Żeby życie miało smaczek, raz dziewczynka, raz chłopaczek

Rafał SZTENCEL

Chłopcy rodzą się nieco częściej niż dziewczynki. Pierwsze dane na ten temat pochodzą od Johna Graunta (1662), który stwierdził, że częstość urodzeń chłopców to $f_m = \frac{14}{27} = 0{,}5185$. Laplace na podstawie danych z Francji (1800–1802) podaje wartość $f_m = 0{,}5166$. Choć Laplace badał wpływ klimatu na wartość f_m, musiał być przekonany o jej stabilności, gdy bowiem okazało się, że w londyńskich i paryskich przytułkach wśród podrzutków jest podejrzanie mało chłopców, poszukał najpierw zdroworozsądkowego wyjaśnienia. Okazało się, że chłopi z okolicznych wsi częściej podrzucają dziewczynki.

Częstość narodzin chłopców jest zadziwiająco stabilna, co potwierdzają kolejne wartości [3, 4]: 0,5173 (Polska, 1957–1967), 0,5146 (Polska, 1999–2005), 0,5175 (Szwecja, 1935).

Prezydenci USA [1, 5, 6] mieli 87 synów i 63 córki[1], co daje $f_m = 0{,}58$. Jak rozstrzygnąć, czy to odchylenie od spodziewanej wartości można przypisać niedużej próbce, czy przyczynom natury biologicznej? Biolog mógłby argumentować, że męscy potomkowie zapewniają sukces reprodukcyjny[2], a tego w końcu spodziewamy się po ludzkim odpowiedniku samca alfa (inaczej: dominanta).

Powszechnie przyjmowanym (nie wiadomo, czy słusznie) modelem probabilistycznym procesu urodzin jest schemat Bernoulliego n niezależnych doświadczeń. W każdym z nich szansa sukcesu jest równa p. Średnio otrzymamy np sukcesów, a znając wariancję liczby sukcesów ($np(1-p)$) ocenimy, czy obserwowane odchylenie liczby sukcesów S_n od średniej jest małe, czy duże. Do tego celu służy pierwiastek z wariancji, zwany odchyleniem standardowym.

Twierdzenie de Moivre'a–Laplace'a mówi, że przy $n \to \infty$ zmienna losowa $\frac{S_n - np}{\sqrt{np(1-p)}}$ ma asymptotycznie standardowy rozkład normalny $N(0,1)$. Wynika stąd, że dla dużych n

$$P\left(-t < \frac{S_n - np}{\sqrt{np(1-p)}} < t\right) =$$

$$= P\left(np - t\sqrt{np(1-p)} < S_n < np + t\sqrt{np(1-p)}\right) \approx$$

$$\approx \Phi(t) - \Phi(-t),$$

gdzie Φ jest (stablicowaną) dystrybuantą rozkładu $N(0,1)$. Jeśli np. $n = 360000$ i $p = \frac{1}{2}$, to $\sqrt{np(1-p)} = 300$, i $|S_n - np|$ nie przekroczy 300 z prawdopodobieństwem $\Phi(1) - \Phi(-1) \approx 0{,}68$. To zaledwie 0,08% liczby doświadczeń, więc zmienna losowa S_n niezwykle silnie koncentruje się wokół wartości średniej. W tablicach rozkładu normalnego można sprawdzić, że przy trzech odchyleniach standardowych szansa na jeszcze większe odchylenie spada do 0,003.

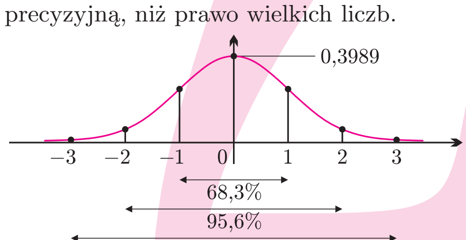
W latach 1999–2005 w Polsce przyszło na świat 2 553 906 dzieci, w tym 1 314 333 chłopców, co daje $f_m = 0{,}514636$. W tabeli widzimy odchylenia częstości od tej liczby w poszczególnych latach. W kolumnie oznaczonej literą σ przeliczono je na odchylenia standardowe (dokładniej: przyjęto, że $p = f_m$, co jest usprawiedliwione wielkością próby). Jak widać, odchylenia nie są zbyt duże. Łatwo uwierzyć, że spowodowane są działaniem czystego przypadku.

Rok	m	k	$m+k$	f_m	σ	f_m po k latach
1999	196411	185591	382002	0,514162	−0,5864	0,514162
2000	194824	183524	378348	0,514933	0,3654	0,514546
2001	189521	178684	368205	0,514716	0,0966	0,514601
2002	182136	171629	353765	0,514850	0,2545	0,514661
2003	180634	170438	351072	0,514521	−0,1365	0,514634
2004	183422	172709	356131	0,515041	0,4829	0,514700
2005	187385	176998	364383	0,514253	−0,4632	0,514636
Razem	1314333	1239573	2553906	0,514636		

Ale w latach 1957–1967 w Polsce urodziło się ponad 6 milionów dzieci. Przyjmijmy, że szansa urodzenia chłopca jest równa 0,5173 (czyli częstości) i zobaczmy, czy dane z lat 1999–2005 są zgodne z tym założeniem. Powinno było przyjść na świat 1 321 220 chłopców, czyli prawie o 7000 więcej. Mamy odchylenie standardowe około $\sqrt{2553906/4} \approx 799$. Zatem z pozoru drobna różnica częstości, wynosząca 0,27%, przekłada się na ponad 8 odchyleń standardowych! To nie może być przypadek (szansa przypadkowego otrzymania takiego lub większego odchylenia jest mniejsza niż 10^{-17}) i należałoby poszukać wyjaśnienia.

Podobnie, przewaga chłopców wśród dzieci prezydentów USA jest być może nieprzypadkowa. Szansa na większą przewagę, oszacowana jak wyżej, jest równa $1 - \Phi(1{,}54) \approx 0{,}064$, a obliczona dokładnie (za pomocą arkusza kalkulacyjnego) jest równa 5,17%. Statystyk, przyzwyczajony do poziomu istotności 5%, miałby problem z decyzją. Musiałby mieć większą próbkę. O to nietrudno: jak twierdzi Robin Baker [1], w wyższych warstwach społecznych wśród dzieci istotnie przeważają chłopcy. Analiza socjobiologiczna tego zjawiska wykracza poza ramy niniejszego artykułu.

Spójrzmy jeszcze na ostatnią kolumnę tabeli. Są tam częstości f_m po k latach, $k = 1, 2, \ldots, 7$. Można zauważyć tendencję do stabilizacji. Jest to przejaw działania prawa wielkich liczb: zmienne losowe $\frac{S_n - np}{n} = \frac{S_n}{n} - p$ zmierzają do zera z prawdopodobieństwem 1. Innymi słowy, częstość sukcesu zmierza do jego prawdopodobieństwa. Statystyk powie, że dla dużych n jest ona dobrym estymatorem p (z czego korzystaliśmy). Zamiana n w mianowniku pierwszego ułamka na wielkość rzędu $\sqrt{n}$ powoduje, że rozkład graniczny już nie jest jednopunktowy – jest rozkładem normalnym, co daje informację o wiele bardziej precyzyjną, niż prawo wielkich liczb.



$0{,}68 \approx \Phi(1) - \Phi(-1) = \int_{-1}^{1} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$

[1] Mowa o dzieciach urodzonych w małżeństwie. Odpowiednie komisje śledcze zbadały DNA Tomasza Jeffersona i jego niewolnicy, Sally Hemings, z którą przypuszczalnie miał 7 dzieci [5, 6].

[2] Rekordzistami w tej dziedzinie są – może ze względów organizacyjnych – władcy muzułmańscy. Mulaj Ismail Krwiopijca z Maroka miał 888 dzieci [1, 2]; August II Sas miał według paszkwilantów 356 nieślubnych dzieci. Rekord w kategorii kobiet wygląda skromnie. Pewna Rosjanka urodziła 69 dzieci.

Bibliografia

[1] Robin Baker, *Wojny plemników*, Rebis, Poznań 2000

[2] Jared Diamond, *Trzeci szympans*, PIW, Warszawa 1996

[3] Polska Statystyka Publiczna, http://www.stat.gov.pl/demografia/index.html

[4] Wiesław Szlenk, *Rachunek prawdopodobieństwa*, PZWS, Warszawa 1970

[5] Wikipedia: *List of children of the Presidents of the United States*, http://en.wikipedia.org/wiki

[6] Wives and Children of the Presidents, http://www.infoplease.com/ipa/A0194051.html