UNIWERSYTET IM. ADAMA MICKIEWICZA W POZNANIU
WYDZIAŁ MATEMATYKI I INFORMATYKI

Imię Nazwisko

Kierunek: ???
Specjalność: ???
Numer albumu: 99999

Tytuł po polsku

Tytuł po angielsku

Praca magisterska
napisana pod kierunkiem
prof. UAM dra hab. Tomasza Góreckiego

POZNAŃ 2021

Spis treści

Streszczenie

Abstract

Rozdział 1

Wprowadzenie

Na podstawie książki prof. Bobrowskiego. Teoria prawdopodobieństwa tradycyjnie nazwana rachunkiem prawdopodobieństwa zajmuje się modelami zjawisk losowych (przypadkowych). Opiera się ona na sformalizowaniu pojęcia częstości, gdy dąży ona do stałej wartości, czyli prawdopodobieństwa.

Pojęcie to po raz pierwszy zostało sprecyzowane przez Jakuba Bernoulliego (1654–1705) w dziele „Ars consectandi”, wydanym w 1713 r. Bernoulli rozumie tu prawdopodobieństwo jako stopień pewności.

Powstanie i rozwój rachunku prawdopodobieństwa związany był głównie z grami hazardowymi, mimo że, jak się okazało, ma on bardzo duże możliwości zastosowań. Nadal jednak przy objaśnianiu prawidłowości bardzo przejrzystą interpretację uzyskują takie zjawiska jak: rzut monetą, rzut kostką do gry, losowanie określonych kart z talii itp. Te ilustracje mają na celu eliminacje drugorzędnych aspektów zagadnienia i koncentrację uwagi na czynnikach istotnych dla zjawisk losowych.

Powstanie nowoczesnej teorii prawdopodobieństwa zawdzięczamy przede wszystkim Andrejowi Kołmogorowowi (1901–1982), który w ramach ogólnej teorii miary stworzył tzw. aksjomatyczny model teorii prawdopodobieństwa, pozwalający między innymi uniknąć sprzeczności przy budowie modeli zjawisk losowych.

Aby wyjaśnić otaczającą nas rzeczywistość i występujące w niej zjawiska, nauka oraz praktyka dnia codziennego tworzą odpowiednie modele. Znakomita większość z nich to modele matematyczne. Korzystają one nie tylko z praw odpowiedniej nauki, ale zbudowane są w ramach wybranej dyscypliny matematycznej. Cel budowy takich modeli sprowadza się do tego, aby na podstawie właściwych dla danej dyscypliny matematycznej metod otrzymać wyniki, których interpretacja pozwala lepiej poznać, ocenić, objaśnić, przewidzieć ewolucję otaczającej nas rzeczywistości.

Modele matematyczne można podzielić na dwie zasadnicze grupy: modele deterministyczne formułowane głównie w języku analizy matematycznej i modele losowe (probabilistyczne[1]) korzystające z metod rachunku prawdopodo-

[1] *Probabilis* znaczy w języku łacińskim godny uznania przyjęcia.

bieństwa. Dla tego samego zjawiska mogą istnieć różne modele. Przykładem jest fazowa i korpuskularna teoria propagacji (rozchodzenia się) światła.

Tak więc podstawowym terminem współczesnego rachunku prawdopodobieństwa jest pojęcie zdarzenia losowego lub przypadkowego. Zrozumienie tego, co znaczy „przypadkowe", konieczne jest dla dobrego poznania koncepcji rachunku prawdopodobieństwa.

Rozpatrzmy najpierw następujący przykład. Codziennie odbywam spacer po różnych ścieżkach miejskiego parku. Zdarza się, od czasu do czasu, że spotykam koleżankę, mającą taki sam zwyczaj. Są to spotkania niezamierzone, których nie można z góry przewidzieć, a do których dochodzi w wyniku wielu, na ogół nieuchwytnych, a nawet nieuświadamianych sobie powodów, zarówno ze strony koleżanki, jak i mojej. Mówimy, że są to spotkania przypadkowe.

Już w starożytności bardzo wnikliwie analizowano, co to znaczy, że zdarzenie jest przypadkowe w odróżnieniu od zdarzenia koniecznego (deterministycznego), tzn. takiego, które jest następstwem występowania ustalonej przyczyny. Natomiast o zdarzeniach przypadkowych mówiono, że ich przyczyną jest traf, fortuna, fatum (po grecku *tyche* po polsku *los*). Arystoteles (388–332) tak pisał: „... Wiele rzeczy powstaje i istnieje dzięki przypadkowi i trafowi... Wszystko jednak co nazywamy przypadkiem lub trafem ma swoją określoną przyczynę". Isaac sir Newton (1643–1727) utrzymywał, że zjawiska przypadkowe to wynik naszej ignorancji. Wobec tego, aby wyjaśnić jakiś fragment rzeczywistości, należy stworzyć jego mechaniczną teorię. Pogląd ten przetrwał wiek XVIII i XIX i został zachwiany dopiero w wieku XX przez mechanikę kwantową. Do tego czasu panowało przekonanie, że w miarę postępu nauki zdarzenia przypadkowe znikną zupełnie.

Reminiscencje, na szczęście coraz rzadsze, tego poglądu panują u wielu do dziś. Uważają oni bowiem, że rachunkiem prawdopodobieństwa muszą posługiwać się z tego powodu, że nie potrafią skonstruować modelu deterministycznego.

Zwróćmy jeszcze uwagę na fakt, że ponieważ nie zawsze jesteśmy w stanie wyświetlić przyczynę zdarzenia przypadkowego, zatem może pojawić się mniej lub bardziej nieracjonalny pogląd, iż zdarzenia takie mają zupełnie nieuzasadnione przyczyny, np. konfiguracja planet wpływa na los konkretnego człowieka, szczęśliwe liczby w dacie na dobrą prognozę przedsięwzięcia itp.

W przeszłości dość często mniemano, że zjawiska przypadkowe są wynikiem czarów i wobec tego należy je w sposób magiczny odczarować. Iluż naiwnych ludzi także obecnie ulega takim sugestiom?

Nawet jeśli występowanie wielu zjawisk przypadkowych można uzasadnić w sposób naukowy, to liczba przyczyn może być tak wielka, że ścisłe ich wyjaśnienie staje się niemożliwe. Tak np. mimo że zmiany atmosferyczne podlegają znanym prawom fizyki, to przewidywanie pogody nie może mieć charakteru deterministycznego. W otaczającym nas świecie nie występuje zjawisko, które nie charakteryzowałoby się, w mniejszym lub większym stopniu, przypadkowością. Jeśli jednak z olbrzymiego zbioru czynników można wybrać te, które pozwolą opisać dane zjawisko w uproszczony sposób, to mówimy, że stworzyliśmy jego deterministyczny model. Tak postępuje się w wielu naukach: fizyce, technice, biologii.

Dla wielu problemów naukowych budowa takiego modelu deterministycznego jest niewystarczająca. Wówczas zagadnienie udaje się analizować przy zastosowaniu „modelu probabilistycznego" uwzględniającego także wpływ czynników drugorzędnych, a więc przypadkowych. Udaje się to wówczas, gdy potrafimy ustalić, jak często one występują i jaki jest zakres ich wpływu. Tak np. okazuje się, że w naszych warunkach klimatycznych, mimo że do dziś nie wiadomo, jaka jest tego przyczyna, częstość urodzin chłopca wynosi 0,51 a dziewczynki 0,49.

Podrzucam okrągłą monetę, moneta obraca się w powietrzu i upada na stół, odkrywając jedną stronę – awers („orzeł") lub rewers („reszka"). Czynność tę powtarzam wielokrotnie. Niezależnie od tego, jak bardzo starałbym się powtarzać warunki (wysokość podrzucania, początkowa prędkość i moment obrotu), wynik będzie różny, raz wypadnie "orzeł", a w innym rzucie "reszka". Jest to wynik nieuchwytnych przyczyn np. drobnych nierówności powierzchni stołu itp. Zasadnicze warunki pozostają niezmienne, natomiast drugorzędne czynniki zmieniają się przypadkowo między jednym a drugim rzutem.

Powróćmy do naszego przykładu. Mogę sukcesywnie liczyć, jaką częstość mają nasze spotkania i interesująca będzie niewątpliwie sytuacja, gdy częstość ta będzie się stabilizowała na jakimś poziomie. Świadczyć to będzie o tym, że przyczyny powodujące przypadkowe spotkania cechuje co prawda nie determinizm, ale mimo wszystko jakaś prawidłowość, która różni je od zupełnie chaotycznych spotkań, w których częstość w miarę upływu czasu może zmieniać się w sposób niemożliwy do przewidzenia. Zjawiska charakteryzujące się stabilną częstością nazywamy losowymi (przypadkowymi). Modelami matematycznymi takich zjawisk zajmuje się rachunek prawdopodobieństwa.

Tak życie codzienne, jak i doświadczenia naukowe potwierdzają, że stabilność częstości stanowi własność świata, która powoduje, iż jest on probabilistycznie poznawalny. Posługując się modelami probabilistycznymi, nie możemy określić (obliczyć), jaką wartość przyjmie interesująca nas wielkość (np. temperatura, ciśnienie atmosferyczne), a możemy tylko ocenić parametry charakteryzujące rozkład prawdopodobieństwa przyjmowanych wartości.

Tak więc pomimo skomplikowanych przyczyn występowania pewnych zdarzeń można obserwować prawidłowość częstościową. Dla takich zdarzeń nie możemy nadal przewidzieć z całą pewnością ich wystąpienia, ale możemy przewidywać ich wystąpienie w przyszłości opierając na zaobserwowanej prawidłowości częstościowej. Oczywiście mogą pojawiać się zdarzenia, które w ogóle nie wykazują nawet prawidłowości częstościowej. Zjawiska takie nazywamy chaotycznymi. Zajmuje się nimi od kilkudziesięciu lat teoria tzw. chaosu deterministycznego. Natomiast teoria prawdopodobieństwa dotyczy tylko takich zjawisk o niewiadomej ewolucji, dla których można przyjąć, że charakteryzują się stabilnością częstości. Pojedyncze zdarzenie pozostaje nadal nieprzewidywalnym, natomiast dla wielu analogicznych zdarzeń dają się zaobserwować określone prawidłowości polegające na tym, że wartości będą z określoną częstością występować w ustalonych granicach.

Ogólnie biorąc model probabilistyczny jest na pewno mniej korzystny w zastosowaniach od modelu deterministycznego ze względu na rodzaj i ilość informacji. Z drugiej strony oparcie się na deterministycznej prognozie (np. znajomość dokładnej daty wypadku samochodowego) może mieć negatywne znaczenie psychiczne.

Współczesna teoria prawdopodobieństwa jest jednym z działów matematyki tak samo ścisłym jak inne dyscypliny matematyczne. Z drugiej strony dla użytkownika probabilistyki ważna jest nie tyle precyzja matematyczna ile umiejętność rozpoznania w konkretnym zagadnieniu, jakie są jego probabilistyczne aspekty i rozumienie zastosowanych metod w jego rozwiązywaniu tak, aby otrzymać wynik możliwie najbardziej wiarogodny.

Rozdział 2

Ogólna teoria

2.1. Zdarzenia niezależne

Pojęcie zdarzeń niezależnych i zależnych probabilistycznie lub stochastycznie jest jednym z bardzo ważnych pojęć teorii prawdopodobieństwa. Na samym wstępie trzeba zaznaczyć, że zależność probabilistyczna nie jest równoważna zależności przyczynowo–skutkowej. W praktyce jednak, gdy nie ma możliwości stwierdzenia metodami probabilistycznymi niezależności zdarzeń, a widoczny jest brak zależności przyczynowo–skutkowej, zakładamy również niezależność probabilistyczną tych zdarzeń.

Definicja 1. *Dwa zdarzenia A i B nazywamy* **niezależnymi** *jeżeli spełniają warunek:*

$$\mathrm{P}(A \cap B) = \mathrm{P}(A) \cdot \mathrm{P}(B). \tag{2.1}$$

Można łatwo sprawdzić, że warunek (2.1) jest równoważny, intuicyjnie bardziej zrozumiałemu, lecz mniej wygodnemu przy sprawdzaniu, warunkowi:

$$\mathrm{P}(A|B) = \mathrm{P}(A).$$

Niezależność zdarzeń ściśle uwarunkowana jest przyjętą miarą probabilistyczną i dwa zdarzenia niezależne przy jednej mierze probabilistycznej mogą być zależne przy innej mierze probabilistycznej określonej na tym samym ciele zdarzeń. Nie należy mylić zdarzeń rozłącznych, dla których $A \cap B = \emptyset$, $\mathrm{P}(A \cap B) = 0$ i zdarzeń niezależnych, dla których zachodzi (2.1).

2.2. Prawdopodobieństwo całkowite

Zajmiemy się zagadnieniem obliczania prawdopodobieństwa zdarzenia A, które wystąpić może łącznie z jednym ze zdarzeń H_i $(i = 1, 2, \ldots, N)$ tworzących podział przestrzeni zdarzeń elementarnych.

Definicja 2. *Prawdopodobieństwo dowolnego zdarzenia $A \in \mathcal{F}$, obliczane na podstawie znajomości prawdopodobieństw warunkowych $\mathrm{P}(A|H_i)$, nazywamy* **prawdopodobieństwem całkowitym** *tego zdarzenia.*

Prawdopodobieństwo całkowite nie jest inną miarą probabilistyczną (jaką było np. prawdopodobieństwo warunkowe), lecz wskazuje na sposób jego obliczania.

Twierdzenie 1. *Jeżeli zdarzenia H_i $(i = 1, 2, \ldots, N)$ tworzą podział przestrzeni zdarzeń elementarnych, przy czym N jest dowolną liczbą naturalną lub oznacza symbol ∞ i dla każdego H_i jest $\mathrm{P}(H_i) > 0$, wówczas dla dowolnego zdarzenia A zachodzi równość*

$$\mathrm{P}(A) = \sum_{i=1}^{N} \mathrm{P}(A|H_i)\mathrm{P}(H_i). \tag{2.2}$$

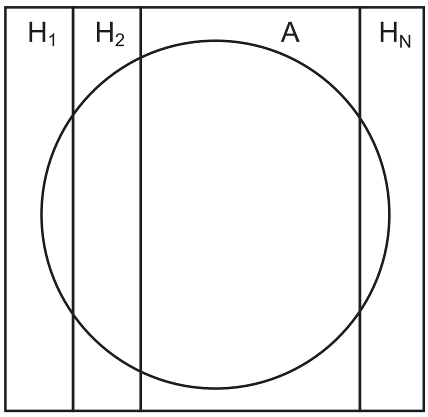
Dowód: Ponieważ z założenia $\mathrm{P}[(\bigcup_{i=1}^{N} H_i)'] = 0$, więc

$$\mathrm{P}(A) = \mathrm{P}(A \cap \bigcup_{i=1}^{N} H_i) + \mathrm{P}[A \cap (\bigcup_{i=1}^{N} H_i)'] =$$

$$\mathrm{P}[\bigcup_{i=1}^{N} (A \cap H_i)] = \sum_{i=1}^{N} \mathrm{P}(A \cap H_i) = \sum_{i=1}^{N} \mathrm{P}(A|H_i)\mathrm{P}(H_i).$$

$\square$

Zależność między zdarzeniami, o których mowa w twierdzeniu (1) ilustruje Rysunek 2.1.



Rysunek 2.1. Ilustracja twierdzenia (1)

Rozdział 3

Metodologia

Rozdział 4

Opis funkcji R (Python)

Rozdział 5

Przykłady

5.1. Przykład 1

5.2. Przykład 2

Porównanie metod znajduje się w Tabeli 5.2.

Tabela 5.1. Wyniki działania algorytmów

Zbiór danych	Algorytm 1	Algorytm 2
A	95%	96%
B	97%	89%

Bibliografia

[1] Górecki, T. (2011): *Podstawy statystyki z przykładami w R.* BTC, Legionowo.

[2] Górecki, T., Łuczak, M. (2013): Using derivatives in time series classification. *Data Mining and Knowledge Discovery* **26(2)***, 310–331.*

[3] Olszewski, R.T. (2001): *Generalized Feature Extraction for Structural Pattern Recognition in Time-Series Data.* Ph.D. Thesis, Carnegie Mellon University, Pittsburgh, PA.

Skorowidz