ABSTAT17

IGC, April 10-13, 2017

EXERCISE: EM Algoritm

The main locus for the blood type of mice is called Ag-B (B). Several alleles are associated to this locus but for some crossovers Mendel's laws do not seem to hold. A mating $AaBb \times AaBb \equiv F_1 \times F_1$, originated a F_2 progeny, yielding

Genotype	Frequency	Probability
$AABB$	11	$(1-\theta)^2/4$
$AABb$	14	$\theta(1-\theta)/2$
$AAbb$	1	$\theta^2/4$
$AaBB$	10	$\theta(1-\theta)/2$
$AaBb$	27	$(\theta^2/2)+[(1-\theta)^2]/2$
$Aabb$	12	$\theta(1-\theta)/2$
$aaBB$	3	$\theta^2/4$
$aaBb$	13	$\theta(1-\theta)/2$
$aabb$	11	$(1-\theta)^2/4$

Estimate the recombination fraction, θ, from these data by the EM algorithm.

Step 1

Read the data and state ao many recombinant gametes are there for each genotype.

```
nAABB<-11 # 0 recombinant gametes
nAABb<-14 # 1 recombinant gamete
nAAbb<-1 # 2 recombinant gametes
nAaBB<-10 # 1 recombinant gamete
nAaBb<-27 # 0 or 2 recombinant gametes
nAabb<-12 # 1 recombinant gamete
naaBB<-3 # 2 recombinant gametes
naaBb<-13 # 1 recombinant gamete
naabb<-11 # 0 recombinant gametes
```

Calculate n_1, the number of individuals from 1 recombinant gametes (`n1`).

Calculate n_2, the number of individuals from 2 recombinant gametes (`n2`). Note that $n_{AaBb} = n_2^* + n_0^*$ (`nAaBb = n2.star + n0.star`).

Calculate n, the total number of individuals (`n`).

Step 2

Initialize $\theta \in]0, 0.5[$ (`r`).

Step 3 - E (Expectation)

Create function `expected` in order to calculate the expected value for N_2^*:

N_2^*: random variable representing the number of individuals from 2 recombinant gametes, among n_{AaBb} individuals.

$$N_2^* \frown Binomial(n_{AaBb}, p) \quad \text{with} \quad p = \frac{\theta^2}{\theta^2 + (1-\theta)^2}$$

then, $n_2^* = E(N_2^*) = n_{AaBb} \times p$

Step 4 - M (Maximization)

Create function `update.theta` in order to update θ according to:

$$\theta = \frac{n_1 + 2(n_{AAbb} + n_{aaBB} + n_2^*)}{2n}$$

meaning that the proportion of recombinant gametes is calculated as the total number of recombinant gametes (0, 1 or 2 for each individual) over the total number of gametes for n individuals.

Step 5

EM algorithm: Iterative procedure.

Step 6

Print the results.