II Liceum Ogólnokształcące im. Króla Jana III Sobieskiego

ul. Sobieskiego 9, 31-136 Kraków

tel. 12 633-73-92

Rozkład liczby serii i jego zastosowania

Anna Szczepańska

Praca zgłoszona do Konkursu Prac Matematycznych dla młodzieży szkół ponadgimnazjalnych, gimnazjów i szkół podstawowych organizowanego przez Krakowskie Młodzieżowe Towarzystwo Przyjaciół Nauk i Sztuk w Krakowie.

Kraków 2012

Spis treści

Wstęp

W oparciu o dostępną literaturę [1], [2] przedstawiamy zagadnienie związane z badaniem rozkładu liczby serii i jego zastosowanie w statystyce do badania, czy rozkład pewnych wielkości w dwóch populacjach jest jednakowy.

Rozważania rozpoczniemy od obserwacji liczby możliwych sposobów przejścia pomiędzy dwoma wskazanymi polami w prostokątnej tabeli. Zauważymy, że każdą drogę można jednoznacznie opisać za pomocą ciągu znaków p (w prawo) oraz d (w dół). Obliczymy na tej podstawie liczbę najkrótszych dróg w tabeli o k kolumnach i w wierszach, za pomocą wzoru:

$$D = \frac{(k-1+w-1)!}{(k-1)!\,(w-1)!}$$

Zauważymy, że liczba dróg jest największa w tabelach, w których różnica $|k-w|$ jest mała.

Zbadamy rozkład liczby serii w ciągach złożonych ze znaków p oraz d.

Zaproponujemy zastosowanie rozkładu liczby serii do testu Walda-Wolfowitza. Za pomocą tego testu sprawdzimy, czy rozkład ocen szkolnych z pewnego przedmiotu nie zależy od płci.

Pewna łamigłówka

Zastanówmy się, na ile sposobów można przejść od górnego lewego rogu w prostokątnej tabeli do jej dolnego prawego rogu tak, aby po drodze przeczytać słowo DOMEK (w przykładzie 1 i 2) lub KRAKÓW (w przykładzie 3 i 4).

Przykład 1:

D	O	M	E
O	M	E	K

Przykład 2:

D	O	M
O	M	E
M	E	K

Przykład 3:

K	R	A	K	Ó
R	A	K	Ó	W

Przykład 4:

K	R	A	K
R	A	K	Ó
A	K	Ó	W

Zauważmy, że liczba możliwości w każdym z przykładów nie zależy jedynie od liczby pól w danej tabeli, ale również od sposobu ich rozmieszczenia. W pierwszym przykładzie wyraz DOMEK możemy przeczytać na cztery sposoby:

D	O	M	E
O	M	E	K

D	O	M	E
O	M	E	K

D	O	M	E
O	M	E	K

D	O	M	E
O	M	E	K

W drugim przykładzie, w którym tabela jest kwadratowa, takich dróg jest już sześć:

D	**O**	**M**
O	M	**E**
M	E	**K**

D	**O**	M
O	**M**	**E**
M	E	**K**

D	**O**	M
O	**M**	E
M	**E**	**K**

D	O	M
O	**M**	**E**
M	E	**K**

D	O	M
O	**M**	E
M	**E**	**K**

D	O	M
O	M	E
M	**E**	**K**

W trzecim przykładzie wyraz KRAKÓW możemy ułożyć na pięć sposobów:

K	**R**	**A**	**K**	**Ó**
R	A	K	Ó	**W**

K	**R**	**A**	**K**	Ó
R	A	K	**Ó**	**W**

K	**R**	**A**	K	Ó
R	A	**K**	**Ó**	**W**

K	**R**	A	K	Ó
R	**A**	**K**	**Ó**	**W**

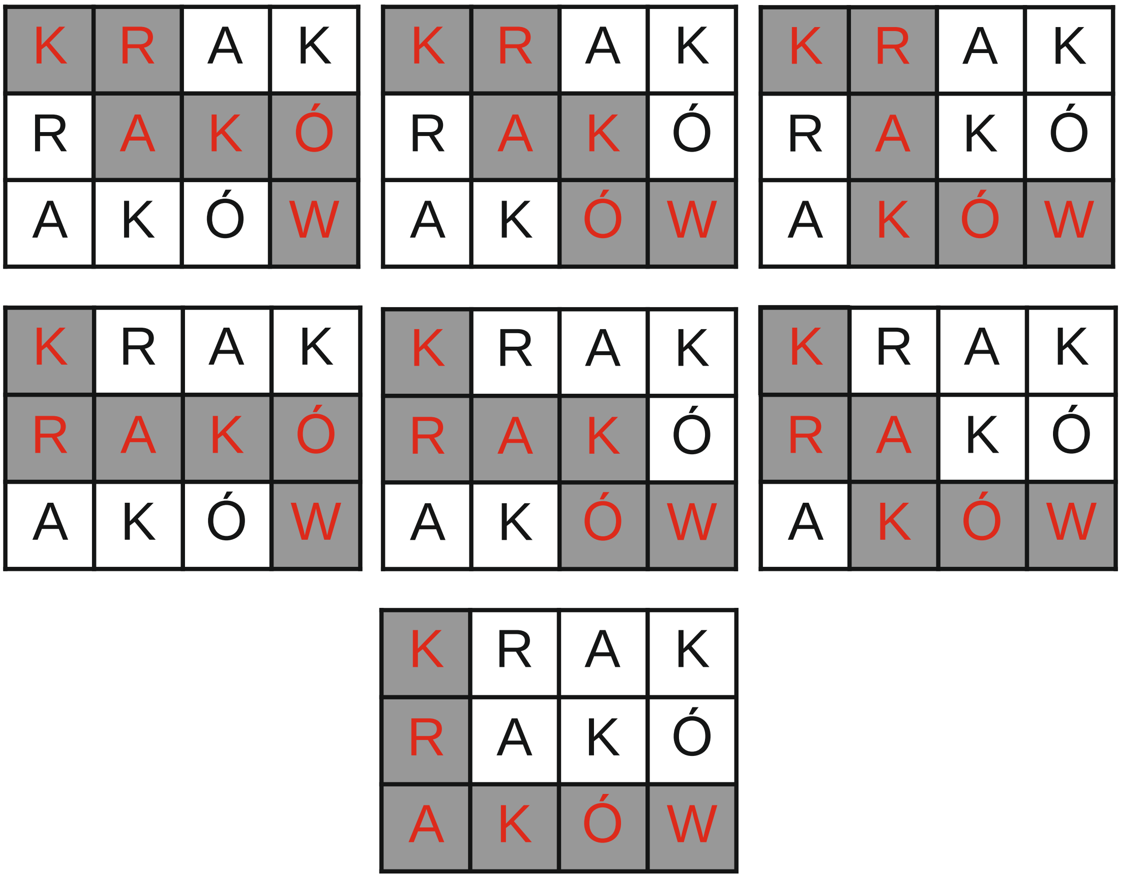
K	R	A	K	Ó
R	**A**	**K**	**Ó**	**W**

W czwartym przykładzie wyraz KRAKÓW możemy ułożyć na dziesięć sposobów:

K	**R**	**A**	**K**
R	A	K	**Ó**
A	K	Ó	**W**

K	**R**	**A**	K
R	A	**K**	**Ó**
A	K	Ó	**W**

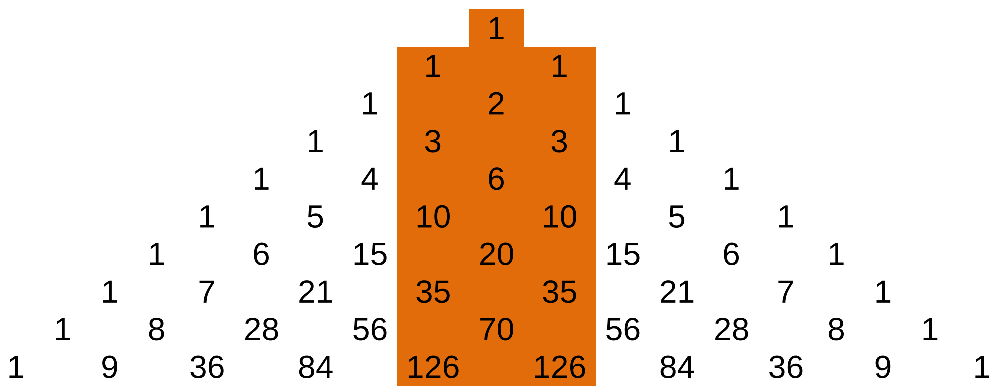
K	**R**	**A**	K
R	A	**K**	Ó
A	K	**Ó**	**W**



Zauważmy, że jeśli każdy krok w prawo oznaczymy literą p, a każdy krok w dół – literą d, to liczba wszystkich ciągów złożonych z $k-1$ znaków p oraz $w-1$ znaków d wyraża wzór:

$$D = \frac{(k-1+w-1)!}{(k-1)!\,(w-1)!} = \binom{k+w-2}{k-1} = \binom{k+w-2}{w-1}.$$

Liczba ta wyraża, więc taką liczbę wszystkich sposobów przejścia od lewego górnego rogu w tabeli o k kolumnach i w wierszach do jej dolnego prawego rogu. Zauważmy, że tym więcej jest sposobów przejścia od lewego górnego rogu do prawego dolnego rogu, im mniejsza jest różnica między liczbą kolumn i wierszy w prostokątnej tabeli. Pamiętajmy, że przy ustalonej wartości sumy $k-1+w-1$, symbol Newtona osiąga największą wartość wtedy, gdy $k=w$ lub kiedy liczby k i w różnią się o jeden:

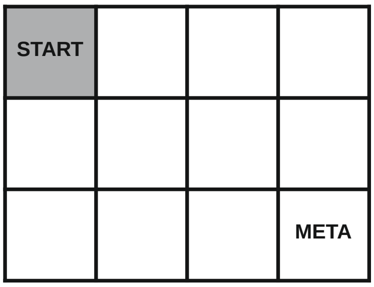


Badanie liczby serii

Rozważmy prostokątną tabelę złożoną z k kolumn i w wierszy. Zauważyliśmy, że liczba najkrótszych dróg, które prowadzą od pola w lewym górnym rogu z napisem START, do prawego dolnego rogu z napisem META, wyraża wzór:

$$D = \frac{(n_p + n_d)!}{n_p!\, n_d!},$$

w którym $n_p = k - 1$ jest liczbą kroków w prawo, a $n_d = w - 1$ jest liczbą kroków w dół.



Rysunek 1.

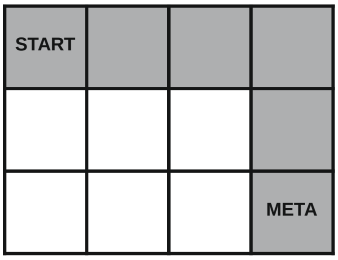
Na rysunku 1 przedstawiamy tabelę o czterech kolumnach i trzech wierszach, czyli $n_p = 3$, a $n_d = 2$. W przypadku, gdy $n_p = 3$, a $n_d = 2$ mamy $D = 10$.

Możliwości te wypiszemy w tabeli 1, zapisując każdy krok w prawo literą p, a każdy krok w dół literą d.

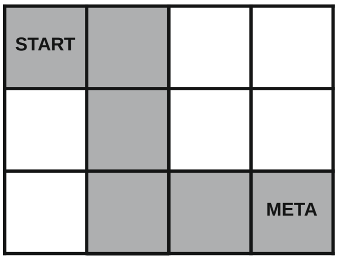
pppdd	*pddpp*
ppdpd	*dpppd*
ppddp	*dppdp*
pdppd	*dpdpp*
pdpdp	*ddppp*

Tabela 1.

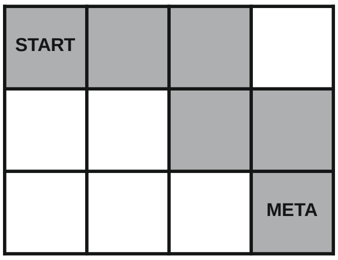
Poniżej znajdują się ilustracje, które przedstawiają wypisane przez nas drogi:



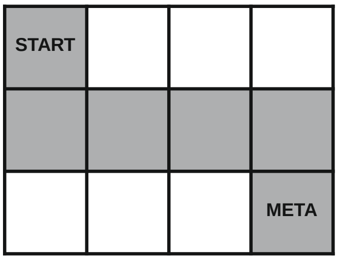
pppdd



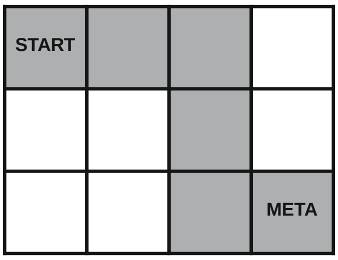
pddpp



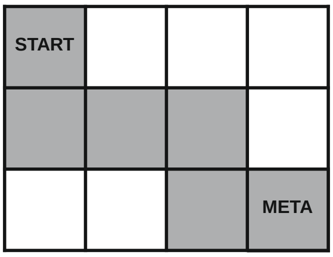
ppdpd



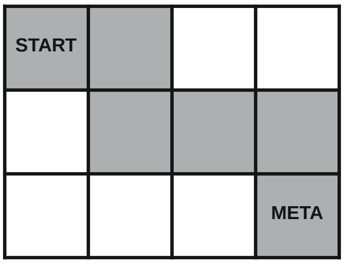
dpppd



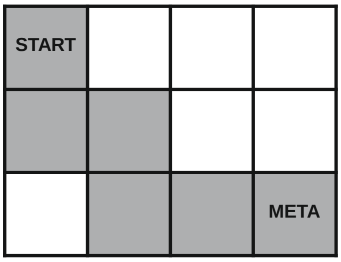
ppddp



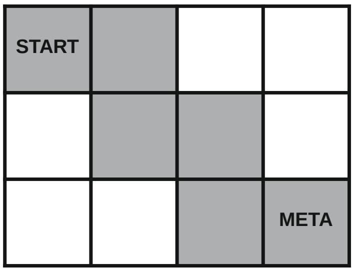
dppdp



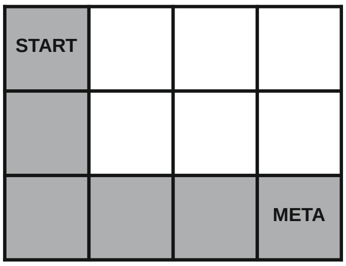
pdppd



dpdpp



pdpdp



ddppp

Mimo że w każdym z przypadków wykonujemy dokładnie 5 kroków, z tego 3 w prawo i 2 w dół, to drogi te różnią się między sobą. Różnią się kolejnością wykonywania kroków w prawo i do dołu oraz liczbą zmian kierunku.

Przykładowo, idąc drogą *pppdd* zmieniamy tylko raz kierunek. Natomiast, gdy wybierzemy drogę *pdpdp,* będziemy zmieniać kierunek po każdym kroku, czyli cztery razy.

Spójrzmy jeszcze raz na wypisane drogi i przypiszmy każdej z nich liczbę występujących serii s (tabela 2). Przez *serię* rozumiemy ciąg jednakowych znaków występujących bezpośrednio po sobie.

droga	**liczba serii**	**droga**	**liczba serii**
pppdd	2	pddpp	3
ppdpd	4	dpppd	3
ppddp	3	dppdp	4
pdppd	4	dpdpp	4
pdpdp	5	ddppp	2

Tabela 2.

Niech D_k oznacza liczbę dróg, w których występuje dokładnie k serii. Zauważmy, że liczba serii zawsze wynosi co najmniej 2. W rozważanym przykładzie mamy:

$$D_2 = 2$$
$$D_3 = 3$$
$$D_4 = 4$$
$$D_5 = 1.$$

Widzimy, że więcej jest dróg, w których występują trzy lub cztery serie, niż dróg, gdzie mamy dwie lub pięć serii.

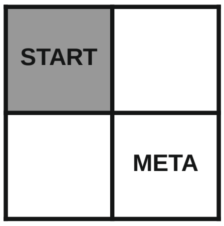
Spójrzmy również na wartości D_2, D_3, D_4, D_5, dla tabel o trzech wierszach (tzn. $n_d = 2$) i o $n_p + 1$ kolumnach (zob. tabela 3), gdzie $n_p \in \{1, 2, 3, 4, 5\}$:

n_p	1	2	3	4	5
D_2	2	2	2	2	2
D_3	1	2	3	4	5
D_4	-	2	4	6	8
D_5	-	-	1	3	6

Tabela 3. *Wartości D_2, D_3, D_4, D_5 dla tabel o trzech wierszach ($n_d = 2$).*

Zauważmy, że bez względu na to, ile wynosi n_p, liczba D_2 zawsze jest równa 2. Możemy zatem wnioskować, że istnieją co najmniej dwie drogi, w których są dwie serie, w każdym przypadku gdy ciąg utworzony jest z dwóch rodzajów znaków. Spójrzmy na inne przykłady.

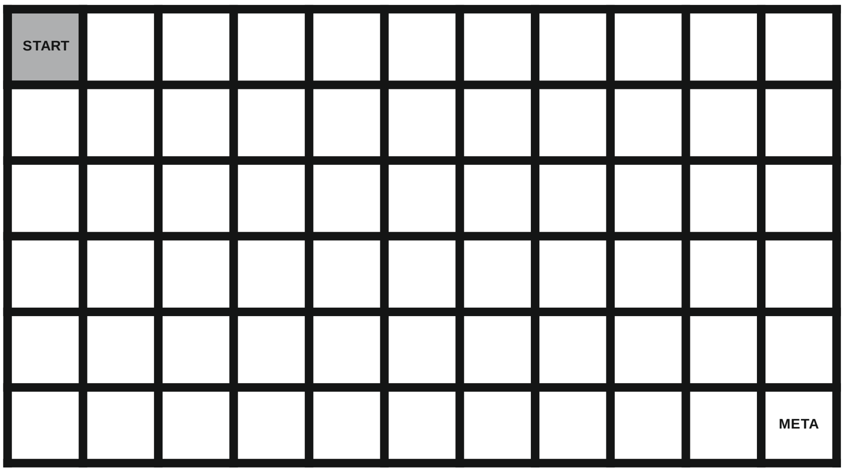
Gdy $n_d = 1$ i $n_p = 1$ (rysunek 2):



Rysunek 2.

Możliwymi drogami są pd, gdzie $s = 2$ oraz dp, gdzie $s = 2$.

Gdy $n_d = 5$ i $n_p = 10$ (rysunek 3):



Rysunek 3.

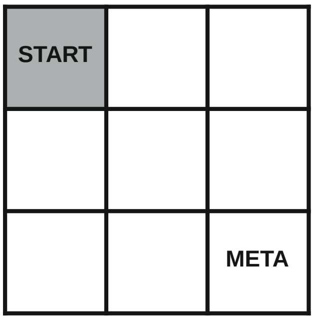
W tym przypadku również mamy jedynie dwie drogi, w których występują dwie serie: *ppppppppppddddd, dddddpppppppppp*.

Maksymalna liczba serii

Zauważyliśmy, że minimalna liczba serii wynosi dwa. Zastanówmy się, czy istnieje górne ograniczenie liczby serii.

Rozpatrzmy trzy przykłady.

Przykład A. Weźmy $n_d = 2$ i $n_p = 2$ (rysunek 4):



Rysunek 4.

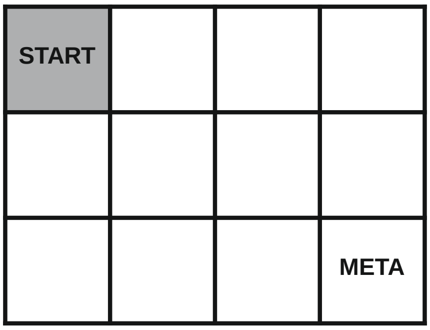
Możliwe drogi i liczbę serii w każdej z dróg podaje poniższa tabela:

droga	**liczba serii**	**droga**	**liczba serii**
ppdd	2	ddpp	2
pdpd	4	dpdp	4
pddp	3	dppd	3

Tabela 4.

Przez s_{max} oznaczmy największą liczbę serii. Zauważmy, że gdy $n_p = 2$ i $n_d = 2$, to maksymalna liczba serii $s_{max} = 4$.

Przykład B. Rozważmy teraz przypadek, w którym $n_d = 2$ i $n_p = 3$ (rysunek 5):



Rysunek 5.

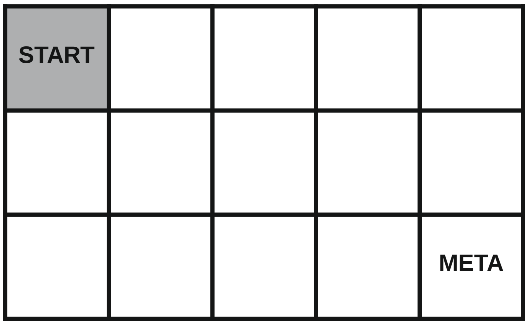
Możliwe drogi i liczbę serii w każdej z dróg podajemy w poniższej tabeli:

droga	**liczba serii**	**droga**	**liczba serii**
pppdd	2	pddpp	3
ppdpd	4	dpppd	3
ppddp	3	dppdp	4
pdppd	4	dpdpp	4
pdpdp	5	ddppp	2

Tabela 5.

Gdy $n_p = 3$ i $n_d = 2$, to maksymalna liczba serii wynosi $s_{max} = 5$.

Przykład C. Rozważmy również przypadek, w którym $n_d = 2$ i $n_p = 4$ (rysunek 6):



Rysunek 6.

Możliwe drogi i liczby serii podajemy w tabeli:

droga	**liczba serii**	**droga**	**liczba serii**
ppppdd	2	*pdpdpp*	5
pppdpd	4	*pddppp*	3
pppddp	3	*dppppd*	3
ppdppd	4	*dpppdp*	4
ppdpdp	5	*dppdpp*	4
ppddpp	3	*dpdppp*	4
pdpppd	4	*ddpppp*	2
pdppdp	5	-	-

Tabela 6.

Gdy $n_p = 4$ i $n_d = 2$, to maksymalna liczba serii wynosi $s_{max} = 5$.

Zauważyliśmy, że gdy $n_p = n_d = 2$, to maksymalna liczba serii s_{max} równa się 4, a w przypadku gdy n_p równa się 3, 4, 5 lub więcej, maksymalna liczba serii jest równa 5.

Wniosek. Jeśli $n_p = n_d$, to maksymalna liczba serii jest równa $n_p + n_d$. Jeśli liczby n_p i n_d są różne, to maksymalna liczba serii jest równa dwukrotności mniejszej z liczb n_p i n_d powiększonej o jeden:

$$s_{max}(n_p, n_d) = \begin{Bmatrix} n_p + n_d, \ gdy \ n_p = n_d \\ 2\min\{n_p, n_d\} + 1, \ gdy \ n_p \neq n_d \end{Bmatrix}.$$

Przyjrzyjmy się dokładniej sytuacji, gdy $n_p = n_d$. Wiemy, że s_{max} jest liczbą parzystą i równa się $n_p + n_d$.

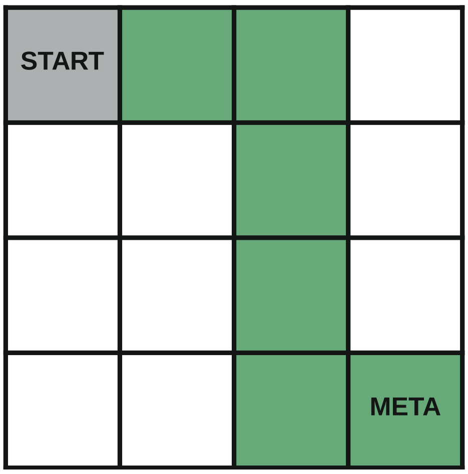
Spójrzmy na przypadek gdy $n_p = 3$ i $n_d = 3$. Wówczas liczba wszystkich dróg wynosi:

$$D = \frac{6!}{3!\,3!} = 20.$$

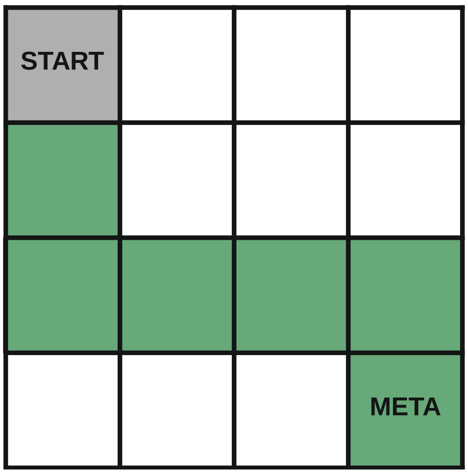
droga	**liczba serii**	**droga**	**liczba serii**
pppddd	2	*dddppp*	2
ppdddp	3	*ddpppd*	3
ppdpdd	4	*ddpdpp*	4
ppddpd	4	*ddppdp*	4
pdddpp	3	*dpppdd*	3
pddppd	4	*dppddp*	4
pddpdp	5	*dppdpd*	5
pdppdd	4	*dpddpp*	4
pdpdpd	6	*dpdpdp*	6
pdpddp	5	*dpdppd*	5

Tabela 7. *Możliwe drogi i liczby serii każdej z nich, gdy* $n_p = n_d = 3$.

Zauważmy, że drogi te są w stosunku do siebie symetryczne w tym sensie, że np. drodze *ppdddp* w kolumnie lewej odpowiada droga *ddpppd* w kolumnie prawej. Sytuacje tę przedstawiamy na rysunku 7 i 8.



Rysunek 7. *Droga ppdddp.*



Rysunek 8. *Droga ddpppd.*

Możemy również dostrzec zależność między liczbą dróg, w których zawarte są dwie, trzy, cztery, pięć lub sześć serii:

$$D_2 = 2$$
$$D_3 = 4$$
$$D_4 = 8$$
$$D_5 = 4$$
$$D_6 = 2.$$

Widzimy, że mamy dokładnie tyle samo dróg, w których p i d układają się w dwie serie ($D_2 = 2$) i w sześć serii ($D_6 = 2$) oraz mamy dokładnie tyle samo dróg, które zawierają trzy serie i pięć serii.

Podobne zależności możemy zaobserwować również w przykładzie rozważanym wcześniej, w którym $n_d = 2$ i $n_p = 2$.

Rozkład liczby serii

Zastanówmy się, jak policzyć prawdopodobieństwo wystąpienia danego ciągu znaków p i d, w zależności od liczby i rodzaju serii, które tworzą i preferencji co do wyboru p lub d (np. chętniej wybieramy p niż d).

Wprowadźmy funkcję $G(s_d, s_p)$, w której s_d oznaczać będzie liczbę serii złożonych ze znaków d, a s_p liczbę serii złożonych ze znaków p. Funkcja ta określa liczbę sposobów, na które możemy otrzymać s_p serii wyrazów p i s_d serii wyrazów d (zob. [1], str. 350-351).

$$G(s_d, s_p) = \begin{cases} 0, \ gdy \ |s_d - s_p| > 1, \\ 1, \ gdy \ |s_d - s_p| = 1, \\ 2, \ gdy \ s_d = s_p. \end{cases}$$

Zauważmy, że funkcja $G(s_d, s_p)$, przyjmuje wartość 0 dla $|s_d - s_p| > 1$, ponieważ liczba s_d serii znaków d nie może się różnić więcej niż o jeden od liczby s_p serii znaków p. W przypadku, gdy $|s_d - s_p| = 1$, czyli gdy liczba s wszystkich serii jest liczbą nieparzystą, funkcja $G(s_d, s_p)$ przyjmuje wartość 1, ponieważ na jeden sposób możemy ułożyć serie znaków p i serie znaków d. Natomiast w przypadku, gdy liczba serii s jest liczbą parzystą (czyli, gdy $s_d = s_p$), to funkcja $G(s_d, s_p)$ przyjmuje wartość 2, ponieważ możemy w dwojaki sposób ułożyć serie znaków p i serie znaków d.

Oznaczmy przez A zdarzenie, że wystąpi dany ciąg serii znaków p i d. W ogólnym przypadku prawdopodobieństwo tego zdarzenia wyraża wzór (zob. [1], str. 351):

$$P(A) = \frac{s_d!}{s_{d1}!\, s_{d2}! \dots s_{dn}!} \cdot \frac{s_p!}{s_{p1}!\, s_{p2}! \dots s_{pn}!} G(s_d, s_p) q^{n_p} (1-q)^{n_d},$$

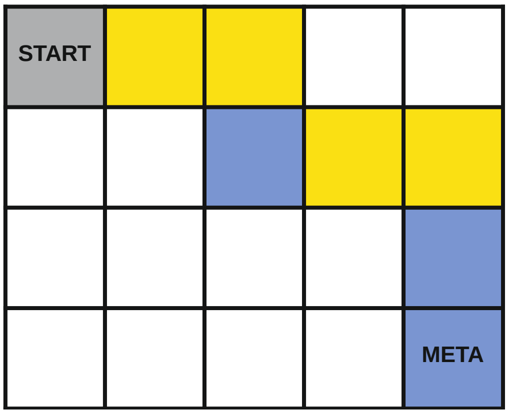
w którym q wyznacza liczbę z przedziału (0,1), która określa prawdopodobieństwo wyboru znaku p (w przypadku prostokątnej planszy jest to prawdopodobieństwo tego, że zrobimy krok w prawo, a nie w dół). W sytuacji, gdy prawdopodobieństwo, że wybierzemy znak p lub d jest takie samo, mamy $q = \frac{1}{2}$.

Przez s_{di} oznaczać będziemy liczbę serii złożonych z i znaków d, zaś przez s_{pj} – liczbę serii złożonych z j znaków p. Wówczas:

$$s_{d1} + s_{d2} + \cdots = s_d$$
$$s_{p1} + s_{p2} + \cdots = s_p$$
$$s_p + s_d = s.$$

Przypomnijmy, że przez n_p określamy liczbę znaków p, a przez n_d liczbę znaków d w danym ciągu.

Rozpatrzmy przykład, w którym weźmiemy tabelę o 4 wierszach i 5 kolumnach (rysunek 8), tzn. aby dojść do mety, musimy zrobić 4 kroki w prawo i 3 kroki w dół, czyli $n_p = 4$, $n_d = 3$. Policzmy ile wynosi prawdopodobieństwo, że przejdziemy drogę *ppdppdd*:



Rysunek 8.

Zauważamy, że w ciągu *ppdppdd* występują cztery serie, tzn. $s = 4$. Policzmy liczbę poszczególnych seriis_{di} oraz s_{pj}.W powyższym przykładzie mamy:

$$s_{d1} = 1 \qquad s_{p1} = 0$$
$$s_{d2} = 1 \qquad s_{p2} = 2$$
$$s_{d3} = 0 \qquad s_{p3} = 0$$
$$s_{d4} = 0 \qquad s_{p4} = 0$$
$$\ldots\ldots\ldots\ldots \qquad \ldots\ldots\ldots\ldots$$

Zauważmy, że liczba serii s jest liczbą parzystą, czyli wartością funkcji $G(s_d, s_p)$ jest 2.

Policzmy prawdopodobieństwo zdarzenia A, że wybierzemy drogę *ppdppdd* i sprawdźmy jak zmienia się ono w zależności od kierunku, który preferujemy.

Gdy równie chętnie wybieramy drogę w prawo i w dół, mamy $q = \frac{1}{2}$. Wówczas szukane prawdopodobieństwo wynosi:

$$P(A) = \frac{2!}{1!\,1!} \cdot \frac{2!}{2!} \cdot 2 \cdot \left(\frac{1}{2}\right)^4 \cdot \left(\frac{1}{2}\right)^3 = \frac{1}{32} = 0{,}03125.$$

Gdy chętniej wybieramy drogę w prawo, na przykład, gdy $q = \frac{2}{3}$, szukane prawdopodobieństwo wynosi:

$$P(A) = \frac{2!}{1!\,1!} \cdot \frac{2!}{2!} \cdot 2 \cdot \left(\frac{2}{3}\right)^4 \cdot \left(\frac{1}{3}\right)^3 = \frac{64}{2187} = 0{,}02926.$$

Gdy chętniej wybieramy drogę w dół, na przykład gdy $q = \frac{1}{3}$, szukane prawdopodobieństwo wynosi:

$$P(A) = \frac{2!}{1!\,1!} \cdot \frac{2!}{2!} \cdot 2 \cdot \left(\frac{1}{3}\right)^4 \cdot \left(\frac{2}{3}\right)^3 = \frac{64}{6561} = 0{,}00975.$$

Obserwacje. Zauważmy, że wybór drogi w opisanych przykładach jest *zdarzeniem losowym*, któremu przyporządkowujemy liczby serii (np. zob. tabela 7). Takie przyporządkowanie, zdarzeniu losowemu pewnej liczby rzeczywistej nazywamy *zmienną losową*. Przez *zmienną losową* będziemy rozumieć funkcję określona na zbiorze zdarzeń losowych o wartościach w zbiorze liczb rzeczywistych.

Funkcję, która określa z jakim prawdopodobieństwem zmienna losowa osiąga daną wartość, nazywamy *rozkładem zmiennej losowej*. Ważną wielkością związaną z analizą zmiennej losowej i jej rozkładu są kwantyle. *Kwantylem* l_p rzędu $100p\%$ (gdzie $0 \leq p \leq 1$) rozkładu zmiennej losowej, nazywamy taką wartość tej zmiennej, że zmienna przyjmuje wartości nie większe niż l_p z prawdopodobieństwem p oraz nie mniejsze niż l_p z prawdopodobieństwem $1 - p$.

Test Walda i Wolfowitza

Podamy teraz zastosowanie rozkładu liczby serii do badania równości rozkładów dwóch wielkości w dwóch populacjach. Jest to *test serii* zaproponowany przez Walda i Wolfowitza. Możemy go zastosować np. do zbadania czy rozkład wynagrodzenia kobiet i mężczyzn w danej firmie nie zależy od płci (zob. [2], str. 114-115). Adaptujemy przykład z książki [2] do stwierdzenia, czy rozkład ocen w pewnej szkole z pewnego przedmiotu, jest taki sam w grupie uczennic, jak i uczniów, tzn. czy nie zależy od płci. Jeśli nie występuje preferencja, tzn. oba rozkłady ocen są identyczne, to powinna występować duża liczba serii, nie mniejsza niż wartość *kwantyla* rozkładu liczby serii, podana w tabeli (zob. tabela 8).

Dana jest alfabetyczna lista uczniów. Każdej osobie z klasy przyporządkowujemy średnią arytmetyczną z ocen. Następnie listę porządkujemy rosnąco (albo malejąco), w zależności od średniej ocen. W przypadku, gdy dwie osoby mają dokładnie taką samą średnią, to porządkujemy je w sposób losowy, opierając przyporządkowanie o rzut monetą. Każdej osobie przyporządkowujemy literę p lub d (podobnie jak w zagadnieniu związanym z rozkładem liczby serii), np. p – każdej uczennicy, a d – każdemu uczniowi. Następnie wyznaczamy liczbę serii.

Jeśli nauczyciel ocenia uczennice wyżej od uczniów (albo uczniów wyżej niż uczennice), to liczba serii będzie mała. Natomiast jeśli rozkład ocen wśród uczniów i uczennic jest jednakowy, to wystąpi duża liczba serii, gdyż dane przemieszają się w trakcie sortowania.

Minimalną, graniczną liczbę, która każe snuć przypuszczenia, że rozkład ocen w grupie uczennic nie jest taki sam jak w grupie uczniów, podajemy w tabeli *kwantyli rozkładu liczby serii* (tabela 8), która została zaczerpnięta z książki [2], str. 297.

Kwantyle l(0,05, p, d) rozkładu liczby serii

	7	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
6	4	4	4	5	5	5	5	5	6	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7
7		4	5	5	5	6	6	6	6	6	7	7	7	7	7	7	8	8	8	8	8	8	8	8
8			5	6	6	6	6	7	7	7	7	8	8	8	8	8	8	8	9	9	9	9	9	9
9				6	6	7	7	7	8	8	8	8	8	9	9	9	9	9	10	10	10	10	10	10
10					7	7	8	8	8	8	9	9	9	9	10	10	10	10	10	10	11	11	11	11
11						8	8	8	9	9	9	10	10	10	10	10	11	11	11	11	11	12	12	12
12							9	9	9	10	10	10	10	11	11	11	11	12	12	12	12	12	13	13
13								9	10	10	10	11	11	11	12	12	12	12	13	13	13	13	13	14
14									10	11	11	11	12	12	12	12	13	13	13	14	14	14	14	14
15										11	11	12	12	12	13	13	13	14	14	14	14	15	15	15
16											12	12	13	13	13	14	14	14	15	15	15	15	16	16
17												13	13	13	14	14	15	15	15	15	16	16	16	17
18													14	14	14	15	15	15	16	16	16	17	17	17
19														14	15	15	16	16	16	17	17	17	18	18
20															15	16	16	16	17	17	18	18	18	18
21																16	17	17	17	18	18	18	19	19
22																	17	17	18	18	19	19	19	20
23																		18	18	19	19	20	20	20
24																			19	19	20	20	20	21
25																				20	20	21	21	21
26																					21	21	21	22
27																						21	22	22
28																							22	23
29																								23

Tabela 8. *Tabela kwantyli rozkładu liczby serii*

Powyższa tabela została zaczerpnięta z książki [2], str. 297.

Numer kolumny powyższej tabeli odpowiada liczbie symboli p, natomiast numer wiersza odpowiada liczbie symboli d.

płeć	średnia arytmetyczna z ocen
p	2,00
d	2,28
d	2,50
p	2,74
p	3,28
p	3,32
p	3,32
p	3,56
d	3,60
p	3,64
d	3,68
d	3,68
p	3,80
d	3,83
d	4,00
d	4,20
p	4,28
d	4,30
p	4,47
p	4,67
d	4,75
p	4,90
p	4,93
p	4,93
d	5,00
d	5,20

Tabela 9.

Spójrzmy na przykładową tabelę (zob. tabela 9), w której zostały przedstawione średnie arytmetyczne ocen uczniów pewnej klasy z pewnego przedmiotu i wartości te zostały uporządkowane według opisanej powyżej zasady. Do klasy tej uczęszcza 14 dziewcząt i 12 chłopców, z tabeli kwantyli rozkładu liczby serii odczytujemy, że minimalna liczba serii powinna wynieść 9, abyśmy mogli wnioskować, że rozkład ocen nie zależy od płci. W opisanym przykładzie liczba serii wynosi 14. Zatem możemy stwierdzić, że nauczyciel najprawdopodobniej jednakowo ocenia uczniów i uczennice.

Literatura

[1] Marek Fisz, *Rachunek prawdopodobieństwa i statystyka matematyczna*, Państwowe Wydawnictwo Naukowe, Warszawa 1958

[2] Lesław Gajek, Marek Kałuszka, *Wnioskowanie statystyczne*, Wydawnictwa Naukowo-Techniczne, Warszawa 1999