Lemat o głosowaniu i test rangowy

Rafał SZTENCEL

Przypuśćmy, że n roślin poddanych doświadczeniu urosło do wysokości a_1, a_2, …, a_n, podczas gdy egzemplarze kontrolne uzyskały wysokości b_1, b_2, …, b_n. Jeśli wypiszemy otrzymane liczby w porządku malejącym i wszystkie a pojawią się przed wszystkimi b, to uznamy, że doświadczenie silnie wpłynęło na wzrost roślin. Metoda nieskuteczna powinna natomiast dać uporządkowanie losowe.

Zagadnieniem prowadzenia przy obliczaniu głosów zajmowaliśmy się w poprzednim odcinku.

Uporządkowanie liter a i b można zinterpretować jako wynik głosowania, w którym obaj kandydaci uzyskali tyle samo głosów. W trakcie obliczania głosów jeden z kandydatów mógł prowadzić $2k$ razy, $0 \leqslant k \leqslant n$, co zajdzie wtedy i tylko wtedy, gdy nierówność $a_i > b_i$ ma miejsce dla k wskaźników i (należy założyć, że wszystkie a_i oraz b_i są różne).

Szansa takiego zdarzenia jest równa $1/(n+1)$; szansa na $2k$ lub więcej prowadzeń wynosi $(n-k+1)/(n+1)$. W roku 1876 Galton [2] przeprowadził przedstawione wyżej rozumowanie nie znając wartości prawdopodobieństw. Dane doświadczalne pochodziły od Darwina, a prowadziły 26 razy przy $n = 15$. Galton wnioskował, że metoda była skuteczna. My możemy dodać, że nawet nieskuteczna metoda dałaby 26 lub więcej prowadzeń z prawdopodobieństwem $3/16$. Obecnie podobne testy rangowe stosuje się powszechnie, bowiem potrafimy obliczyć szansę przypadkowego uznania metody nieskutecznej za skuteczną.

Udowodnimy teraz to, co zapowiadaliśmy w poprzednim odcinku i czego nie wiedział Galton: jeśli głosowanie zakończyło się remisem, to szansa, że kandydat A prowadził $2k$ razy, nie zależy od k i jest równa $1/(n+1)$. Przyjmijmy oznaczenie

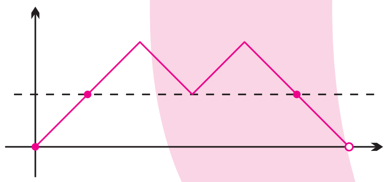
$$L_{2n} = \frac{1}{n+1}\binom{2n}{n}.$$

Wyniki liczenia głosów, jak pamiętamy, rejestrujemy w postaci dróg wychodzących z punktu $(0,0)$.

Twierdzenie 1. Jeśli głosowanie zakończyło się remisem $n:n$, to wśród $\binom{2n}{n}$ możliwych wyników liczenia głosów jest (a) L_{2n-2} takich, że kandydat A miał w chwilach 1, 2, …, $n-1$ przewagę; (b) L_{2n} takich, że kandydat A prowadził $2n$ razy.

Dowód. Punkt (a) wynika z lematu o głosowaniu z poprzedniego odcinka: każda droga spełniająca (a) przechodzi przez punkt $(2n-1,1)$, więc liczba takich dróg jest równa

$$\frac{1}{2n-1}\binom{2n-1}{n-1} = \frac{1}{n}\binom{2n-2}{n-1} = L_{2n-2}.$$



Punkt (b) otrzymujemy zauważając, że droga z (a) po odrzuceniu pierwszego i ostatniego odcinka prowadzi z $(1,1)$ do $(2n-1,1)$, nie schodząc poniżej prostej $y = 1$ (p. rys.). Przesunięcie całości o wektor $(-1,-1)$ pokazuje, że istnieje wzajemnie jednoznaczne przyporządkowanie między drogami nieujemnymi z $(0,0)$ do $(2n-2,0)$ a drogami spełniającymi (a).

Twierdzenie 2. Niech $L_{2k,2n}$ oznacza liczbę dróg z $(0,0)$ do $(2n,0)$ takich, że $2k$ odcinków leży ponad osią OX, a $2n-2k$ odcinków leży pod osią OX, $k = 0,1,2,\ldots,n$. Wtedy $L_{2k,2n} = L_{2n}$ niezależnie od k.

Dowód. Teza jest oczywista dla $n = 1$. Dla indukcji założymy, że $L_{2k,2m} = L_{2m}$, $1 \leqslant m \leqslant n-1$, $0 \leqslant k \leqslant m$.

Przypuśćmy, że droga ma $2k$ odcinków powyżej osi OX, $0 \leqslant k \leqslant n-1$. Jej pierwsza „wycieczka” z zera mogła być dodatnia lub ujemna i zakończyła się w chwili $2r$.

W pierwszym przypadku wycieczka mogła zostać odbyta na L_{2r-2} sposobów, a dalsza część drogi na $L_{2k-2r,2n-2r} = L_{2n-2r}$ sposobów. W drugim przypadku pierwsza wycieczka jest ujemna, ale liczbę dróg obliczamy analogicznie, co po zsumowaniu względem dopuszczalnych r daje:

$$L_{2k,2n} = \sum_{r=1}^{k} L_{2r-2}L_{2n-2r} + \sum_{r=1}^{n-k} L_{2r-2}L_{2n-2r}.$$

Zamiana wskaźnika sumowania w drugiej sumie na $s = n-r+1$ pozwala zobaczyć, że

$$L_{2k,2n} = \sum_{r=1}^{n} L_{2r-2}L_{2n-2r},$$

a to nie zależy od k.

Symetria i tw. 1 dają $L_{2n,2n} = L_{0,2n} = L_{2n}$, a ponieważ wszystkich dróg z $(0,0)$ do $(2n,0)$ jest $(n+1)L_{2n}$, więc $L_{2k,2n} = L_{2n}$ dla $0 \leqslant k \leqslant n$.

Bibliografia

[1] W. Feller, *Wstęp do rachunku prawdopodobieństwa*, t. I, wyd. II, PWN, Warszawa 1966

[2] J.L. Hodges, Jr., *Golton's rank-order test*, Biometrika 42 (1955), ss. 261–262.