

Rachunek Prawdopodobieństwa MAP1181

Wydział Matematyki, Matematyka Stosowana

Projekt - zastosowania rachunku prawdopodobieństwa w genetyce

Opracowanie: Antonina Urbaniak

1 Podstawowe pojęcia z zakresu genetyki

Niekwestionowanym prekursorem genetyki był Grzegorz Mendel (1822-1884) - niemiecko-czeski zakonnik i przyrodnik. Sformułował on podstawowe prawa dziedziczenia, przeprowadzając badania nad krzyżowaniem roślin, głównie grochu zwyczajnego.

Teoria dziedziczości jest przykładem na zastosowanie prostych modeli probabilistycznych. Cechy dziedziczne zależą od specjalnych nośników, zwanych genami. Wszystkie komórki ciała (poza komórkami rozrodczymi) rodzą kopie o tej samej strukturze genów, a co najważniejsze, geny zawsze występują parami.

W najprostszym przypadku każdy gen pojedynczej pary może przybierać dwie formy (allele): A oraz a , gdzie A oznacza cechę dominującą, a a recesywną. Mamy zatem trzy możliwości: AA , Aa oraz aa i jedynie w ostatnim przypadku osobniki wykazują obserwowalny wpływ genów a .

Pierwsze prawo Mendla (prawo czystości gamet) mówi nam, że każda gameta wytwarzana przez organizm posiada tylko jeden allel z danej pary alleli genu. Wynika z tego, że każda komórka płciowa musi zawierać po jednym allelu z każdej pary. Z kolei drugie prawo Mendla (prawo niezależnej segregacji cech) mówi o tym, że geny należące do jednej pary alleli są dziedziczone niezależnie od genów należących do drugiej pary alleli.

2 Podstawowy model dziedziczenia

Załóżmy, że trzy genotypy AA , Aa oraz aa występują wśród osobników płci męskiej i żeńskiej w tym samym stosunku $u : 2v : w$. Zakładamy, że $u + 2v + w = 1$, a u , v oraz w to częstości występowania genotypów. Przyjmijmy również

$$p = u + v, \quad q = v + w.$$

Widzimy, że liczby genów A oraz a mają się do siebie jak p do q , a ponieważ $p + q = 1$, to p i q to częstości występowania genów A oraz a .

Zatem, jeśli gen A jest wybierany z prawdopodobieństwem p (zakładamy niezależność na podstawie drugiego prawa Mendla), to prawdopodobieństwo, że potomek będzie typu AA , wynosi p^2 . Genotyp Aa można wybrać na dwa sposoby, więc jego prawdopodobieństwo wynosi $2pq$.

Tak więc, w warunkach kojarzenia się losowego, potomek w pierwszym pokoleniu należy do genotypu AA , Aa oraz aa odpowiednio z prawdopodobieństwem

$$u_1 = p^2, \quad 2v_1 = 2pq, \quad w_1 = q^2.$$

W zależności od warunków występujących w populacji, jak i od rodzaju dziedziczonych cech, a także wielu innych czynników, dany model ulega zmianom. W kolejnych rozdziałach będziemy omawiać specyficzne przypadki powyższego modelu dziedziczenia.

3 Prawo Hardy'ego-Weinberga

3.1 Najważniejsze założenia

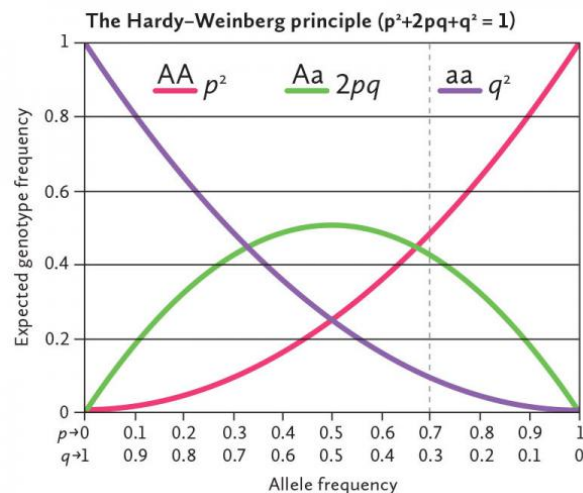
Częstości alleli i genotypów w populacji będą kształtować się zgodnie z przewidywaniami prawa Hardy'ego-Weinberga, jeżeli:

1. liczebność populacji jest nieskończenie wielka
2. nie zachodzą modyfikacje genetyczne
3. kojarzenia w populacji są losowe
4. osobniki rozmnażają się wyłącznie płciowo

3.2 Definicja

Jeżeli populacja genetyczna spełnia powyższe założenia i dwa allele A oraz a występują odpowiednio z częstościami p i q , gdzie $p + q = 1$, to:

1. częstość alleli w populacji nie zmienia się z pokolenia na pokolenie
2. częstość genotypów zależy tylko od częstości alleli genów i ustala się w kolejnych pokoleniach, niezależnie od początkowej częstości genotypów, w proporcjach p^2 , $2pq$, q^2 , będących rozwinięciem dwumianu $(p + q)^2$. Proporcję tę określa się mianem rozkładu Hardy'ego-Weinberga częstości genotypów.



Rysunek 1: Źródło ilustracji: <https://bigpictureeducation.com/evolutionary-forces>

3.3 Przykład

Albinizm jest cechą determinowaną przez allel recesywny. Częstość występowania albinizmu u szympansa wynosi raz na 20 tysięcy razy. Jaka jest częstość allela na albinizm oraz jaka jest częstość nosicieli allela na albinizm? Załóżmy, że gen albinizmu znajduje się w populacji w stanie równowagi genetycznej, czyli spełnia prawo Hardy'ego-Weinberga.

Rozwiązanie:

a - allel na albinizm

A - allel normalny (na produkcję ciemnego barwnika)

Wiadomo, że albinosami są tylko osobniki o genotypie aa .

Częstość występowania genu wynosi $w = \frac{1}{20000}$, zatem

$$q = \sqrt{w} = \sqrt{\frac{1}{20000}} \approx \frac{1}{141}$$

Widzimy zatem, że jeden allel na 141 jest zmutowany na albinizm. Widać od razu, iż $p = 1 - q = \frac{140}{141}$. Częstość nosicieli allele na albinizm wynosi

$$2v = 2pq = 2 \frac{1}{141} \frac{140}{141} = \frac{280}{19881} \approx \frac{1}{71}$$

Zatem wśród 71 osobników aż jedna jest nosicielem genu na albinizm, chociaż albinosem jest tylko jeden szympan na 20 000. Wiedząc, że dana populacja jest populacją w stanie równowagi, nie tylko mogliśmy wyliczyć powyższe zadanie, ale także wiemy, że powyższy rozkład częstości genów oraz liczba nosicieli genu na albinizm i albinosów będzie w przybliżeniu taka sama w każdym pokoleniu.

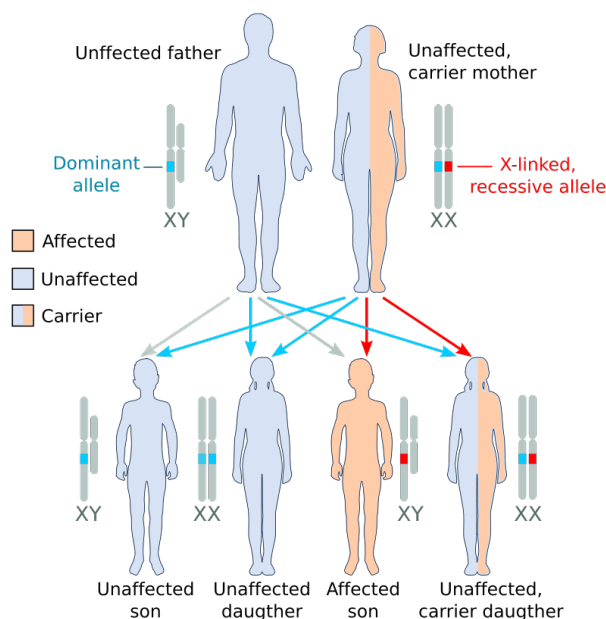
4 Cechy związane z płcią

Wiadomo, że geny leżą na chromosomach, które występują parami i są przekazywane jako odrębne jednostki. Płeć człowieka jest zdeterminowana przez dwa chromosomy: żeńska - XX i męska - XY . Matka przekazuje zawsze chromosomy X , zatem płeć dziecka zależy tylko i wyłącznie od chromosomu przekazanego przez ojca.

Daltonizm oraz hemofilia to jedne z najbardziej znanych chorób sprzężonych z płcią. Mutacje powodujące chorobę znajdują się tylko na chromosomach X .

Widzimy zatem, że osobniki płci żeńskiej mają trzy możliwe układy genotypów (AA , Aa , aa), a osobniki płci męskiej mają tylko jeden gen, zatem mogą należeć tylko do jednego z dwóch genotypów (A lub a), czyli mężczyźni nigdy nie będą nosicielami omawianych chorób.

Zauważmy, że jeśli ojciec jest daltonistą, to jego syn nigdy nie odziedziczy po nim danej choroby, jako że ojciec przekazuje mu tylko chromosom Y . To, czy syn będzie odróżniał kolory czy nie, zależy od genotypu matki.



Rysunek 2: Źródło ilustracji: <https://en.wikipedia.org/wiki/Haemophiliat>

Niech częstości genotypów AA , Aa , aa w żeńskiej populacji wynoszą, tak jak wcześniej, $u, 2v, w$. Przyjmijmy również $p = u + v$ oraz $q = v + w$. W męskiej populacji natomiast oznaczmy częstość A oraz a odpowiednio p' i q' ($p' + q' = 1$). Zatem p i p' to częstości genu A odpowiednio w populacji żeńskiej i męskiej.

Prawdopodobieństwo należenia potomka płci żeńskiej do genotypów AA , Aa , aa oznaczmy odpowiednio $u_1, 2v_1, w_1$, analogicznie dla płci męskiej mamy p'_1 oraz q'_1 . Wiemy też, że potomek męski otrzymuje swój chromosom X od matki, zatem

$$p'_1 = p, \quad q'_1 = q.$$

Dla genotypów żeńskich mamy:

$$u_1 = pp', \quad 2v_1 = pq' + p'q, \quad w_1 = q'q,$$

czyli

$$p_1 = u_1 + v_1 = pp' + \frac{1}{2}(pq' + p'q) = pp' + \frac{1}{2}(p(1-p') + p'(1-p)) = \frac{1}{2}(p + p')$$

$$q_1 = v_1 + w_1 = \frac{1}{2}(q + q')$$

Co mówią nam następujące zależności? Widzimy, że częstości genów wśród kobiet wynoszą w przybliżeniu średnią arytmetyczną częstości odpowiednich genów w populacji matek i ojców, a wśród mężczyzn geny występują z częstościami p i q populacji matek. Można zatem spostrzec tendencję do zrównania się częstości genów.

Postępując w taki sam sposób, dla drugiego pokolenia otrzymujemy (nie mamy tutaj do czynienia ze stabilną sytuacją już po jednym pokoleniu, tak jak w prawie Hardy'ego-Weinberga):

$$p_2 = \frac{1}{2}(p_1 + p'_1) = \frac{3}{4}p + \frac{1}{4}p'$$

$$q_2 = \frac{1}{2}(q_1 + q'_1) = \frac{3}{4}q + \frac{1}{4}q'$$

$$p'_2 = p_1 \quad q'_2 = q_1.$$

Przyjmijmy teraz

$$\alpha = \frac{1}{3}(2p + p'), \quad \beta = \frac{1}{3}(2q + q').$$

Ogólne wyrażenie na p_n oraz q_n to:

$$p_n = \frac{p_{n-1} + p'_{n-1}}{2} = \alpha + (-1)^n \frac{p - p'}{3 \cdot 2^n}$$

$$q_n = \frac{q_{n-1} + q'_{n-1}}{2} = \beta + (-1)^n \frac{q - q'}{3 \cdot 2^n}$$

$$p'_n = p_{n-1} \quad q'_n = q_{n-1}.$$

Widzimy zatem, że gdy n dąży do nieskończoności, to:

$$p_n \rightarrow \alpha, \quad p'_n \rightarrow \alpha, \quad q_n \rightarrow \beta, \quad q'_n \rightarrow \beta.$$

Częstości genotypów w żeńskiej populacji wynoszą:

$$u_n = p_{n-1} p'_{n-1}, \quad 2v_n = p_{n-1} q'_{n-1} + p'_{n-1} q_{n-1}, \quad w_n = q_{n-1} q'_{n-1},$$

zatem

$$u_n \rightarrow \alpha^2, \quad 2v_n \rightarrow 2\alpha\beta, \quad w_n \rightarrow \beta^2.$$

Widzimy, że przy losowym kojarzeniu się, można oczekiwać, iż płciowo skojarzone genotypy A i a wśród mężczyzn oraz AA , Aa i aa wśród kobiet będą występowały w przybliżeniu z częstościami odpowiednio α , β , α^2 , $2\alpha\beta$, β^2 , gdzie $\alpha + \beta = 1$.

Powróćmy teraz do wspomnianych wcześniej chorób sprzężonych z płcią. Wiemy, że geny związane z daltonizmem i hemofilią są recesywne. Niech a będzie genem powodującym daltonizm, wówczas defekt wykazują mężczyźni typu a oraz kobiety typu aa . Osobniki płci żeńskiej o genotypie Aa są nosicielkami mutacji, mogą przekazać defekt swoim potomkom, jednak same są "zdrowe". Widzimy zatem, na podstawie przeprowadzonych wcześniej obliczeń, że jeśli zmutowany gen a będzie występował w populacji męskiej z częstością α , to w populacji żeńskiej będzie występował z częstością α^2 . Na przykład, jeśli jeden na 100 mężczyzn jest daltonistą, to wśród kobiet wadę będzie wykazywać jedna osoba na 10000. Wyjaśnia to w prosty sposób, dlaczego daltonizm i hemofilia dotyczą przede wszystkim mężczyzn.

5 Selekcje

Selekcja to zjawisko polegające na eliminowaniu pewnych osobników z populacji lub uniemożliwieniu im pozostawiania potomstwa (przekazywania swoich genów następnemu pokoleniu). Z punktu widzenia genetyki jest to po prostu eliminacja pewnych alleli z populacji. Jako przykład weźmy przypadek, gdy osobniki typu aa nie mogą się rozmnażać (gen a jest recesywny i zgubny, czyli osobniki aa rodzą się, ale nie mogą żyć).

Załóżmy losowe kojarzenie się z osobnikami typu AA i Aa bez kojarzenia się z osobnikami typu aa . Standaardowo, niech częstości występowania powyższych genotypów wynoszą odpowiednio u , $2v$, w . Odpowiednie częstości dla rodziców wynoszą zatem:

$$u^* = \frac{u}{1-w}, \quad 2v^* = \frac{2v}{1-w}, \quad w^* = 0.$$

Częstości genów A i a można zatem zapisać jako:

$$p = \frac{u+v}{1-w}, \quad q = \frac{v}{1-w}.$$

Postępując w analogiczny sposób jak w poprzednim paragrafie, otrzymamy ogólne wzory na prawdopodobieństwa w n pokoleniu:

$$p_n = \frac{u_n + v_n}{1 - w_n}, \quad q_n = \frac{v_n}{1 - w_n}$$

oraz

$$u_{n+1} = p_n^2, \quad 2v_{n+1} = 2p_n q_n, \quad w_{n+1} = q_n^2.$$

Porównując ze sobą powyższe wzory i pamiętając, że $p_n + q_n = 1$, otrzymujemy następującą zależność:

$$p_{n+1} = \frac{u_{n+1} + v_{n+1}}{1 - w_{n+1}} = \frac{p_n^2 + p_n q_n}{1 - q_n^2} = \frac{p_n(p_n + q_n)}{1 - q_n^2} = \frac{1 - q_n}{1 - q_n^2} = \frac{1}{1 + q_n}.$$

Otrzymujemy również:

$$q_{n+1} = \frac{v_{n+1}}{1 - w_{n+1}} = \frac{q_n}{1 + q_n}.$$

Kontynuując operacje na otrzymanej zależności, mamy:

$$\frac{1}{q_{n+1}} = 1 + \frac{1}{q_n},$$

czyli

$$\frac{1}{q_1} = 1 + \frac{1}{q}, \quad \frac{1}{q_2} = 2 + \frac{1}{q}, \quad \frac{1}{q_3} = 3 + \frac{1}{q}, \quad \dots, \quad \frac{1}{q_n} = n + \frac{1}{q},$$

a zatem widzimy, że:

$$\frac{1}{q_n} = n + \frac{1}{q} = \frac{nq + 1}{q},$$

czyli

$$q_n = \frac{q}{nq + 1},$$

i zarazem

$$w_{n+1} = \left(\frac{q}{nq + 1}\right)^2$$

Widzimy, że niepożądany (niepożądanym) genotyp ulega stopniowemu zanikowi, proces ten jest jednak bardzo powolny. Jeśli przyjmiemy, że częstość allelu a wynosi $q = 0,1$, to widzimy, że musi przejść 10 pokoleń, żeby zredukować częstość tych genów o połowę (dla $q = 0,1$ i $n = 10$ mamy $q_{10} = 0,05$). $w_{10} = 0,0025 = 0,25\%$, gdzie $w_1 = 1\%$, czyli redukuje to częstość typów aa raptem o $0,75\%$.

Bibliografia

Praca została napisana na podstawie książki Williama Fellera *Wstęp do rachunku prawdopodobieństwa* oraz artykułu *Hardy-Weinberg principle* z Wikipedii.