



Ilu mamy przodków?

Piotr RÓŻAŃSKI*

Rozwiązanie zadania M 1310.

Załóżmy, że liczbę n zapisano jako $(k+1) + (k+2) + \dots + l$ dla pewnych liczb naturalnych k i l , takich że $l - k \geq 2$. To oznacza, że

$$n = \frac{l(l+1)}{2} - \frac{k(k+1)}{2} = \frac{(l-k)(l+k+1)}{2}.$$

Zauważmy, iż jedna z liczb $l-k$, $l+k+1$ jest nieparzysta i większa od 1, czyli n ma czynnik nieparzysty, więc nie jest potęgą dwójki. Z drugiej strony, jeżeli n nie jest potęgą dwójki, to liczba $2n$ zapisuje się jako iloczyn liczby parzystej i nieparzystej pq . Jeśli p oznacza większą z tych liczb, to możemy przyjąć, iż

$$k = \frac{p-q-1}{2}, \quad l = \frac{p+q-1}{2}.$$

Jak wiadomo, każdy człowiek ma dwoje rodziców. Skoro każdy z rodziców też jest człowiekiem, ta rekurencyjna zależność pozwala w prosty sposób wyznaczyć liczbę przodków dowolnej osoby w linii prostej w kolejnych pokoleniach: czworo babć i dziadków, ośmioro prababć i pradziadków, a zatem ogólnie $(\text{pra})^n$ dziadków obojga płci mamy 2^{n+2} .

Na nieszczęście, powyższa zależność w dłuższym przedziale czasu prowadzi do paradoksu: jak bowiem możemy mieć ponad miliard przodków w trzydziestym pokoleniu, jeśli przewidywana liczba ludności świata w tym okresie (około roku 1300 n.e.) nie przekraczała 500 milionów? Aby wyjaśnić ten paradoks, wystarczy zauważyć, że w każdej ilościowo ograniczonej populacji pary zawsze tworzone są pomiędzy osobami spokrewnionymi, choćby w minimalnym stopniu. W związku z tym każdy ze wspólnych przodków tej pary będzie liczony dwukrotnie w drzewie genealogicznym każdego z jej potomków. Przy rozpatrywaniu wielu pokoleń (a więc wielkiej liczby przodków) zjawisko to zdarza się bardzo często.

Niestety, wyjaśnienie to nie zbliża nas do ilościowej odpowiedzi na zadane w tytule pytanie. Aby takiej odpowiedzi udzielić, należy dokonać pewnych założeń dotyczących przyjętego modelu populacji i dziedziczenia.

Założenia modelu

Pierwszym uproszczeniem, którego dokonamy, modelując populację, będzie podzielenie jej na pokolenia. Zakładamy, że pary czy też związki, mogą być zawierane tylko przez osoby należące do tego samego pokolenia. Dzięki temu ciągłe zagadnienie modelowania liczby ludności sprowadza się do zagadnienia dyskretnego. Przyjmujemy ponadto, że tworzone pary są ściśle monogamiczne oraz niezmiennie. W rozpatrywanym modelu zakładamy, że pary tworzone są w sposób losowy wewnątrz pokolenia, pomijając natomiast będziemy osoby niewchodzące w skład żadnej pary.

Dzięki przyjętym założeniom możemy za podstawową jednostkę modelu przyjąć nie pojedynczą osobę, lecz parę złożoną z dwóch osób przeciwnej płci. W tym ujęciu każda jednostka (para) pochodzi od dwóch jednostek (dwóch par rodziców) z poprzedniego pokolenia. Dzięki takiemu przedstawieniu sytuacji całkowicie unikamy problemów związanych z kategorią płci jako takiej. Od tego momentu (do odwołania) określenia „potomek” oraz „rodzic” będą dotyczyć jednostek, czyli *de facto* par.

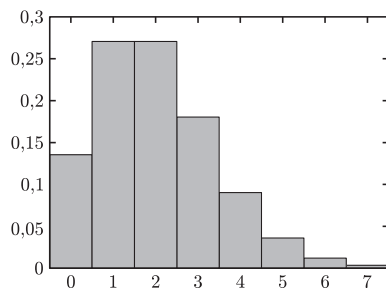
„Magiczne” założenie

Dotychczasowe założenia modelu wynikały w naturalny sposób z określonych wniosków dotyczących realnego świata. Dla odmiany, ostatnie przyjęte w modelu założenie będzie jego cechą charakterystyczną. Nazywamy je „magicznym”, ponieważ mimo pozornie trywialnej treści ma znaczące konsekwencje. Oto ono:

Dopasowania rodziców dla poszczególnych dzieci są od siebie niezależne.

Jak to rozumieć? Wyobraźmy sobie, że w pokoleniu dzieci wskazujemy dwie jednostki i oznaczamy je jako A i B. W tym momencie rodziców A może stanowić z równym prawdopodobieństwem każda nieuporządkowana dwójka jednostek z pokolenia rodziców. Istota założenia polega na tym, że niezależnie od wskazania rodziców dla jednostki A wybór rodziców dla jednostki B nadal jest losowy z równym prawdopodobieństwem dla każdej dwójki.

Okazuje się, że przy takich założeniach jedynymi parametrami modelu są liczebności poszczególnych pokoleń. W szczególności, znając liczebność pokolenia rodziców (R) i liczebność pokolenia dzieci (D), można wyznaczyć strukturę dzietności, czyli rozkład zmiennej losowej T opisującej liczbę potomków wybranej jednostki z pokolenia rodziców. Otrzymany wynik ma postać rozkładu dwumianowego z ilością prób równą D i parametrem $p = \frac{2}{R}$. Biorąc pod uwagę, że liczebności populacji są zwykle dość duże, dobrym przybliżeniem tego rozkładu staje się rozkład Poissona z parametrem $\lambda = \frac{2D}{R}$. Rozkład w szczególnym przypadku populacji o stałej liczebności (dla $\lambda = 2$) przedstawiony jest na wykresie.



Rys. 1. Rozkład dzietności (T) przy stałej liczebności populacji ($\lambda = 2$).

*

Funkcja przejścia

Założenia modelu mówią, że każda jednostka pochodzi od dwóch rodziców. Zastanówmy się w takim razie, od ilu rodziców pochodzą dwie jednostki? Naturalna wydaje się odpowiedź: „od czterech”, ale przecież zachodzi to tylko wtedy, gdy jednostki te nie mają wspólnych rodziców.

Rozważmy bardziej ogólną sytuację, w której spośród pokolenia dzieci wybieramy d jednostek i badamy wartość r – moc zbioru rodziców wybranych dzieci. Rozwiązaniem tego zagadnienia dla zadanego d jest zmienna losowa o rozkładzie $p_d(r)$, przyjmującym niezerowe wartości jedynie dla $2 \leq r \leq \min(2d, R)$.

Rozkłady $p_d(r)$, wyznaczone kombinatorycznie i zapisane w postaci rekurencyjnej, nie wyglądają zachęcająco:

$$p_d(r) = p_{d-1}(r-2) \cdot \frac{(R-r+1)(R-r+2)}{R(R-1)} + \\ + p_{d-1}(r-1) \cdot \frac{2(R-r+1)(r-1)}{R(R-1)} + p_{d-1}(r) \cdot \frac{r(r-1)}{R(R-1)}.$$

Jak się jednak okazuje po prostym badaniu metodami numerycznymi, każdy z nich przyjmuje znacząco niezerowe wartości tylko w niewielkim przedziale otaczającym maksimum. Na dodatek, szerokość tego przedziału zmniejsza się szybko wraz ze wzrostem liczebności populacji. Wobec tego przybliżymy poszczególne rozkłady $p_d(r)$ rozkładami jednopunktowymi zlokalizowanymi w maksimach – położenia owych maksimów oznaczmy przez r_0 . Będziemy zatem poszukiwać wartości r_0 w zależności od parametru d rozkładu. Poszukiwana funkcja $r_0(d)$ ma tutaj bardzo sensowną interpretację: jest to oczekiwana ilość rodziców dla d jednostek z pokolenia dzieci.

Wróćmy na chwilę do zmiennej losowej \mathbf{T} przedstawiającej strukturę dietności populacji. Zauważmy najpierw, że bezpośrednio z rozkładu tej zmiennej możemy odczytać wartość $r_0(D)$, czyli przewidywaną liczbę rodziców całego pokolenia. Wartość ta będzie równa $R \cdot P(\mathbf{T} > 0) = R(1 - e^{-\frac{2D}{R}})$.

Następnie zauważmy, że jeśli z pokolenia dzieci wyodrębnimy d jednostek, to (na podstawie „magicznego” założenia) jakiegokolwiek przyporządkowania dokonane dla nich będą niezależne od przyporządkowań dla pozostałych jednostek. Zatem, wynik dla d jednostek wybranych z pokolenia o liczebności D , czyli $r_0(d)$, powinien być taki sam, jak wynik dla całego pokolenia o liczebności d . Na mocy tej obserwacji otrzymujemy

$$r_0(d) = R(1 - e^{-\frac{2d}{R}}).$$

Można sprawdzić, że dla odpowiednio licznych populacji (R dużo większe od d) jest $r_0(d) \approx 2d$.

Zliczanie

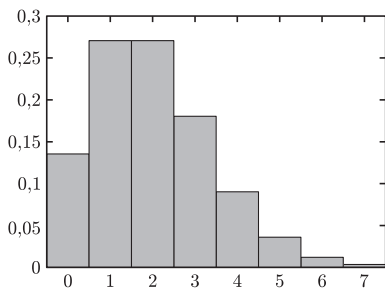
Aby policzyć przodków z poszczególnych pokoleń, musimy przyjąć określony model liczebności. Powracając do liczenia osób, a nie par, określić należy ciąg N_1, N_2, N_3, \dots wyrażający liczebności poszczególnych pokoleń (liczonych wstecz). Analogicznie, przez P_1, P_2, P_3, \dots oznaczmy liczby przodków kolejnych stopni (rodziców, dziadków, pradziadków...). Zauważmy, że dla dowolnego $n > 1$, przy rozpatrywaniu związku pomiędzy pokoleniami $n-1$ i n , liczba par z pokolenia dzieci równa jest $\frac{1}{2}P_{n-1}$, liczba zaś par rodziców $\frac{1}{2}P_n$. Jednocześnie, ilość wszystkich par z pokolenia rodziców, czyli R , wynosi $\frac{1}{2}N_n$. Zatem, korzystając z obliczonej poprzednio postaci funkcji $r_0(d)$ mamy

$$\frac{1}{2}P_n = r_0\left(\frac{1}{2}P_{n-1}\right)$$

dla dowolnego $n > 1$.

Dodatkowo, biorąc pod uwagę, że mamy dwoje rodziców ($P_1 = 2$), oraz podstawiając do powyższego wzoru jawną postać funkcji $r_0(d)$, możemy napisać następującą rekurencję:

$$P_1 = 2, \quad P_n = N_n \cdot \left[1 - \exp\left(-\frac{2P_{n-1}}{N_n}\right)\right].$$



Rys. 2. Wyniki dla modelu jednoparametrowego z $N_1 = 100000$ oraz $\alpha = 1,03$

Jednoparametrowy model liczebności

Przyjmując prosty model liczebności scharakteryzowany jednym parametrem $\alpha = N_n/N_{n+1} = \text{const}$, można jakościowo scharakteryzować wyniki otrzymane przez numeryczne rozwiązanie powyższej rekurencji. Wyróżniamy trzy fazy:

- faza wzrostu wykładniczego
dla najbliższych pokoleń, w której $P_n \approx 2P_{n-1}$;
- faza przejściowa,
w której $P_n < 2P_{n-1}$, występująca w okolicy $P_n \approx \frac{1}{2}N_n$;
- faza nasycenia
charakteryzująca się stałym ilorazem (współczynnikiem nasycenia) $\tau = \frac{P_n}{N_n}$.
Współczynnik nasycenia w powyższym modelu można obliczyć, szukając granicy ciągu (P_n/N_n) , a więc rozwiązując równanie $\tau = 1 - e^{-2\alpha\tau}$.

Przy populacji o stałej liczebności ($\alpha = 1$) współczynnik ten osiąga wartość $\tau \approx 0,797$. Oznacza to, że dla takiej populacji, w odległych pokoleniach, przodkowie stanowią niezmiennie prawie 80% całego pokolenia.

Co poza tym?

Nie da się nie zauważyć, że przedstawiony tu model obliczeń jest znacząco uproszczony. Dla rzeczywistych, dużych populacji, dobór osób w pary nie odbywa się całkowicie losowo, lecz zazwyczaj istnieje skłonność do szukania partnerów wśród bliższego otoczenia. Uzasadnione więc byłoby wyodrębnienie grup wewnątrz populacji, dopuszczając jednocześnie (z określonym prawdopodobieństwem) możliwość krzyżowania pomiędzy grupami. Ponadto, podział na pokolenia nie jest do końca naturalny, szczególnie w populacjach, w których nie obowiązują ściśle reguły (np. kulturowe) dotyczące różnic wieku.

Bardziej złożony model, uwzględniający rzeczywiste migracje ludności i izolacje poszczególnych grup (geograficzne, językowe...) wykorzystali Rohde, Olson i Chang w swoim artykule *Modelling the recent common ancestry of all living humans* opublikowanym w Nature. Oszacowali oni czas pojawienia się tzw. Ostatniego Wspólnego Przodka wszystkich obecnie żyjących ludzi na Ziemi na pierwsze lub drugie tysiąclecie p.n.e. Genealodzy, do pracy!



Zadania

Redaguje Ewa CZUCHRY

F 785. W naczyniu z wodą zanurzamy pionowo drewniany klocek o przekroju kwadratowym $a \times a$, ($a = 10$ cm) i długości $l = 20$ cm. Jaka ilość ciepła wydzieli się w wyniku przewrócenia się tego klocka? Przyjmujemy, że gęstość drewna jest równa połowie gęstości wody.

Rozwiązanie na str. 15

F 786. Jak należy napełnić naczynie wodą, by środek ciężkości leżał możliwie najniżej?

Rozwiązanie na str. 7

Redaguje Przemysław MAZUR

M 1309. Na bokach AB i AC trójkąta ABC zbudowano po zewnętrznej stronie podobne trójkąty prostokątne ADB i AEC , w których kąty przy wierzchołkach D i E są proste. Punkt M jest środkiem odcinka BC . Udowodnić, że $DM = EM$.
Rozwiązanie na str. 24

M 1310. Znaleźć wszystkie liczby naturalne, których nie da się zapisać jako sumy co najmniej dwóch kolejnych liczb całkowitych dodatnich.

Rozwiązanie na str. 9

M 1311. Dane jest słowo złożone z liter a, b (np. $baabab$). Na takim słowie możemy wykonać następujące operacje:

- dopisać lub wykreślić dwie kolejne litery a (np. $baabab \leftrightarrow bbab$),
- dopisać lub wykreślić trzy kolejne litery b (np. $bbab \leftrightarrow bbabbbb$),
- zamienić sekwencję liter ab na bba lub na odwrot (np. $bbabbbb \leftrightarrow abbbb$).

Rozstrzygnąć, czy za pomocą wielokrotnego wykonywania tych operacji można ze słowa a otrzymać słowo ba .

Rozwiązanie na str. 24