

Естествен отбор в диплоидна популация с три фенотипа

Допускаме, че в популацията са възможни три фенотипа, в зависимост от разпределението на W , w -алели в дадения локус:

- ▶ WW белокрили молци
- ▶ Ww сивокрили молци
- ▶ ww чернокрили молци

Въпроси: При какви условия ще просъществуват и трите фенотипа?

Възможно ли е сивокрилите молци да изчезнат? При какви условия?

Ако сивокрилите изчезнат, кой от другите два фенотипа ще просъществува?

Параметри на предимство при селекцията

Относителни дялове от всеки вид, които оцеляват:

- ▶ α за белокрилите
- ▶ β за сивокрилите
- ▶ γ за чернокрилите

отбор в популация с три фенотипа

$$p_{n+1} = f(p_n) = \frac{(\alpha - \beta)p_n^2 + \beta p_n}{(\alpha - 2\beta + \gamma)p_n^2 + 2(\beta - \gamma)p_n + \gamma}$$

четири случая

случай	параметри	равновесни точки
1	$\alpha > \beta > \gamma$	
2	$\alpha < \beta < \gamma$	
3	$\beta > \gamma > \alpha, \beta > \alpha > \gamma$	
4	$\beta < \gamma < \alpha, \beta < \alpha < \gamma$	

четири случая

случай	параметри	равновесни точки	поведение $n \rightarrow \infty$
1	$\alpha > \beta > \gamma$	0, 1	$p \rightarrow 1$
2	$\alpha < \beta < \gamma$	0, 1	$p \rightarrow 0$
3	$\beta > \gamma > \alpha, \beta > \alpha > \gamma$	0, 1, p_*	$p \rightarrow p_*$
4	$\beta < \gamma < \alpha, \beta < \alpha < \gamma$	0, 1, p_*	$p \rightarrow 0$ или $p \rightarrow 1$

$$p_* = \frac{\gamma - \beta}{\alpha - 2\beta + \gamma}$$

числено изследване на устойчивостта на равновесните точки

метод на паяжината:

1. изобразяване на графиката на f в интервала $(0,1)$
2. изобразяване на точките $(x, f(x))$
3. избор на начално условие $p_0 \in (0,1)$
4. последователно изобразяване на точките $(p_n, f(p_n))$, (p_{n+1}, p_{n+1}) , имайки предвид рекурентната зависимост $p_{n+1} = f(p_n)$

упражнение

решете задачата за следните стойности

1. $\alpha = 0.8, \beta = 0.5, \gamma = 0.1$
2. $\alpha = 0.1, \beta = 0.5, \gamma = 0.7$
3. $\alpha = 0.1, \beta = 0.9, \gamma = 0.3$
4. $\alpha = 0.9, \beta = 0.1, \gamma = 0.7$

експериментирайте с няколко различни начални условия за p_0 .
в кой случай резултатът зависи качествено от началното
условие?

Ефект на случайността върху генетичната структура на популации

Генетичният дрейф представлява промяна в честотата на гените (алелите), която е случайна и няма приспособителен характер.

Генетичната структура на природните популации е силно зависими от случайния генетичен дрейф, тъй като случайността може

- ▶ да унищожат генетичното разнообразие, изградено чрез мутация,
- ▶ да противодействат на естествения отбор,
- ▶ да изградят статистически асоциации между различни локуси в генома

Затова изясняването на повечето въпроси в еволюционната биология изисква да се вземат под внимание *случайните ефекти*.

Практическа важност

Изясняването на случайните ефекти има и висока практическа важност. Например, фиксирането на мутации за лекарствена устойчивост в организми, причиняващи заболявания (независимо дали от нововъзникнали или вече съществуващи мутации) зависи в значителна степен от генетичния дрейф. Да разгледаме нововъзникналата мутация: в много голяма популация се генерират мутации за лекарствена устойчивост във всяко поколение и случайността има малка роля. При малка популация обаче такива мутации възникват с доста малка вероятност и затова времето на появата им е случайно.

Пример

Тази случайност е необходима да обясни факта, че времето до появата на лекарствена устойчивост на вируса на ХИВ варира до голяма степен от пациент на пациент (Nijhuis et al., *Proc Natl Acad Sci USA*, 1998).

Значението на случайните ефекти зависи от много параметри на модела (размера на популацията или скоростта на мутация). В това упражнение ще изведем прости случайни модели, за да разграничим ролята различните фактори и за да получим по-ясна представа за генетичния дрейф.

Основен модел

В основния модел разглеждаме само един локус с алели A и a .
Двата алела имат честоти f_A и $f_a = 1 - f_A$.

За простота предполагаме, че поколенията са дискретни и във всяко има *мутация* и *отбор* (селекция).

По принцип трябва да симулираме мутацията и естествения отбор като случайни процеси.

За да опростим модела, разглеждаме приближение, в което

- ▶ мутацията и отборът са едновременно детерминистични,
- ▶ крайният размер на популацията се определя чрез вземане на извадка от N генома (N е размерът на популацията), след като са се случили мутацията и естественият отбор.

Стъпките на модела 1

Мутация

Нека относителен дял m , $0 < m < 1$ от всички a -алели да мутира в A -алел, както и относителен дял m , $0 < m < 1$ от всички A -алели да мутира в a -алел.

Съответно остатъчният дял $1 - m$ от всички A -алели не се променя. Тогава честотата на A -алела след мутация е зададена с изрази:

$$f_A^{[1]} = mf_a + (1 - m)f_A = m(1 - f_A) + (1 - m)f_A$$

Стъпките на модела 2

Естествен отбор

Нека двата алела имат приспособеност съответно $w_a = 1$ и $w_A = 1 - s$, където s е коефициентът на естествен отбор.

С $\bar{w} = w_A f_A^{[1]} + w_a f_a^{[1]}$ означаваме средната приспособеност в популацията.

Конкретно в случая

$$\bar{w} = (1 - s)f_A^{[1]} + 1(1 - f_A^{[1]}).$$

След стъпката на естествен отбор A -алелът има нова честота в популацията

$$f_A^{[2]} = f_A^{[1]} \cdot \frac{w_A}{\bar{w}} = f_A^{[1]} \frac{(1 - s)}{(1 - s)f_A^{[1]} + (1 - f_A^{[1]})}$$

Стъпките на модела 3

Извадка

Получаваме броя на A -алели след извадката, като теглим произволно число N_A съгласно *биномно разпределение* с параметри N и $f_A^{[2]}$.

Това съответства на случайна извадка от общо N потомствени алела от генофонд, в който делът на A -алелите е $f_A^{[2]}$, а остатъкът е a -алели.

Честотата на A -алелите след извадката е $f_A^{[3]} = \frac{N_A}{N}$.

Целта ни е да проследим промените в честотата на A -алела в последователни поколения в дадената популация.

Въпрос. Този модел хаплоидна (безполова) или диплоидна (полова) популация ли описва?

Алгоритъм

Симулацията на честотните промени на A -алела в едно поколение се състои в последователно извършване на стъпките

1. мутация,
2. отбор/селекция
3. извадка

и със задаване на новата честота на A -алел в следващото поколение

$$f_A := f_A^{[3]}$$

Забележка. Обърнете внимание, че не трябва да проследяваме изрично f_a , защото винаги имаме следната зависимост

$$f_a = 1 - f_A.$$

Биномно разпределение

Matlab има функция `binornd()`, която генерира случайни числа от биномно разпределение с входни параметри N (брой опити), P (вероятност за успех).

В R тази функция е `rbinom()`.

Внимание! Необходимо е функцията да върне само една стойност от биномно разпределение. С кой аргумент на функцията трябва да зададем това?

Код за R: инициализация на параметри

```
# implements the basic model of stochastic population
genetics
#--set parameters
populationSize<-10000 # corresponds to N
mutationRate<-0.001 # corresponds to m
selectionStrength<-0 # corresponds to s
#-- set initial condition
frequency1<-0.1 # frequency of the allele to be
followed (A)
numberOfGenerations<-40000
```

Код за R: цикъл

Инициализираме списък, в който ще се записват честотите във всяко поколение

```
frequencies<-numeric(numberOfGenerations)  
time<-1:numberOfGenerations
```

Код за R: цикъл

Цикъл с трите стъпки на модела

```
for(i in time){  
  ##--mutation  
  frequency1<-...  
  ##--selection  
  frequency1<-...  
  ##--sampling  
  frequency1<-...  
  ##--saving values  
  frequencies[i]<-frequency1  
}
```

Графично изобразяване на резултата

С команда

```
plot(frequencies, time,type="l")
```

Задачи

- ▶ Изпробвайте различни параметри и проверете как те влияят върху кривата на развитие.
- ▶ Започнете с неутрален модел (т.е. положете $s = 0$) и променяйте N и m . След това въведете и отбор ($s > 0$) в модела.
- ▶ За някои случаи обмислете и граничния детерминистичен случай, като
 - ▶ зададете много голяма стойност за N **или**
 - ▶ елиминирате стъпката *извадка* от симулацията.

Проверете дали и как първият подход се доближава до втория, когато увеличавате N .

Важна забележка

Тъй като симулацията включва случаен процес (стъпка *извадка* в модела), всяко нейно изпълнение ще бъде различно от предишното.

За да получите надеждна представа за възможните вариации на поведението на модел, трябва да изпълните няколко симулации със същите стойности на параметрите (и да ги запазите, за да анализирате разпределението на резултатите).

Упътване.

За да стартирате по-лесно симулации с различни параметри, напишете обвивка около основния код, която да го извиква за различни стойности, например

```
sim(initialFrequency=0.1,numberOfGenerations=10000,  
selectionStrength=0,  
mutationRate=0.01,populationSize=100,plotTimeCourse=TRUE)
```


Разширен модел: кръстосване

Какъв е ефектът от кръстосването?

Това е процес, при който нишки ДНК се прекъсват и съединяват наново, като образуват нови комбинации от алели. Нека включим в модела два локуса и два алела – получаваме като възможни комбинации следните генотипи, които ще означим условно като:

1. ab (див щам)
2. aB и Ab (единични мутанти)
3. AB (двоен мутант)

Тази постановка ни позволява да изследваме как кръстосването влияе върху хода на еволюцията.

Разширен модел: кръстосване

За повече сведения за ролята на кръстосването, може да се консултирате с обзорната статия (Otto & Lenormand, *Nature Rev Genet*, 2002).

Разширен модел: кръстосване

За повече сведения за ролята на кръстосването, може да се консултирате с обзорната статия (Otto & Lenormand, *Nature Rev Genet*, 2002).

Разглеждането на два локуса изисква малки промени в модела!

Стъпките на модела 1

Мутация. За улеснение допускаме, че и двата алела мутират с еднаква вероятност.

Да разгледаме генотипа ab . С вероятност $(1 - m)^2$ той не мутира.

С вероятността $(1 - m)m$ генотип aB или Ab мутира до ab (единият алел в локуса мутира, а другият не).

За да се измени AB в ab , двата локуса трябва да мутират едновременно (с вероятност m^2).

По този начин честотата на генотипа ab след мутация става

$$f_{ab}^{[1]} = (1 - m)^2 f_{ab} + (1 - m)m(f_{Ab} + f_{aB}) + m^2 f_{AB}.$$

За другите три генотипа формулите могат да бъдат получени по подобен начин. Направете го сами.

Стъпките на модела 2

Естествен отбор

Без промяна:

$$f_{xy}^{[2]} = f_{xy}^{[1]} \frac{w_{xy}}{\bar{w}}, xy \in \{ab, aB, Ab, AB\}.$$

където

$$w_{ab}, w_{aB}, w_{Ab}, w_{AB}$$

са някакви положителни стойности.

Как трябва да се сметне средната приспособеност $\bar{w}$?

Стъпките на модела 3 (нова)

Кръстосване

Можем да запишем честотите на генотипите и по следния начин:

$$\begin{aligned}f_{ab} &= p_a p_b + D \\f_{Ab} &= p_A p_b - D \\f_{aB} &= p_a p_B - D \\f_{AB} &= p_A p_B + D,\end{aligned}$$

където с p се означават честотите на алелите и $D = f_{ab}f_{AB} - f_{Ab}f_{aB}$ е коефициентът на *неравновесие на свързване* на гените.

Неравновесие на свързване

Два или повече локуса са в неравновесие на свързване, ако честотата на различните свързани алели в тях е различна от очакваната честота, ако локусите биха били независими и свързани случайно.

Ако честотата на алел A в популация от гамети е f_A , а на алел B е f_B , а честотата на генотипа AB е f_{AB} , съответните локуси са в неравновесие на свързване, ако $f_{AB} \neq p_A p_B$.

В нашия случай алели a, A могат да бъдат само в първия локус, а b, B само във втория, което извежда израза за коефициента

$$D := f_{AB} - p_A p_B$$

Неравновесие на свързване

В нашия случай алели a, A могат да бъдат само на първия локус, а b, B само във втория, което означава, че

$$\begin{aligned} D &= f_{AB} - p_A p_B = f_{AB} - (f_{Ab} + f_{AB})(f_{aB} + f_{AB}) \\ &= f_{AB}(1 - (f_{Ab} + f_{aB} + f_{AB})) - f_{Ab} f_{aB} \\ &= f_{AB} f_{ab} - f_{Ab} f_{aB} \end{aligned}$$

Кръстосване

Да допуснем, че r е скоростта на кръстосване на случаен принцип. Тогава в следващото поколение честотата на генотипа AB ще се промени, както следва

$$f_{AB}^{next} = rp_{APB} + (1 - r)f_{AB}$$

Кръстосването не засяга честотите на алелите, следователно неравновесието на свързване се променя така:

$$\begin{aligned} D^{next} &= f_{AB}^{next} - p_{APB} \\ &= rp_{APB} + (1 - r)f_{AB} - p_{APB} \\ &= (1 - r)D \end{aligned}$$

Кръстосване

Да допуснем, че r е скоростта на кръстосване на случаен принцип. Тогава в следващото поколение честотата на генотипа AB ще се промени, както следва

$$f_{AB}^{next} = rp_{APB} + (1 - r)f_{AB}$$

Кръстосването не засяга честотите на алелите, следователно неравновесието на свързване се променя така:

$$\begin{aligned} D^{next} &= f_{AB}^{next} - p_{APB} \\ &= rp_{APB} + (1 - r)f_{AB} - p_{APB} \\ &= (1 - r)D \end{aligned}$$

Извод: кръстосването намалява D с коефициент $1 - r$, където r е скоростта на кръстосване.

Стъпките на модела 3

Затова честотите на генотипите в модела след кръстосване стават

$$f_{ab}^{[3]} = f_{ab}^{[2]} - rD$$

$$f_{Ab}^{[3]} = f_{Ab}^{[2]} + rD$$

$$f_{aB}^{[3]} = f_{aB}^{[2]} + rD$$

$$f_{AB}^{[3]} = f_{AB}^{[2]} - rD$$

Стъпките на модела 4

Извадка.

Полиномното разпределение (зададено с `mnrnd()` в Matlab и с `rmultinom()` в R) предоставя естествено обобщение на биномното разпределение, когато трябва да се направи извадка от повече от два генотипа.

Кръстосване и полезни мутации

Нека се опитаме да разберем как кръстосването влияе върху фиксацията на полезни мутации. Ще разгледаме следната постановка.

Да допуснем, че стойностите на приспособеност на генотипите удовлетворяват

$$w_{ab} < w_{Ab} = w_{aB} < w_{AB},$$

т.е. и за двата локуса, алелът на дивия щам има по-ниска приспособеност (предположението $w_{Ab} = w_{aB}$ е направено за простота).

Кръстосване и полезни мутации

Нека да допуснем, че

$$w_{ab} = 1, w_{Ab} = w_{aB} = 1 + s, w_{AB} = (1 + s)^2.$$

Начални условия: Започнете с честоти

$$(f_{ab}, f_{Ab}, f_{aB}, f_{AB}) = (1, 0, 0, 0),$$

т.е. популацията се състои първоначално само от неприспособения див щам.

Кръстосване и полезни мутации

Тази постановка може да се разглежда като опростяване на процеса на възникване на мутации за устойчивост към две различни лекарства (ако допуснем, че мутациите, обуславящи устойчивостта, са кодирани с алелите A и B). Така горното допускане за приспособеността на генотипите означава, че в началната популация се състои от чувствителния и към двете лекарства див щам ab , а най-устойчивият генотип (с най-висока приспособеност) е двойноустойчивият мутант AB .

Задачи

- ▶ Въведете тези условия във вашия скрипт за разширения модел. Използвайте например следните стойности

$$m = 10^{-4}, w_{ab} = 1, w_{aB} = w_{Ab} = 1.01, w_{AB} = 1.012, \\ r = 0.1, N = 10^4$$

- ▶ Направете няколко симулации за няколкостотин поколения и пресметнете средната стойност на честотите във всяко поколение. Изобразете резултатите графично.
- ▶ Изобразете няколко симулации за различни стойности на параметрите. Откривате ли някаква закономерност в промяната в генотипното разпределение?

Упътване

Използвайте масив, за да запишете стойностите на честотите на всеки генотип.

```
# initial condition frequency<-c(1,0,0,0)
numberOfGenerations<-1000
frequencies<-matrix(0,nrow=numberOfGenerations,ncol=4)
#--matrix array
time<-1:numberOfGenerations
```

Матрица: ускоряване на пресмятането

За да ускорим пресмятането на модела е добре да въведем матрица за вероятностите при мутация

```
mutationMatrix<-  
(1-m)^2, m(1-m), m(1-m),m^2  
m(1-m), (1-m)^2, m^2, m(1-m)  
m(1-m),m^2, (1-m)^2, m(1-m)  
m^2, m(1-m), m(1-m),(1-m)^2
```

с която ще умножим вектора от честоти, за да пресметнем промените при стъпка 1.

```
frequency<-(mutationMatrix %*% frequency)
```