Matematyka dla biologów — Zajęcia nr 10.

Dariusz Wrzosek

4 grudnia 2023

Podstawy kombinatoryki: permutacje, wariacje, kombinacje

Co to jest ciąg? -przypomnienie.

Niech X będzie zbiorem skończonym. Ciągiem elementów ze zbioru X nazywamy funkcję określoną na dowolnym podzbiorze L zbioru liczb naturalnych $\mathbb{N}$ o wartościach w X czyli każdej liczbie ze zbioru L przyporządkowany jest dokładnie jeden element zbioru X.

Dla podkreślenia tego, że elementy ciągu są uporządkowane (bo są ponumerowane) nazywa się je wyrazami.

Jeśli funkcja określająca ciąg jest różnowartościowa to wtedy wyrazy ciągu są różne (nie powtarzają się).

Rozróżnienie -ciągi a podzbiory

Rozpatrzmy zbiór liter $X = \{a\,, b\,, c\}$. Oto wszystkie podzbiory dwuelementowe tego zbioru (pary nieuporządkowane)

$$\{a, b\}\,,\ \{a, c\}\ \{b\,, c\}$$

a wszystkie ciągi dwuwyrazowe elementów ze zbioru X

(pary uporządkowane) to

$$(a\,, a)\,,\ (a\,, b)\,,\ (a\,, c)\,,\ (b\,, b)\,,\ (b\,, a)\,,\ (b\,, c)\,,\ (c\,, a)\,,\ (c\,, b)\,,(c\,, c)\,.$$

Wsród pytań, które można zadać odnośnie ciągów skończonych wybierzmy dwa;

1. Ile jest wszystkich ciągów p-wyrazowych o wyrazach ze zbioru m-elementowego? Czyli przyjmując oznaczenie, że znaczek $\sharp L$ oznacza liczebność zbioru L mamy

$$\sharp L = p\,, \quad \sharp X = m\,.$$

(Uwaga: wyrazy ciągu mogą się powtarzać)

Odpowiedź. Jest ich tyle ile jest wszystkich funkcji określonych na zbiorze p-elementowym o wartościach w zbiorze m elementowym czyli $m^p = \overbrace{m \cdot m \cdot \ldots m}^{p}$ gdyż pierwszy wyraz ciągu można wybrać na m sposobów, drugi również na m sposobów i tak dalej na każdym spośród p miejsc.

Wariacja bez powtórzeń

2. Ile jest wszystkich ciągów p wyrazowych ze zbioru m-elementowego gdy $m \geqslant p$ i takich, że wyrazy w ciągu nie powtarzają się?

Innymi słowy, ile jest funkcji różnowartościowych o dziedzinie p-elementowej o wartościach ze zbioru m-elementowego. Każdy taki ciąg nazywa się w kombinatoryce wariacją bez powtórzeń.

Odpowiedź. Zbiór wszystkich ciągów p wyrazowych i różnowartościowych ze zbioru m-elementowego (czyli wariacji bez powtórzeń) ma

$$m \cdot (m-1) \cdot (m-2) \cdot (m-3) \cdot \ldots (m-(p-1)) \overset{ozn.}{=} V_m^p \quad \text{elementów,}$$

gdyż pierwszy wyraz ciągu można wybrać na m sposobów, ale drugi, skoro nie mogą się powtarzać, już tylko na $m-1$ sposobów. Kolejny na $m-2$ sposobów i tak dalej p-ty wyraz może być wybrany na $m-(p-1)$ sposobów, bo dokładnie $p-1$ elementów zostało już wybranych i ustalonych jako kolejne wyrazy ciągu.

Permutacje

W przypadku szczególnym gdy $m = p$, wariację bez powtórzeń nazywamy permutacją i wtedy

$$V_m^m = 1 \cdot 2 \cdot \ldots (m-1)m \overset{ozn.}{=} m!$$

Przykłady

Ile jest wszystkich możliwych ciągów zasad azotowych w 10 elementowym fragmencie łańcucha RNA?

Odpowiedź. Zbiór zasad azotowych UACG jest czteroelementowy. Na każdym miejscu łańcucha RNA może być dowolna spośród czterech zasad, a więc wszystkich możliwych jest tyle ile ciągów 10-wyrazowych o wartościach ze zbioru czteroelementowego czyli $4^{10} = 1048576$.

Ograniczmy się teraz do czteroelementowych fragmentów łańcucha i zapytajmy **ile jest 4-elementowych łańcuchów, w których żadne dwie zasady się nie powtarzają.** Jest ich oczywiście $1 \cdot 2 \cdot 3 \cdot 4 = 4! = 24$.

Kombinacje

Kombinacja

k-elementową kombinacją ze zbioru n-elementowego nazywamy każdy k elementowy podzbiór tego zbioru. **Liczbę kombinacji k-elementowych ze zbioru n-elementowego** , oznaczamy przez C_n^k.

Udowodnimy, że

$$C_n^k = \binom{n}{k} \overset{ozn.}{=} \frac{n!}{(n-k)!k!} .$$

Symbol $\binom{n}{k}$ to tzw. symbol Newtona, który czytamy "n po k".

Dowód:

Spróbujmy najpierw znaleźć liczbę wszystkich ciągów k-wyrazowych o elementach ze zbioru n elementowego, $n \geqslant k$, czyli liczbę wszystkich wariacji k-wyrazowych ze zbioru n-elementowego. Pierwszy wyraz ciągu możemy wybrać na n sposobów drugi już tylko na $n-1$ sposobów i tak k-ty wyraz na $n-(k-1)$ sposobów. Teraz pozostaje uwzględnić, że rozróżnialiśmy ciągi w których występują te same elementy w różnej kolejności. Liczbę wszystkich wariacji bez powtórzeń musimy zmniejszyć tyle razy ile jest permutacji k-wyrazowych. Wszystkich kombinacji jest zatem

$$\frac{n(n-1)\cdot\ldots(n-(k-1))}{k!} = \frac{n!}{(n-k)!k!}.$$

Słowem nazywać będziemy dowolny ciąg liter (znaków) jakiegoś alfabetu. Oto kilka pytań, które można postawić odnośnie liczby słów mających określone własności.

1. Ile różnych słów dziesięcioliterowych można utworzyć z dziesięciu różnych liter tak aby litery nie powtarzały się w słowie?
2. Na ile sposobów można wybrać trójki (nie uporządkowane) niepowtarzających się liter spośród 10 znaków alfabetu?
3. Na ile sposobów można wybrać słowa trzyliterowe o niepowtarzających się literach spośród 10 znaków alfabetu?
4. Ile słów trzyliterowych można utworzyć z 10 liter alfabetu?

W pierwszej chwili może się wydawać, że trzy ostatnie pytania dotyczą tego samego zagadnienia. Tak jednak nie jest. Spróbujmy wyrazić te pytania w języku matematyki. Pierwsze pytanie dotyczy w istocie liczby wszystkich ciągów 10-wyrazowych i różnowartościowych lub używając specyficznej terminologii kombinatorycznej liczby wszystkich **permutacji** 10 wyrazowych. Jest ich

$$V_{10}^{10} = 1 \cdot 2 \cdot 3 \cdot \ldots 7 \cdot 8 \cdot 9 \cdot \cdot 10 = 10! = 3\,628800\,.$$

W drugim pytaniu chodzi o liczbę wszystkich podzbiorów 3-elementowych zbioru 10-elementowego bez ustalenia jakiegokolwiek porządku ich występowania.

W naszym przypadku

$$C_{10}^{3} := \binom{10}{3} = \frac{10!}{7!\,3!} = \frac{1\cdot2\cdot3\cdot\ldots7\cdot8\cdot9\cdot\cdot10}{(1\cdot2\cdot3\cdot\ldots7)\cdot2\cdot3} = \frac{8\cdot9\cdot10}{6} = 4\cdot3\cdot10 = 120$$

Powyższe rozumowanie zawiera odpowiedź na pytanie trzecie. Słów trzyliterowych o niepowtarzających się literach jest tyle ile jest wariacji trzywyrazowych bez powtórzeń czyli $V_{3}^{10} = 10\cdot9\cdot8 = 720$.

W pytaniu czwartym chodzi zaś o liczbę wszystkich ciągów trzywyrazowych o wyrazach ze zbioru 10-elementowego (litery mogą się powtarzać) czyli liczbę wszystkich funkcji określonych na zbiorze trzyelementowym w zbiór 10-elementowy. Jest ich tyle ile jest wszystkich trzywyrazowych wariacji z powtórzeniami ze zbioru 10-elementowego czyli $10\cdot10\cdot10 = 10^3$.

Pary alleli na chromosomach

Przyjmijmy, że w pewnej populacji organizmów diploidalnych występuje n odmian danego genu czyli alleli. Najczęściej w szkolnych zadaniach spotykamy się z przypadkiem dwóch alleli. W ogólności alleli może być więcej, jest tak na przykład u grzybów podstawczaków w genach odpowiedzialnych za rozmnażanie. Pojawia się pytanie

Ile jest par alleli jednego genu ulokowanych w odpowiadających sobie miejscach na chromosomach homologicznych, jeśli wszystkich alleli danego genu jest n?

Odpowiedź. Zwróćmy uwagę, że z punktu widzenia klasycznych praw genetyki kolejność występowania alleli w parze nie ma znaczenia (podobnie do liczby oczek na kościach domina). Wszystkich ciągów dwuwyrazowych (par) o wyrazach ze zbioru n elementowego jest n^2. I taka byłaby odpowiedź jeśli rozróżnialibyśmy pary typu (A,a) i (a,A). By wyeliminować te przypadki odejmiemy najpierw od n^2 liczbę wszystkich par mających te same alle na obu miejscach np. (A,A). Jest ich oczywiście n. Ponieważ każde dwie pary różniące się tylko kolejnością traktujemy jako jedną to jest ich $\frac{n^2-n}{2}$. Do tej liczby trzeba tylko dodać pary mające te same alle na obu miejscach. Ostateczną odpowiedzią jest zatem

$$\frac{n^2-n}{2}+n=\frac{n(n+1)}{2}=\binom{n+1}{2}\ .$$

W znanym przypadku dwóch alleli mamy trzy ”genotypy”: AA, Aa, aa. Dla pięciu alleli możliwych 15 genotypów.

Przewidywalne/nieprzewidywalne

Co to jest rachunek prawdopodobieństwa?

Rachunek prawdopodobieństwa to dziedzina matematyki zajmująca się badaniem modeli zjawisk i procesów o przebiegu nie dającym się przewidzieć z całkowitą pewnością. Takie zjawiska nazywamy **zjawiskami losowymi**. Nieprzewidywalny jest wynik rzutu monetą lub kością, wynik meczu piłkarskiego, znalezienie wadliwej żarówki wśród nowo wyprodukowanych, wynik pomiaru wysokości przypadkowo wybranego studenta a także wystąpienie mutacji genetycznej.

Rzut monetą symetryczną

W pustym pokoju podrzucamy w górę symetryczną monetę. Zanim moneta upadnie wyrażamy dwie opinie;

1) moneta spadnie na podłogę

2) moneta spadnie reszką do góry .

W pierwszym przypadku mamy pewność, że sąd jest prawdziwy, a w drugim nie !

Chcąc mimo wszystko znaleźć jakąś regułę opisującą wyniki rzutu monetą i chcąc uwolnić się od konsekwencji oddziaływania wielkiej liczby czynników wpływających na przebieg każdego rzutu z osobna należy wykonać wielką liczbę powtórzeń i przeformułować sąd 1), który teraz brzmiałby następująco

1') W trakcie 1000 rzutów tą samą monetą w tych samych warunkach liczba reszek które wypadną mieści się w przedziale [490,510].

Rachunek prawdopodobieństwa umożliwia określenie stopnia pewności z jakim można uznać, że sąd 1') przewiduje rezultat konkretnej serii rzutów.

Odpowiednikiem wyrzucenia reszki bądź orła może być wystąpienie bądź niewystąpienie w ciągu tysiąca pokoleń, mutacji genetycznej w ustalonej części łańcucha DNA, w komórce generatywnej jakiegoś organizmu.

Wystąpienie mutacji w ciągu kolejnego miliona lat to jakby kolejny rzut monetą itd.

Różnica pomiędzy tym przykładem i serią rzutów monetą polega na tym, że w odróżnieniu od tego ostatniego nie możemy powtarzać wielokrotnie przebiegów ewolucji aby oszacować prawdopodobieństwo pojawienia się mutacji.

Trzy powody dla których biolog winien znać podstawy rachunku prawdopodobieństwa:

1. Podstawowym źródłem przypadkowej zmienności w populacjach organizmów dwupłciowych są dwa procesy towarzyszące mejozie, w trakcie pierwszego z nich następuje wymiana części chromatyd chromosomów homologicznych pochodzących od obojga rodziców (crossing-over), w drugim dochodzi do losowego rozchodzenie się chromatyd do biegunów komórki. Oba procesy są od siebie niezależne a ich rezultat nie jest przewidywalny o czym przekonać może się każdy kto ma rodzeństwo pochodzące od tej samej pary rodzicielskiej.
2. Drugim czynnikiem powodującym przypadkową zmienność (a także specjację) jest występowanie mutacji genetycznych. Zarówno miejsce w genomie jaki moment w którym wystąpi, są zjawiskami losowymi.
3. Powyższe czynniki losowe wydają się być immanentną cechą procesów naturalnych. Innej natury czynnikiem losowym, nieprzewidywalnym, jest wpływ nieświadomych błędów i niedokładności towarzyszących każdemu aktowi pomiaru.

Przestrzeń zdarzeń elementarnych

W każdym z przytoczonych wyżej przykładów zjawisk losowych potrafimy sporządzić listę wszystkich możliwych sytuacji, które mogą się pojawić w taki sposób by każda z nich wykluczała pozostałe. Każda taka sytuacja nazywa się **zdarzeniem elementarnym**. Jest to pojęcie pierwotne rachunku prawdopodobieństwa tzn. nie definiuje się go w języku teorii, której dotyczy.

Wszystkie zdarzenia elementarne dotyczące wyników danego doświadczenia losowego tworzą **zbiór zdarzeń elementarnych**. Jeśli zbiór zdarzeń elementarnych jest zbiorem skończonym to **zdarzeniem** może być dowolny podzbiór zbioru zdarzeń elementarnych, a w przypadku gdy zbiór zdarzeń elementarnych jest zbiorem nieskończonym, trzeba wprowadzić dodatkowe warunki ograniczające, które musi spełniać podzbiór będący zdarzeniem. Jest to niezbędne dla matematycznej spójności teorii ale nie ma bezpośredniego praktycznego znaczenia.

Przykłady przestrzeni zdarzeń

- dla doświadczenia polegającego na jednokrotnym rzucie monetą zbiór zdarzeń elementarnych jest dwuelementowy

$$\Omega_1 = \{0\,,R\}$$

  gdzie 0 reprezentuje wyrzucenie orła a R reszki.
- dla doświadczenia polegającego na dwukrotnym rzucie monetą zbiór zdarzeń elementarnych jest czteroelementowy i składa się z wszystkich możliwych serii wyników

$$\Omega_2 = \{(0, R)\,,(0, 0)\,,(R\,, R)\,,(R\,, 0)\}\,,$$

- w przypadku pojedyńczego rzutu kością zbiór zdarzeń elementarnych jest sześcioelementowy.

Wszystkie operacje, które znamy z rachunku zbiorów mają swoje odpowiedniki w rachunku zdarzeń.

Dla przykładu jeśli Ω jest zbiorem zdarzeń elementarnych i $A \subset \Omega$ to zbiór $\bar{A} = \Omega \setminus A$ nazywamy zdarzeniem przeciwnym. Jeśli zachodzi zarówno zdarzenie A jaki zdarzenie B to takie zdarzenie nazywa się, podobnie jak w rachunku zbiorów, iloczynem (częścią wspólną) zdarzeń $A \cap B$ i.t.d. W przypadku dwukrotnego rzutu monetą jeśli przez A oznaczymy zdarzenie polegające na wyrzuceniu orła w pierwszym rzucie a przez B zdarzenie polegające na wyrzuceniu orła w drugim rzucie to zdarzenie polegające na wyrzuceniu orła w obu rzutach to

$$A \cap B = \{(0, R)\,, (0, 0)\} \cap \{(0, 0)\,, (R\,, 0)\} = \{(0\,, 0)\}\,.$$

Prawdopodobieństwo a częstość

Pojęcie prawdopodobieństwa jest uogólnieniem pojęcia częstości. Jeśli przy n-krotnym powtórzeniu danego doświadczenia zdarzenie A zachodzi n_A razy to liczbę $c_A = n_A/n$ nazywamy **częstością występowania** zdarzenia A. Rozważmy dla przykładu sekwencję zasad nukleinowych w łańcuchu DNA. Przypomnijmy, że ze względu na podobieństwo budowy cząsteczek wyróżnia się puryny: adeninę ozn. A i guaninę ozn. G oraz pyrymidyny cytozynę ozn. C i tyminę ozn. T. Zasady A i T oraz G i C tworzą pary komplementarne. W 12-wyrazowej sekwencji zasad

$$AGCTGGCGACTA$$

częstość występowania adeniny A wynosi $\frac{1}{4}$, częstość występowania guaniny G wynosi $\frac{1}{3}$, a częstość występowania puryn wynosi $\frac{1}{4} + \frac{1}{3} = \frac{7}{12}$. Widzimy zatem, że częstości zdarzeń rozłącznych dodają się do siebie.

Aksjomaty rachunku prawdopodobieństwa

Przyjmując aksjomaty sformułowane przez Andrieja Kołmogorowa (1896-1945) można uwolnić się od źródłowego pojęcia częstości.

Niech Ω będzie przestrzenią zdarzeń elementarnych. Prawdopodobieństwo zdarzenia A oznaczamy przez $P(A)$.

1. Dla dowolnego zdarzenia $A \subset \Omega\,,\ 0 \leqslant P(A) \leqslant 1\,.$
2. $P(\Omega) = 1$
3. Dla każdej pary rozłącznych zdarzeń A i B

$$P(A \cup B) = P(A) + P(B)\,.$$

Drugi aksjomat mówi, że zajście któregoś ze zdarzeń spośród wszystkich możliwych jest pewne tzn. że zbiór Ω uwzględnia wszystkie możliwe zdarzenia, które mogą zajść w danych okolicznościach.

W przypadku rzutu monetą spadającą na płaską poziomą powierzchnię są tylko dwie możliwości i po wykonaniu rzutu na pewno któraś z nich się zrealizuje czyli prawdopodobieństwo tego, że wypadnie orzeł lub reszka równe jest 1. Z aksjomatów 2 i 3 wynika, że dla dowolnego zdarzenia $A \subset \Omega$

$$P(A) + P(\Omega \setminus A) = 1\,. \tag{1}$$

A więc jest pewne, że zdarzy się A lub zdarzy się zdarzenie przeciwne $\Omega \setminus A$.

W pełnej matematycznej teorii prawdopodobieństwa uwzględniającej przypadek gdy zbiór Ω jest zbiorem nieskończonym np. zbiorem liczb rzeczywistych, ostatni aksjomat jest niewystarczający i rozszerza się go także na nieskończone (przeliczalne) sumy zdarzeń parami rozłącznych. Konsekwencji tego nie będziemy dalej rozważać ograniczając się na razie do przypadku **prawdopodobieństwa dyskretnego** tzn. sytuacji gdy zbiór Ω ma skończenie wiele elementów.

Trzeba podkreślić, że prawdopodobieństwo jest funkcją określoną na zbiorze zdarzeń, czyli podzbiorów zbioru Ω i z tego powodu, bez wcześniejszego ustalenia konwencji notacyjnej, zapis $P(\omega)$ dla określenia prawdopodobieństwa zdarzenia elementarnego $\omega \in \Omega$ jest niepoprawny. Powinno zapisać się $P(\{\omega\})$

W danym zbiorze zdarzeń elementarnych prawdopodobieństwo można wprowadzić na wiele różnych sposobów byle spełnione były aksjomaty. Najprostsza sytuacja jest wtedy, gdy zdarzenia elementarne są jednakowo prawdopodobne. Rozpatrzmy n-elementową przestrzeń zdarzeń elementarnych

$$\Omega = \{\omega_1 , \omega_2 , \ldots , \omega_n\}$$

Wtedy dla każdego $i \in \{1 , \ldots n\}$ mamy

$$P(\{\omega_i\}) = \frac{1}{n} .$$

Konsekwentnie jeśli jakieś zdarzenie $A \subset \Omega$ ma m elementów to

$$P(A) = \frac{m}{n} .$$

W dwuelementowym zbiorze zdarzeń elementarnych $\Omega_{0R} \overset{ozn.}{=} \{0 , R\}$ możemy zadać prawdopodobieństwo P_1 przyjmując, że zdarzenia elementarne są jednakowo prawdopodobne

$$P_1(\{0\}) = \frac{1}{2} = P_1(\{R\})$$

lub przyjąć, że

$$P_2(\{0\}) = 0.499\,, \quad P_2(\{R\}) = 0.501\,.$$

Przestrzeń probabilistyczna $(\Omega_{0R}\,, P_1)$ jest **modelem probabilistycznym** dla doświadczenia losowego polegającego na jednokrotnym rzucie monetą idealnie symetryczną.

Czy to jest dobry model przekonać można się tylko wykonując doświadczenia w postaci długiej serii powtórzeń rzutu monetą w tych samych warunkach. Matematyczna teoria prawdopodobieństwa nie zajmuje się w zasadzie tym czy dany model probabilistyczny dobrze opisuje przebieg konkretnego doświadczenia. Dostarcza jedynie ogólnej teorii dającej podstawy do prowadzenia takich badań i daje także podstawy do badań statystycznych. Jednym z celów statystki jest oszacowanie na podstawie danych empirycznych prawdopodobieństw różnych zjawisk. Tak otrzymane prawdopodobieństwa konkretyzują model probabilistyczny. Po przeprowadzeniu dostatecznie wielu prób może się okazać, że przestrzeń probabilistyczna (Ω_{0R}, P_2) lepiej opisuje rzeczywisty przebieg doświadczenia, a więc moneta nie jest idealnie symetryczna.

1. Zadanie do zrobienia na zajęciach

1. Rozważmy zbiór zdarzeń elementarnych Ω mający n elementów. Ile jest zatem wszystkich możliwych zdarzeń. Uzasadnić, że jest ich tyle ile jest wszystkich funkcji ze zbioru n-elementowego w zbiór dwuelementowy czyli ?
2. Korzystając z powyższego uzasadnić, że

$$2^n = \sum_{k=0}^{n} \binom{n}{k} \ .$$

2. Zadanie do zrobienia na zajęciach

Cztery małżeństwa chcemy usadzić na ośmiu krzesłach stojących w szeregu. Na ile sposobów można to zrobić,

1. nie bacząc kto koło kogo siedzi,
2. tak aby małżonkowie siedzieli obok siebie,
3. tak aby mężowie siedzieli razem w jednej grupie obok siebie, a żony w drugiej,
4. tak aby żona siedziała obok swojego męża i jakiejś innej żony.