Dla potrzeb pola fermionowego, z którym mamy do czynienia w przypadku równania Diraca, zapisujemy (3.55) w postaci niezmienniczej ze względu na transformacje izometrii⁵ działające w $N/2$ -wymiarowej, zespolonej przestrzeni pól ψ rangi N .

Definicja 3.6.5. (I w funkcji ψ_n z pochodną D_μ oraz tensorem metrycznym $\eta^{\mu\nu}$)

$$I = 4N \sum_{n=1}^{N/2} \int_{\chi} d^4\mathbf{x} \sum_{\mu,\nu=0}^3 (D_\mu \psi_n^*(\mathbf{x})) \eta^{\mu\nu} (D_\nu \psi_n(\mathbf{x})) \quad (3.56)$$

w skrócie tensor metryczny, dla wektorowego indeksu Minkowskiego $\nu = 0, 1, 2, 3$ przyjmuje postać $(\eta_{\nu\mu}) = \text{diag}(1, -1, -1, -1)$ (co oznacza macierz diagonalną z niezerowymi elementami na przekątnej głównej oraz zerami poza nią).

Macierze Diraca γ^μ , $\mu = 0, 1, 2, 3$, które składają się na tensor metryczny, tworzą dla pola fermionowego spinorową reprezentację ortogonalnej bazy (dla spinorów rangi N baza jest 16 wymiarowa) i spełniają tożsamość:

$$\eta^{\mu\nu} = \frac{1}{2} \{\gamma^\mu, \gamma^\nu\}, \quad \mu, \nu = 0, 1, 2, 3. \quad (3.57)$$

Jedyna całkowita fizyczna informacja TFI (3.14) (z $\kappa = 1$) dla równania Kleina-Gordona (KG) i każdego pola typu ψ rangi N , gdzie k_{KG} jest gęstością TFI dla równania KG, ma postać (złożoną z sumy części kinetycznej i części strukturalnej):

$$K = K_{KG} = \int_{\chi} d^4\mathbf{x} k_{KG} \equiv 4N \int_{\chi} d^4\mathbf{x} \sum_{n=1}^{N/2} \sum_{\mu=0}^3 (D_\mu \psi_n)^* D^\mu \psi_n + Q. \quad (3.58)$$

Postać ta jest formą podstawową dla przypadku N -skalarów (szerzej opisanych w [26]). W przypadku pola fermionowego i spinorów rangi N , można dokonać (za Friedenem [28]) rozkładu fizycznej informacji K_{KG} (3.58) na składowe, podstawiając wzór (3.57) w (3.56). Uwzględniając pola cechowania w pochodnej kowariantnej otrzymano następujący rozkład fizycznej informacji K_{KG} :

$$K = K_D \equiv 4N \int_{\chi} d^4\mathbf{x} \sum_{n=1}^{N/2} \sum_{\mu=0}^3 (D_\mu \psi_n)^* D^\mu \psi_n + Q, \quad (3.59)$$

gdzie informacja strukturalna Q ma postać podaną poniżej.

Definicja 3.6.6. (Informacja strukturalna Q dla pola fermionowego)

$$Q = Q_D = \int_{\chi} d^4 \mathbf{x} q_D \equiv -4N \int_{\chi} d^4 \mathbf{x} \sum_{n=1}^{N/2} \left[v_{1n} v_{2n} + \left(\frac{mc}{\hbar} \right)^2 \psi_n^* \psi_n \right] + (\text{pozostałe człony}), \quad (3.60)$$

gdzie q_D jest gęstością informacji strukturalnej, a $\frac{N}{2} = 4$ wymiarowe wektory kolumnowe $v_i = (v_{i1}, v_{i2}, v_{i3}, v_{i4})^T$ dla $i = 1, 2$ mają składowe:

$$v_{1n} = \sum_{n'=1}^4 \left(i\mathbb{I}D_0 - \beta \frac{mc}{\hbar} + \sum_{l=1}^3 i\alpha^l D_l \right)_{nn'} \psi_{n'}, \quad n = 1, 2, 3, 4 \quad (3.61)$$

oraz

$$v_{2n} = \sum_{n'=1}^4 \left(-i\mathbb{I}D_0 + \beta^* \frac{mc}{\hbar} + \sum_{l=1}^3 i\alpha^{l*} D_l \right)_{nn'} \psi_{n'}^*, \quad n = 1, 2, 3, 4, \quad (3.62)$$

gdzie macierze $\alpha^l, l = 1, 2, 3$ oraz β są macierzami Diraca, a \mathbb{I} jest 4×4 - wymiarową macierzą jednostkową.

To Frieden [28] przeprowadził dekompozycję K_{KG} do postaci K_D . Pokazał on również, że wyrażenie oznaczone jako *pozostałe człony* zeruje się przy założeniu, że macierze β i α_l spełniają relację algebry Clifforda (czyli, że iloczyn nie jest przemienny, patrz Dodatek 4.2.1 oraz Dodatek 4.2.2).

Zasady informacyjne, zmodyfikowana informacyjna zasada strukturalna (3.16) dla $\kappa = 1$ i gęstości informacji strukturalnej (3.60) oraz zasada wariacyjna (3.17) (czyli $\delta(I + Q) = 0$) dla równania Diraca w postaci:

Definicja 3.6.7. (Zasady informacyjne dla równania Diraca)

$$\tilde{i}' + \tilde{\mathbf{C}} + q = 0 \text{ oraz } \delta_{(\psi^*)}(I + Q_D) = 0 \quad (3.63)$$

Pole fermionowe jest samospójnym rozwiązaniem zasad informacyjnych, zarówno wariacyjnej jak i strukturalnej. Powyższe zasady dają układ dwóch równań różniczkowych, w wyniku których otrzymano równanie Diraca.

Definicja 3.6.8. (Równanie Diraca)

$$v_1 = \left(iD_0 - \beta \frac{mc}{\hbar} + \sum_{l=1}^3 i\alpha^l D_l \right) \psi = 0, \quad \text{gdzie } \psi = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix}. \quad (3.64)$$

3.7. Omówienie rozwiązania zasad informacyjnych metody EFI dla modelu teorii pola - równanie Diraca

W trakcie rozkładu K_{KG} z członu kinetycznego I (3.58), generowane są wszystkie składniki informacji strukturalnej Q (3.60). Człon masowy generowany jest poprzez sprzężenie Fourierowskie (3.43) pomiędzy reprezentacją położeniową i pędową wynikiem czego masa pola Diracowskiego jest przejawem istnienia samosplątania Fourierowskiego pomiędzy reprezentacją położeniową a pędową.

Definicja 3.7.1. (Całka działania kwadratury równania Diraca) Równanie (3.60) możemy, wprowadzając S_D jako charakterystyczną część Dirakowskiego działania dla kwadratury Diraca, wyznaczoną dla pola Diraca o randze $N = 8$, przepisać w następującej postaci:

$$Q = Q_D \equiv S_q - 4N \int_{\chi} d^4\mathbf{x} \sum_{n=1}^{N/2} \left[\left(\frac{mc}{\hbar} \right)^2 \psi_n^* \psi_n \right]. \quad (3.65)$$

Całka działania (spełniając warunek zerowy) ma tu postać:

$$S_D = -4N \int_{\chi} d^4\mathbf{x} \sum_{n=1}^{N/2} (v_{1n} v_{2n}) + (\text{pozostałe człony}), \quad (3.66)$$

a $K \equiv K_D$ redukuje się i w kwadraturze określa postać K_{KG} dla (3.58) jako:

$$K = K_{KG} \equiv 4N \int_{\chi} d^4\mathbf{x} \sum_{n=1}^{N/2} \left[\sum_{\mu=0}^3 (D_{\mu} \psi_n)^* D^{\mu} \psi_n + \left(\frac{mc}{\hbar} \right)^2 \psi_n^* \psi_n \right]. \quad (3.67)$$

Przedstawiono tu metodę EFI jako typ estymacji procedury statystycznej w ramach teorii pomiaru. Metoda EFI, poprzez wyjście od funkcji wiarygodności oraz pojemności informacyjnej i uzyskanie więzów strukturalnych, dała równania różniczkowe równania Diraca. Wszystkie warunki nałożone na układ traktowano jako związki na odchylenie (fluktuacje) wartości pomiarowych od wartości oczekiwanych. Pomiaru dokonywał sam układ.

Rozdział 4

Podsumowanie

W pracy dokonano przedstawienia konstrukcji relatywistycznego równania Diraca zarówno jego historycznej konstrukcji, jak również bazującej na statystycznym podejściu zaproponowanym przez B. Roya Friedena. Wykorzystano funkcję wiarygodności $P(\Theta)$ do zbudowania metody największej wiarygodności (MNW), oraz informacji Fishera (IF) - obserwowanej i oczekiwanej oraz wprowadzono nierówność Rao-Fishera. W kolejnym kroku MNW i IF zawarte w układzie, poprzez skorzystanie z zasad informacyjnych metody EFI, posłużyły do wyprowadzenia równania Diraca dla cząstki swobodnej.

Dodatek A – Z historii fizyki

4.1. Sylwetki postaci

Ronald A. Fisher (1890 - 1964)

— Narodowość: Anglik

Genetyk i statystyk brytyjski. Zajmował się metodami weryfikacji hipotez za pomocą metod statystycznych, w takich dziedzinach nauki jak antropologia, genetyka, ekologia. Anders Hald¹ określił go mianem *geniusza, który niemalże sam stworzył podstawy współczesnej statystyki*, zaś Richard Dawkins² jako *największego ze spadkobierców Darwina*. Stworzył on między innymi statystyczną metodę największej wiarygodności (MNV), analizę wariancji (ANOVA) oraz liniową analizę dyskryminacyjną. Jego praca naukowa *Statistical Methods for Research Workers* (1925) wznawiana była przez ponad 50 lat. Jest twórcą tzw. nowoczesnej statystyki matematycznej. Metody wprowadzone przez Fishera stosowane są obecnie w każdej dziedzinie nauki.

Paul Dirac (1902 - 1984)

— Narodowość: Anglik

— Nagroda Nobla (1933) przyznana mu (wraz z Erwinem Schrödingerem) w dziedzinie fizyki za *odkrycie nowych, płodnych aspektów teorii atomów i ich zastosowanie*.

Brytyjski fizyk teoretyk. Profesor Uniwersytetu w Cambridge oraz Uniwersytetu Oksfordzkiego. Autor wydanej w latach 30. XX w. książki 'Podstawy mechaniki kwantowej' zawierającej pierwszy systematyczny wykład teorii operatorów liniowych. Sformułował równanie falowe (1928) zwane współcześnie równaniem Diraca, które opisuje elektron w sposób relatywistycznie niezmienniczy. Równanie to pozwoliło na przewidzenie istnienia antycząstki elektronu - pozytronu oraz antymaterii oraz wyjaśnienie pochodzenia spinu elektronu. Wprowadził pojęcie morza Diraca na wyjaśnienie istnienia stanów o ujemnych energiach. Opracował niezależnie statystykę Fermiego-Diraca dla fermionów (o spinie 1/2).

¹ Anders Hald (1913-2007) - duński statystyk, naukowiec oraz znawca historii rachunku prawdopodobieństwa i statystyki.

² Richard Dawkins (1941-) - brytyjski zoolog, etolog, ewolucjonista i publicysta.

B. Roy Frieden (1936 -)

— Narodowość: Amerykanin

Fizyk matematyk, który najbardziej znany jest z prac na temat informacji Fishera i jej wykorzystania, jako podstawowe prawo w fizyce teoretycznej, do wyznaczenia modeli fizycznych takich jak równanie Schrödingera, równanie Kleina-Gordona, rozkład Maxwella-Boltzmana, zasada nieoznaczoności Heisenberga i inne.

Profesor Frieden używa informacji Fishera oraz metody EFI (ekstremalna fizyczna informacja) do wyprowadzenia fundamentalnych praw fizyki, biologii, chemii i ekonomii. Argumentuje, że informacja Fishera I , a w szczególności ubytek informacji oraz metoda EFI, stanowią ogólną metodę do wyprowadzania ścisłych równań i modeli (w tym modeli fizycznych). Istnieje wiele krytyki co do poprawności idei Friedena. Prace jego na temat metody EFI są wciąż analizowane i rozwijane przez niezależne zespoły badawcze.

Dodatek B – Aparat matematyczny

4.2. Algebry

4.2.1. Algebry Grassmanna

Liczby grassmanowskie mają zastosowanie w fizyce, między innymi są wartościami własnymi operatorów fermionowych.

N-wymiarowa liniowa algebra geometryczna określana jako algebra Grassmanna leży u podstaw mechaniki kwantowej (u podstaw idei supersymetrii i supergrawitacji). Liczby Grassmanna jako obiekty należące do algebry Grassmanna mają dobrze zdefiniowane operacje dodawania (łączne i przemienne, element neutralny dodawania - grassmannowskie zero, które utożsamia się z zerem rzeczywistym), odejmowania (jako działanie odwrotne do dodawania) i mnożenia (łączne i antyprzemienne, element neutralny mnożenia nie istnieje). Kwadrat każdej liczby wynosi zero. Nie obowiązuje prawo rozdzielności dodawania względem mnożenia.

Można połączyć algebrę liczb grassmanowskich z algebrą zwykłych liczb rzeczywistych. Iloczyn liczby rzeczywistej i grassmanowskiej definiuje się podobnie, jak mnożenie wektorów przez skalar. Mnożenie przez skalar jest rozdzielne względem dodawania liczb grassmanowskich. Sumę liczby grassmanowskiej i rzeczywistej traktuje się jak wielomian.

4.2.2. Algebry Clifforda

Do opisu algebry Clifforda $\mathfrak{Cl}(n, n-1)$ wystarczą macierze (Pauliego) $\sigma_\mu^{(n)} (\mu = 1, \dots, 2n-1)$ odpowiadające elementom $e_{\mu\eta} \in \mathfrak{Cl}_0(n, n-1)$. Algebra $\mathfrak{Cl}(n, n)$ jest generowana przez macierze (Diraca) $\gamma_\mu^{(n)} (\mu = 1, \dots, 2n)$ [9].

4.3. Komutator i antykomutator

- **Komutator** dwóch elementów a i b $[a, b]$ lub $[a, b]_-$ definiowany jest jako $[a, b] = ab - ba$
- **Antykomutator** dwóch elementów a i b $\{a, b\}$ (z nawiasami Poissona) lub $[a, b]_+$ definiowany jest jako $\{a, b\} = ab + ba$.

Twierdzenie 4.3.1. *Nawias Poissona definiuje mnożenie w przestrzeni funkcji, które nie jest przemienne ani łączne. Nawias Poissona $\{f, g\}$ posiada następujące własności:*

1. *jest biliniowy i antysymetryczny, tzn. $\{f, g\} = -\{g, f\}$,*
2. *spełnia tożsamość Jacobiego, tzn. $\{f, \{g, h\}\} + \{g, \{h, f\}\} + \{h, \{f, g\}\} = 0$,*

3. odwzorowanie $f \rightarrow \{g, h\}$ jest dla każdego g różniczkowaniem $\{g, fh\} = f\{f, g\} + \{f, g\}h$

Dzięki tym definicjom łatwiej jest zapisać reguły komutacyjne i antykomutacyjne stosowane w mechanice kwantowej.

W kwantowej teorii pola dla pól fermionowych stosuje się reguły antykomutacyjne oraz liczby Grassmana, czyli liczby rozpinające algebrę, w której generatory antykomutują (są antyprzemienne) między sobą oraz komutują (są przemienne) ze zwykłymi liczbami [19, 20].

4.4. Czasoprzestrzeń Minkowskiego

Hermann Minkowski wprowadził w 1908 roku pojęcie czasoprzestrzeni poprzez zdefiniowanie przestrzeni liniowej łączącej czas t z przestrzenią \mathbf{x} , mającą trzy współrzędne fizyczne tzw. czasoprzestrzeń M^4 . Miało to miejsce już po ogłoszeniu postulatów szczególnej teorii Einsteina. To nowe pojęcie nadało teorii postać matematyczną, którą można było bezpośrednio uogólnić przechodząc do ogólnej teorii względności. Oś układu współrzędnych oznaczamy przez $x^i, i = \{0, 1, 2, 3\}$. Są to trzy osie współrzędnych przestrzennych x, y, z i czwartej osi czasu t prostopadłych do siebie określając tym samym czterowymiarową przestrzeń - continuum zwane czasoprzestrzenią. Współrzędne przestrzenne i czas przekształcane są w przestrzeni Minkowskiego w następujący sposób: $x^0 = ct, x^1 = x, x^2 = y, x^3 = z$. Interwał (niezmiennik) czasoprzestrzenny definiuje się jako $ds^2 = (dx^0)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2$. Interwał ds jest niezmienniczy względem transformacji Lorentza - ma jednakową wartość liczbową we wszystkich układach odniesienia, poruszających się względem siebie ruchem jednostajnie prostoliniowym.

Wartość interwału czasoprzestrzennego między dwoma zdarzeniami $(\Delta s)^2$ ma taką samą postać jak $(\text{odległość})^2$ pomiędzy dwoma punktami na płaszczyźnie w układzie kartezjańskim, z tą tylko różnicą, że może on przyjąć wartość ujemną. Z postaci niezmiennika s^2 wynika postać tensora metrycznego:

$$\eta_{\mu\nu} = \eta^{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

Nie jest więc to przestrzeń euklidesowa, ze względu na przeciwne znaki przy przy trzech współrzędnych przestrzennych w stosunku do współrzędnej czasowej. Jest to **przestrzeń pseudoeuklidesowa (rozmaitość pseudoriemannowska)**.

Tensor metryczny definiuje się poprzez iloczyn skalarny wektorów bazy:

$$\eta_{\mu\nu} = e_\mu e_\nu \quad \mu, \nu = 0, 1, 2, 3.$$

Jest to postać **kowariantna** (o dolnych indeksach) tensora. Postać **kontrawariantną** (o górnych indeksach) otrzymuje się jako macierz odwrotną z macierzy $(\eta_{\mu\nu})$ czyli: $\eta^{\mu\nu} = \eta_{\mu\nu}^{-1}$.

Współrzędne $\eta_{\mu\nu}$ tensora metrycznego są równe iloczynom skalarnym wektorów bazowych e_i, e_j lokalnego układu współrzędnych. Wektory bazowe oblicza się ze wzoru: $e_\mu = \frac{\partial x}{\partial x^\mu}$, $\mu = 0, 1, 2, 3$, gdzie $x = (x^0, x^1, x^2, x^3)$.

Aby obniżyć (podnieść) wskaźnik wektora X mnożymy go przez tensor metryczny $\eta_{\mu\nu}$ (lub $\eta^{\mu\nu}$):

$$X_\mu = \eta_{\mu\nu} X^\nu,$$

$$X^\mu = \eta^{\mu\nu} X_\nu,$$

Zastosowano tu konwencję sumacyjną Einsteina - sumowanie po powtarzającym się indeksie. Składowe X_μ są tu współrzędnymi kowariantnymi (dolnymi) a składowe X^ν są współrzędnymi kontrawariantnymi (górnymi) wektora X .

4.4.1. Iloczyn skalarny

Stosując konwencję sumacyjną Einsteina iloczyn skalarny dwóch wektorów X i Y wyraża się następująco:

$$\mathbf{X} \cdot \mathbf{Y} = \eta_{\mu\nu} X^\mu Y^\nu = X^\mu Y_\mu = X_\mu Y^\mu.$$

4.4.2. Czterowektor położenia

Czterowektor położenia:

- postać kontrawariantna: $x^\mu = (x^0, x^1, x^2, x^3) = (ct, x, y, z) = (ct, \vec{r})$,
- postać kowariantna: $x_\mu = (x_0, x_1, x_2, x_3) = (ct, -x, -y, -z) = (ct, -\vec{r})$,
- $ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu$.

4.4.3. Czterowektor prędkości

Czterowektor prędkości postać kontrawariantna: $u^\mu = (c\gamma, \vec{v}\gamma)$, gdzie:

- $u^\mu = \frac{dx^\mu}{d\tau}$, gdzie τ jest czasem własnym, który można zapisać jako: $cd\tau = ds$, $d\tau = \frac{dt}{\gamma}$,
- $\|\mathbf{u}\|^2 = \mathbf{u} \cdot \mathbf{u} = \eta_{\mu\nu} u^\mu u^\nu = \pm c^2$ - norma czteroprędkości, jest wielkością uzyskaną przy wykorzystaniu tensora metrycznego $\eta_{\mu\nu}$ czasoprzestrzeni Minkowskiego i wynosi zawsze $\pm c^2$, przy czym znak zależy od przyjętej sygnatury tensora metrycznego.

4.4.4. Czterowektor energii i pędu

Postać kontrawariantna czterowektora energii i pędu ciała o masie spoczynkowej m_0 (za [16, 17]) to $p^\mu = (\frac{E}{c}, \vec{p})$, gdzie:

- $p^\mu = m_0 u^\mu$,
- Składowe czterowektora energii i pędu są to wielkości charakteryzujące cząstkę, które są zachowane (zasada zachowania energii i pędu),
- $\|\mathbf{p}\|^2 = \mathbf{p} \cdot \mathbf{p} = p_\nu p^\nu = \eta_{\mu\nu} p^\mu p^\nu = m_0^2 c^2$ - jest to wzór wiążący współrzędne czterowektora energii i pędu cząstki z jej masą spoczynkową. Wzór ten można przepisać w znanej postaci:

$$\left(\frac{E}{c}\right)^2 - p^2 = m_0^2 c^2 \Rightarrow E^2 = p^2 c^2 + m_0^2 c^4.$$

4.5. Mechanika w ujęciu Lagrange'a

4.5.1. Rachunek wariacyjny

Zasady wariacyjne (za [18]) odgrywają dużą rolę zarówno w mechanice klasycznej, kwantowej, optyce, elektrodynamice i wielu innych. Przykładem zasady wariacyjnej jest zasada Fermata, która wyznacza bieg promienia świetlnego minimalizując jego drogę optyczną (promień świetlny biegnie po drodze, dla której czas przebiegu promienia świetlnego osiąga minimum). Rachunek wariacyjny poszukuje minimum lub maksimum wielkości danej w formie całki.

Mechanika w ujęciu Lagrange'a definiuje nam funkcjonal (zwany działaniem):

$$S = \int_{x_1}^{x_2} f(x, y(x), y'(x)) dx,$$

który liczony wzdłuż krzywej $y = y(x)$ osiąga swój punkt stacjonarny względem wariacji tej krzywej wtedy i tylko wtedy, gdy $y(x)$ spełnia równanie Eulera-Lagrange'a:

$$\frac{\partial f}{\partial y} - \frac{d}{dx} \frac{\partial f}{\partial y'} = 0. \quad (4.1)$$

Zasada mówiąca, że działanie S osiąga swój punkt stacjonarny na prawdziwych trajektoriach nazywa się zasadą Hamiltona lub zasadą najmniejszego działania.

Przypadek wielu zmiennych

Jeśli wyjściowa całka zawiera n zmiennych zależnych, to mamy n równań Eulera-Lagrange'a. Na przykład całka postaci:

$$S = \int_{u_1}^{u_2} L(u, x(u), y(u), x'(u), y'(u)) du$$

z dwiema zmiennymi $x(u)$ i $y(u)$ jest stacjonarna względem wariacji $x(u)$ i $y(u)$ wtedy i tylko wtedy, gdy te dwie funkcje spełniają równania:

$$\begin{aligned} \frac{d}{du} \frac{\partial f}{\partial x'} &= \frac{\partial f}{\partial x} \\ \frac{d}{du} \frac{\partial f}{\partial y'} &= \frac{\partial f}{\partial y}. \end{aligned}$$

4.5.2. Równania Lagrange'a

Lagranżjan

Lagranżjan \mathcal{L} układu, w którym działają jedynie siły zachowawcze, jest zdefiniowany funkcją:

$$\mathcal{L} = T - V, \quad (4.2)$$

gdzie T i V oznaczają odpowiednio energię kinetyczną i potencjalną.

Współrzędne uogólnione

Zbiór n parametrów q_1, \dots, q_n nazywamy współrzędnymi uogólnionymi układu N cząstek, jeśli położenie każdej z cząstek można wyrazić jako funkcje q_1, \dots, q_n (a niekiedy także czasu t) i vice versa, przy czym n jest najmniejszą liczbą parametrów, która pozwala opisać układ w ten sposób. Liczbą stopni swobody układu jest liczba współrzędnych opisujących układ, które mogą się zmieniać niezależnie od siebie. Jeśli liczba stopni swobody jest równa liczbie współrzędnych uogólnionych, to układ nazywamy układem holonomicznym. współrzędne q_1, \dots, q_n nazywamy współrzędnymi naturalnymi, jeśli zależność funkcyjna między \mathbf{r} a q_1, \dots, q_n nie zależy od czasu.

Równanie Lagrange’a

Dla dowolnego układu holonomicznego druga zasada dynamiki Newtona jest równoważna n równaniom Lagrange’a

$$\frac{\partial \mathcal{L}}{\partial q_i} = \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{q}_i}, \quad i = 1, 2, \dots, n \quad (4.3)$$

a równania Lagrange’a są z kolei równoważne zasadzie Hamiltona.

Pęd uogólniony i współrzędne cykliczne

Pęd uogólniony p_i definiuje się jako pochodną

$$p_i = \frac{\partial \mathcal{L}}{\partial \dot{q}_i}. \quad (4.4)$$

Jeśli $\frac{\partial \mathcal{L}}{\partial \dot{q}_i} = 0$ to współrzędna q_i jest współrzędną cykliczną, a odpowiadający jej pęd uogólniony jest zachowany.

Hamiltonian

Hamiltonian jest zdefiniowany jako

$$\mathcal{H} = \sum_{i=1}^n p_i \dot{q}_i - \mathcal{L}. \quad (4.5)$$

Jeśli $\frac{\partial \mathcal{L}}{\partial t} = 0$, to hamiltonian jest zachowany, jeśli współrzędne q_1, \dots, q_n są naturalne, to \mathcal{H} jest po prostu energią układu.

Lagranżjan dla cząstki naładowanej w polu elektromagnetycznym

Lagranżjan dla cząstki o ładunku q , poruszającej się w zewnętrznym polu elektromagnetycznym, ma postać

$$\mathcal{L}(\mathbf{r}, \dot{\mathbf{r}}, t) = \frac{1}{2} m \dot{\mathbf{r}}^2 - q(V - \dot{\mathbf{r}} \cdot \mathbf{A}). \quad (4.6)$$

Bibliografia

- [1] Paul Dirac (1902-1984) - orginał i samotnik, Zofia Gołąb-Meyer, Foton 109, Lato 2010
- [2] Ekonometria, Jerzy Mycielski, Wydział Nauk Ekonomicznych Uniwersytetu Warszawskiego 2009
- [3] Wykłady ze statystyki matematycznej, Agata Boratyńska, 2009 [<http://web.sgh.waw.pl/~aborata/ekonomia/wykladSM.pdf>]
- [4] Statystyka matematyczna. Modele i zadania. Jerzy Greń, PWN Warszawa 1974
- [5] Rachunek prawdopodobieństwa i statystyka, Wojciech Niemiro, Szkoła nauk Ścisłych, Warszawa 1999
- [6] Statystyka dla fizyków, Roman Nowak, PWN, Warszawa 2002
- [7] Wstęp do teorii prawdopodobieństwa, Jacek Jakubowski, Rafał Sztencel, SCRIPT, Warszawa 2000
- [8] Rachunek prawdopodobieństwa dla (prawie) każdego, Jacek Jakubowski, Rafał Sztencel, SCRIPT, Warszawa 2006
- [9] Grupy oraz ich reprezentacje z zastosowaniami w fizyce. Wydanie czwarte rozszerzone, Andrzej Trautman, Instytut Fizyki Teoretycznej Uniwersytet Warszawski, 2011
- [10] Atomy i kwanty. Wprowadzenie do współczesnej spektroskopii atomowej, H. Haken, H.C. Wolf, PWN, Warszawa 2002
- [11] Ramamurti Shankar, Mechanika kwantowa, PWN, Warszawa 2006
- [12] Historia fizyki, Andrzej Kajetan Wróblewski, Wydawnictwo Naukowe PWN, Warszawa 2006
- [13] Introduction to Elementary Particles, D.J. Griffiths, Wiley, New York 1987, s. 221
- [14] Teoria kwantów. Mechanika falowa, Iwo Białynicki-Birula, Marek Cieplak, Jerzy Kamiński, PWN, Warszawa 1991
- [15] Fizyka cząsteczek. Energie i widma., Paweł Kowalczyk, PWN, Warszawa 1999
- [16] Mechanika teoretyczna, Krzysztof Pomorski, Wydawnictwo UMCS, Lublin 2000
- [17] Mechanika, C. Kittel, W.D. Knight, M.A. Ruderman, PWN, Warszawa 1975
- [18] Mechanika klasyczna 1, John R. Taylor, PWN, Warszawa 2012
- [19] Introductory Quantum Mechanics, Richard L. Liboff, Addison-Wesley, 2002
- [20] Introduction to Quantum Mechanics, David J. Griffiths, Prentice Hall, 2004
- [21] Teoria pól kwantowych. Podstawy., Steven Weinberg, PWN, Warszawa 1999
- [22] Teoria pól kwantowych. Supersymetria., Steven Weinberg, PWN, Warszawa 2001
- [23] Klasyczna teoria pola, Krzysztof A. Meissner, PWN, Warszawa 2002
- [24] Teoria Pola, L.D. Landau, E.M. Lifszyc, PWN, Warszawa 1980
- [25] Metoda największej wiarygodności i informacja Fisher'a w fizyce i ekonofizyce (Maximum likelihood method and Fisher's information in physics and econophysics) - skrypt dla studentów ekonofizyki, Jacek Syska, Instytut Fizyki, Uniwersytet Śląski, Katowice

- [26] Likelihood method and Fisher information in construction of physical models, E.W. Piotrowski, J. Sladkowski, J. Syska, S. Zajac, *Phys. Status Solidi B* 246, No. 5, 1033 (2009)
- [27] *Physics from Fisher information. A unification*, B. Roy Frieden, Cambridge University Press 1998
- [28] B.R. Frieden, *Found. Phys.* 16, No.9, A probability law for the fundamental constants, 883-903 (1986). B.R. Frieden, *Phys. Rev. A* 41, Fisher information, disorder, and the equilibrium distributions of physics, 4265-4276 (1990). B.R. Frieden, B.H. Soffer, *Phys. Rev. E* 52, Lagrangians of physics and the game of Fisher-information transfer, 2274-2286 (1995). B.R. Frieden, *Phys. Rev. A* 66, Relations between parameters of adecoherent system and Fisher information, 022107, (2002). B.R. Frieden, A. Plastino, A.R. Plastino and B.H. Soffer, Schrödinger link between non equilibrium thermodynamics and Fisher information, *Phys. Rev. E* 66, 046128 (2002). B.R. Frieden, A. Plastino, A.R. Plastino and B.H. Sffer, Non-equilibrium thermodynamics and Fisher information: An illustrative example, *Phys. Lett. A* 304, (2002), pp.73-78. B.R. Frieden, *Science from Fisher information: A unification*, Cambridge University Press, (2004).
- [29] *Methods of information geometry*, translations of Mathematical monographs, S. Amari, H. Nagaoka, Vol.191, Oxford Univ. Press, (2000). Quantum Fisher metric and estimation for pure state models, A. Fujiwara and H. Nagaoka, *Phys. Lett. A* 201, 119-124, (1995). An estimation theoretical characterization of coherent states, A. Fujiwara and H. Nagaoka, *J. Math. Phys.* 40, 4227-4239, (1999). A geometrical approach to quantum estimation theory, K. Matsumoto, PhD thesis, Univ. of Tokyo, (1998).