

Rachunek prawdopodobieństwa / Data Science [KW/2023]

Zapraszam do zapoznania się z rekomendacjami dotyczącymi proponowanych rozwiązań wybranych zadań

(Uwaga: to tylko szkic, rekomendacja, a nie pełny zapis kompletnego rozwiązania przykładowych problemów. Najlepszym sposobem weryfikacji poprawności rozwiązania jest sprawdzenie, czy dowolnie wybrany czytelnik naszej pracy potrafiłby na tej podstawie je powtórzyć - na podobnym przykładzie).

1. Dwaj studenci umówili się razem na konsultacje u profesora, w ramach których chcieliby poprawić krótkie, 15 minutowe kolokwium. Kto przyjdzie pierwszy, poczeka zatem na drugiego wyłącznie w czasie trwania kolokwium, a wówczas będą mogli współpracować. Jaka jest szansa, że studenci się spotkają i podejmą współpracę, jeśli wiadomo, że przybędą w losowym czasie konsultacji profesora, które odbywają się we czwartki, między 11:45 a 12:45?

Rekomendacja (szkic) rozwiązania:

Punktem wyjścia jest określenie danych (warunków brzegowych). Czas konsultacji wykładowcy obejmuje godziny od 11:45 do 12:45, a zatem 60 minut. Jeśli zatem jako x i y oznaczymy czas przybycia na konsultacje pierwszego $[x]$ i drugiego $[y]$ studenta, to oczywiście wartości x, y należą do $\{0, 60\}$. Dodatkowo wiadomo, że studenci będą mogli podjąć współpracę tylko wówczas, jeśli ich czas przybycia na 15-minutowe kolokwium przynajmniej częściowo będzie się pokrywał, czyli tenże "czas wspólny" będzie wynosił $\{0, 15\}$ minut. Ponieważ pytamy o prawdopodobieństwo, warto rozpocząć od definicji prawdopodobieństwa klasycznego $P(A)$, uwzględniając występowanie dwóch zmiennych x i y , co pozwala na zastosowanie perspektywy prawdopodobieństwa geometrycznego. OMEGA może zatem zostać opisana jako powierzchnia kwadratu $[0, 60] \times [0, 60]$. Z kolei zbiór zdarzeń sprzyjających A zostanie wyznaczony przez wszystkie pary x, y (czas przybycia studentów), oczywiście należące do zbioru OMEGA, takie że wartość bezwzględna $|x-y|$ będzie mniejsza lub (w ostateczności) równa od 15 minut.

W rozpatrywanym przypadku znacznie łatwiej odwołać się do zdarzenia przeciwnego A' , wyznaczonego przez zbiór par poszczególnych, wyznaczonych w minutach czasów x, y : $|x'-y'| > 45$. Zatem $A' = 45 \times 45$; oczywiście $\Omega = 60 \times 60$, dlatego $P(A') = A'/\Omega$, czyli około 56%. W konsekwencji, wyznaczając prawdopodobieństwo zdarzenia przeciwnego $P(A) = 1 - P(A')$ wyznaczymy, że szansa na spotkanie się studentów w trakcie krótkiego kolokwium i nawiązanie ewentualnej współpracy wynosi około 44% [obliczając $P(A) = 100 - 56$].

Czy wskazaliście Państwo metodę? Czy wiadomo, z jakich narzędzi trzeba skorzystać, aby przygotować rozwiązanie? Czy określono jedną jednostkę, będącą następnie podstawą wyznaczenia miary prawdopodobieństwa? Jak wyznaczono i czym jest A' ? Co jest cechą charakterystyczną rozpatrywanego tu prawdopodobieństwa geometrycznego (i jego miarą)?

2. Proszę sobie wyobrazić, że kibicujecie Państwo w turnieju gier, w którym uczestniczą cztery drużyny, w tym Wasza ulubiona. Można w nim zagrać w DARTA lub w karty. Dla uatrakcyjnienia gry w DARTA, w okrągłą tarczę do gry w DARTA wpisano kwadrat, który oznaczono kolorem czerwonym. Rozpoczęto grę, polegającą na rzucie lotką w tarczę przez każdego z czterech zawodników. Nie określono czasu trwania gry. Równolegle, obok, toczy się gra w karty. Gracze (jest ich czterech i posługują się klasyczną talią 52 kart) umówili się, że będzie to uproszczona wersja pokera, w której zwycięży ten z graczy, który pierwszy, w pojedynczym, sprawiedliwym rozdaniu uzyska "karetę" z asów (czyli zdobywa punkty, gdy otrzyma w rozdaniu wszystkie cztery asy). Gracze są zdeterminowani i karty rozdają do skutku...
- W obu grach uczestniczą reprezentanci drużyny, której Państwo kibicujecie. Można też na ich sukcesie zarobić niezłą nagrodę pieniężną, pod warunkiem wskazania zwycięzcy. Proszę wskazać, na którego z Waszych ulubieńców postawicie Państwo swoje pieniądze i uzasadnić dlaczego, wiedząc że:
- * "Wasz" gracz w DARTA regularnie trafia w tarczę. Wygrywa (i zdobywa cenne punkty) wtedy, gdy trafią swoją "lotką" w obszar znajdujący się w tarczy, ale poza oznaczonym na czerwono kwadratem.
 - ** "Wasz" gracz w karty nie ma specjalnych przywilejów, ale jest zdeterminowany grać tak długo, jak będzie toczyła się gra w DARTA.
- Oceniając ryzyko proszę wykorzystać znane Państwu narzędzia kombinatoryki i prawdopodobieństwa.

Rekomendacja (szkic) rozwiązania:

1) GRA W DARTA: Zadanie najlepiej rozwiązać metodą klasyczną. Próba dokonania uproszczenia metodą podstawienia nie daje pełnej satysfakcji. Uniemożliwia identyfikację szczególnej właściwości badanego tu prawdopodobieństwa geometrycznego. Najłatwiej będzie opisać zdarzenie przeciwne, oznaczone A' . Zatem miarą wyników sprzyjających dla A' jest powierzchnia kwadratu wpisanego w koło. Warto rozpocząć od $P(A') = A'/\Omega$. Omega, czyli pole koła to: πr^2 . Poszukujemy "a", czyli długości boku kwadratu. Tymczasem pole kwadratu wyznaczyć można z parametru przekątnej kwadratu, wynoszącej $2r$. Pomaga wybór metody na obliczenie „r” ($1/2a \cdot \sqrt{2}$ z dwóch). To pozwala na wyznaczenie formuły na długość boku kwadratu, czyli a. Wartości te trzeba podstawić, poprawnie przeliczyć, co da w efekcie: $P(A') = 2/\pi$. W przybliżeniu wartość prawdopodobieństwa można wyznaczyć, np. w procentach, zatem $P(A) = 1 - P(A') \approx 100\% - 64\% \approx 36\%$. Proszę zauważyć, że do obliczeń nie są potrzebne żadne dodatkowe założenia, w tym wartości liczbowe.

2) GRA W KARTY: W tym przypadku trzeba wyznaczyć prawdopodobieństwo, z jakim gracz w karty w pierwszym (każdym pojedynczym rozdaniu) ma szansę otrzymać 4 asy. Wiadomo, że posługujemy się klasyczną talią kart, liczącą 52 karty, po 13 figur w każdym z 4 kolorów. Kluczem jest dobry opis zdarzenia losowego (B). Proces rozdania kart warto podzielić na podzbiory, ustalić formułę i wówczas z 52 kart każdy gracz otrzymuje 13, a chciałby dostać [opis zdarzenia sprzyjającego B]:

- * z perspektywy kolorów: (13 po 1) * (13 po 1) * (13 po 1) * (13 po 1), albo
- ** z perspektywy figur [asów]: (4 po 4) * (48 po 9).

Analizowane zdarzenie losowe odbywa się w obrębie jednego zbioru (talia), zatem kombinację bez powtórzeń [formułą obliczeniową jest dwumian Newtona (n po k)] uzupełnia reguła mnożenia. Natomiast OMEGA (Ω) powinna zostać wyznaczona analogicznie, opisuje je każde rozdanie kart w ramach którego każdy gracz otrzymuje 13 kart na (52 po 13) sposobów [obliczonych dwumianem Newtona]. W rezultacie wyznaczamy prawdopodobieństwo klasyczne $P(B) = A/\Omega$. Proszę obliczyć! Teraz pozostaje porównać wyznaczone prawdopodobieństwa.

Co zatem jest większe? $P(A)$ czy też $P(B)$? Czy $P(B)$ może przekroczyć 36%? Stawiamy na DARTA?!

3. Proszę wyznaczyć, jakie jest prawdopodobieństwo wygrania "piątki" w loterii LOTTO, w której losuje się 6 liczb (kul) z bębna zawierającego 49 ponumerowanych kolejno kul. Proszę podać również, ile pieniędzy trzeba zainwestować w losowanie, aby wygranie "piątki" było pewne, przyjmując, że za skreślenie kuponu (w tym przypadku za podanie jednej własnej propozycji "szóstki") zapłacimy 1 EUR. Czy warto podjąć taką grę? Czy są szanse na zysk, czy to inwestycja skazana na porażkę?

Rekomendacja (szkic) rozwiązania:

Podstawą rozwiązania jest przyjęcie perspektywy prawdopodobieństwa klasycznego, a następnie wyznaczenie $P(A)$, gdzie A to ilość wszystkich możliwych kombinacji "piątki" w loterii, zaś Ω oznacza ilość wszystkich możliwych wyników w LOTTO. Owszem, trzeba zaznaczyć, jak dokładnie (formuła) wyznaczenie Państwa " A " (258), a jak Ω (prawie 14 mln). Zatem stosowana jest kombinacja bez powtórzeń dla A (6 po 5)*(43 po 1) wraz z regułą mnożenia, a także dla Ω (49 po 6). Zatem $P(A)$ dla "piątki" jest słabe, tylko 0.002%...

Na zakończenie nie chodzi jednak o to, by wyliczyć tylko konieczną kwotę inwestycji w kupony LOTTO, które pozwolą na wygranie piątki z zachowaniem ryzyka wyznaczonego prawdopodobieństwa. Pewność nie jest prostą odwrotnością prawdopodobieństwa... Aby mieć PEWNOŚĆ wygrania piątki (czyli posiadania jednego pewnego kuponu z trafioną piątką), trzeba ich kupić 13983816-257 (jeden mniej, niż wyznaczone A ; ten jeden to właśnie nasz trafiony los), co razy 1 EUR złotych daje niemal 14 mln EURO i prawie się nie różni od inwestycji niezbędnej dla "szóstki".

Zatem rzeczywiście, inwestycja raczej nie opłaca się... Chociaż, przy takiej determinacji, może się okazać, że celując w "piątkę" przy okazji zgarniemy"

- szóstkę
- przynajmniej jedną piątkę
- kilka czwórek
- kilkanaście trójek

gdyż najmniejszą wygraną, jest kwota wypłacana właśnie za trafienie trzech poprawnie skreślonych liczb z 49. Zatem grać, czy nie grać? Proszę policzyć ewentualny bilans..

Łatwo zauważyć, że w przypadku pytania o „czwórkę”, problem można rozwiązać analogicznie, oczywiście stosując właściwe dane.

Wówczas nasze zdarzenie A , czyli trafienie "czwórki", możliwe jest na $(6 \text{ po } 4) * (43 \text{ po } 2)$ sposobów. Zatem nasza kombinacja $C(A) = (6 \text{ po } 4) * (43 \text{ po } 2) = 13545$, więc na tyle sposobów można w losowaniu typować (wylosować) "czwórkę". Teraz można łatwo obliczyć prawdopodobieństwo "trafienia czwórki", gdzie $P(A) = A/\Omega$. Czyli: $P(A) = (6 \text{ po } 4) * (43 \text{ po } 2) / (49 \text{ po } 6) = 13545 / 13983816 \approx 0.00097 \approx 0,097\% \approx 0,1\%$.

Jak jednak trafić "czwórkę" na pewno? Pewność nie jest prostą konsekwencją prawdopodobieństwa. Dlatego, aby być pewnym czwórki, trzeba kupić przynajmniej: $\{13983816-13545+1\}$ kuponów, czyli $\{13983816-13544\}$, a zatem o jeden mniej, niż szansa na "czwórkę" - gdyż ten "jeden" kupon będzie naszym zwycięskim, on daje PEWNOŚĆ. W efekcie, gdy każdy kupon kosztuje 1 EUR, trzeba zainwestować przynajmniej: 13970272 EUR (ok. 13,97 mln EUR), czyli niewiele mniej, by być pewnym nawet "szóstki". Prawda?

Zatem:

prawdopodobieństwo trafienia "czwórki" w LOTTO oszacowano na poziomie 0,1%,

aby jednak mieć pewność wygrania "czwórki" trzeba zainwestować niemal 14 mln EUR, czyli prawie tyle samo, ile trzeba wydać na zakup pełnej puli kuponów w losowaniu.

Co Państwo na to? Warto zagrać?

4. Zarządca Państwa firmą IT_Vinci, która przejęła nowy biurowiec i została podłączona do nowo uruchamianej wewnętrznej centrali telefonicznej, obsługującej połączenia 6-cyfrowe. Poszczególne numery telefoniczne nie mogą zaczynać się od zera. Proszę ustalić:

Jakie są szanse, że zostanie Państwu przydzielony numer, którego trzy ostatnie cyfry będą oznaczone za pomocą dostępnych oznaczeń liczbowych, jednak innych niż liczby nieparzyste*???

Czy szanse te są większe, niż otwarcie za pierwszym razem drzwi do Waszego nowego biura, które zabezpieczono trzycyfrowym kodem PIN?

Uwaga, niestety nie pamiętacie Państwo tego kodu PIN, oprócz charakterystycznej wskazówki, że dwie z cyfr kodu PIN powtarzały się (występowały dwukrotnie w zapomnianym kodzie)???

Proszę wskazać korzystniejsze rozwiązanie, potwierdzić je obliczeniami, a odpowiedź krótko uzasadnić.

*Liczba nieparzysta to taka liczba całkowita, której nie można podzielić przez 2 (przy dzieleniu przez 2 daje resztę 1).

Rekomendacja (szkic) rozwiązania:

1. Postawiony problem składa się z dwóch zasadniczych części (zdarzeń losowych n. Z1 i Z2) Wypada je na wstępie wskazać, poprawnie opisać. Celem jest zaś porównanie wartości $P(Z1)$ oraz $P(Z2)$ i wskazanie, wraz z uzasadnieniem, rozwiązania korzystniejszego.

2. Pierwsza część - jest pytaniem o prawdopodobieństwo otrzymania numeru telefonu do biura nowej formy, na zadanych warunkach. Zatem, szansa na otrzymanie numeru spełniającego warunek wstępny (dane) wynosi (prawdopodobieństwo klasyczne) odpowiednio $P(A)=112500/900000 \approx 0,125$ czyli 12,5%. Dlaczego? $A=9 \cdot 10 \cdot 10 \cdot 5 \cdot 5$, prawda? Podobnie wyznaczamy $\Omega=9 \cdot 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10$. Trzeba uzasadnić, skąd to się wzięło, potraficie?

3. Druga część - zaprasza do porównania uzyskanego wyniku z innym prawdopodobieństwem, Tym razem jest to pytanie o szansę na odgadnięcie PIN-u do drzwi za pierwszym razem. Prawdopodobieństwo (czyli szansę) tego zdarzenia trzeba wyznaczyć za pomocą permutacji z powtórzeniami (nie wiadomo przecież, które cyfry się powtarzają). Wskazano bowiem (dane), że aż dwie z cyfr kodu powtarzają się dwukrotnie!!! To trzeba poprawnie opisać.

3.1. W zadaniu pojawiła się PUŁAPKA. Mamy 3-cyfrowy PIN, w którym dwie cyfry kodu powtarzają się dwukrotnie. Taka sytuacja występuje tylko wówczas, gdy cyfry kodu są identyczne, zatem jest tylko 10 takich możliwości, gdyż nasze zdarzenie możemy opisać jako $B=\{000, 111, 222, 333, 444, 555, 666, 777, 888, 999\}=10$. Ponieważ $P(B)=B/\Omega$, potrzebujemy jeszcze wyznaczyć zbiór wszystkich możliwych zdarzeń. Zatem Ω wyznaczymy np. za pomocą reguły mnożenia dla 3 pól PINu, które może zapełnić każda z 10 cyfr.

3.2. Wydaje się że będzie to matematycznie tak: $\Omega=10 \cdot 10 \cdot 10$ (wiadomo, dlaczego?), a w konsekwencji $P(B)=10/1000 \approx 0,01$ czyli 1,0%. mamy jednak dodatkową wiedzę, że nasze zdarzenie oczekiwane B może wystąpić tylko jako jedna z 10 możliwości, gdyż $B=\{000, 111, 222, 333, 444, 555, 666, 777, 888, 999\}=10$. Można więc intuicyjnie zawęzić Ω , która w tym wypadku będzie równa B, zatem gdy $\Omega=B=10$, a my tak naprawdę mamy tylko 1 poprawny PIN, choć nie wiemy który, wówczas precyzujemy nowe zdarzenie $C=1$ (tylko jeden PIN jest poprawny, co wynika z zasady, którą zapamiętaliśmy), W efekcie $P(C)=1/10 \approx 0,1$ czyli 10,0%

4. Jaki wniosek wynika z porównania prawdopodobieństw? I tu znów ciekawostka:

Matematycznie $P(A)>P(B)$. Zatem, nieco łatwiej otrzymać wymarzony numer, a bardzo trudno trafić w niby prosty PIN.

Jednak logicznie trzeba uwzględnić nie $P(B)=1,0\%$, ale $P(C)=10,0\%$. Tymczasem nadal $P(A)>P(C)$. Zatem łatwiej uzyskać wymarzony numer, a trudniej trafić nawet dość prosty PIN.

5. Zostaliśmy zaproszeni na turniej gier losowych. Nie chcemy się spóźnić. Na naszym przystanku zatrzymują się 3 autobusy. Są już w drodze, a każdy przyjedzie w losowym czasie między 0 a 15 minut. Jaka jest szansa, że będziemy czekać na pierwszy autobus krócej niż 5 minut?
Uff, dojechaliśmy. Startuje turniej gier losowych. Gra wejściowa polega na tym, że: rzucając trzema uczciwymi, sześciennymi kostkami do gry, analizujemy dwa zdarzenia. Niech A oznacza zdarzenie polegające na tym, że na każdej kostce liczba wyrzuconych oczek jest większa niż 2, zaś B oznacza zdarzenie polegające na tym, że suma oczek jest podzielna przez 9. Proszę wyznaczyć prawdopodobieństwo sumy obu tych zdarzeń.
Czy uda nam się zakwalifikować do turnieju głównego?

Rekomendacja (szkic) rozwiązania:

Zadanie, jak to w życiu, ma swoje elementy składowe. najpierw trzeba zdążyć na turniej, następnie wygrać przynajmniej grę wejściową. Jak w "Milionerach"

Punktem wyjścia jest określenie problemu do rozwiązania, a jest nim określenie szansy (a zatem wyznaczenie prawdopodobieństwa) na udział w turnieju głównym gier losowych (E), pod dwoma warunkami: dotarcia na czas do salonu gier (T) oraz poprawnego rozwiązania zadania wstępnego (C), wymagającego oszacowania prawdopodobieństwa sumy zdarzeń, opisanych jako (A) oraz (B). Dobrze jest więc zapisać w postaci formuły logicznej problem, jaki mamy do rozwiązania.

1. KROK PIERWSZY. Transport publiczny, a w zasadzie każdy transport, zawsze obciążony jest ryzykiem. W naszym przypadku nie jest ono aż takie duże. W naszym przypadku musimy wyznaczyć $P(T)$.

Jak sobie z tym poradzić?

O powodzeniu rozwiązania decyduje poprawny opis zdarzenia losowego (a w szczególności jego elementów składowych). Zdarzeniami elementarnymi są trójki (x, y, z) , określające czasy przyjazdu trzech autobusów. Przestrzeń zdarzeń elementarnych jest więc podzbiorem R^3 :

W pierwszym etapie można przyjąć perspektywę prawdopodobieństwa klasycznego (pamiętamy formułę....). To się uda tylko wówczas, gdy poprawnie określimy, wykorzystując schematy kombinatoryczne: zbiór wszystkich możliwych zdarzeń (tzw. "omegę") oraz zbiór zdarzeń sprzyjających. Dopiero wówczas możliwe jest wyznaczenie prawdopodobieństwa, czyli $P(T)$. Poszukiwane zdarzenie (T) jest dokładnie tym samym, które pozwala zidentyfikować dobrze uchwycone zdarzenie przeciwne $\{A'\}$. Można więc na tej podstawie obliczyć, że szanse dotarcia na czas są znaczne, to ponad 70%! Finalnie wypada dać odpowiedź, którą zrozumie każdy pasażer czekający na autobus. A odpowiedź należy do Państwa: zdążymy zatem, czy nie?

2. KROK DRUGI: Dojechaliśmy wreszcie do salonu gier i startujemy w grze wejściowej (C). Trzeba zatem dobrze opisać zdarzenie losowe (C), które w istocie jest złożone z dwóch podobnych, ale... odrębnych zdarzeń. Poprawny opis sytuacji losowej prowadzi do wniosku, że konieczne jest wyznaczenie prawdopodobieństwa sumy zdarzeń: pierwszego (A) i drugiego (B). Jak sobie z tym poradzić? Przykładowo (uwaga na różnice w symbolice), tak:

Zmierzamy zatem do wyznaczenia $P(D)$, które można wyznaczyć jako sumę prawdopodobieństwa zdarzeń: $P(A)$ oraz $P(B)$. Trzeba podać poprawną formułę umożliwiającą takie obliczenia, uzasadniając, skąd też ona się wzięła.

3. Jeśli zatem:

- zdążyliśmy dojechać do salonu gier
- poprawnie oszacowaliśmy prawdopodobieństwo sumy zdarzeń w turnieju wejściowym na poziomie niemal 40% (czyli $P(C)=39,4\%$),

to z pewnością gospodarze - w nagrodę za poprawną odpowiedź - zaproszą nas do udziału w turnieju głównym. Jakie więc może być $P(E)=?$