

# Raonaments per identificar relacions de parentiu

Memòria escrita

Anna Aumatell, Alicia Chimeno  
Irene de Merlo, Laia Lluís

Juny 2021

**Resum:** Per tal d’afirmar si dues persones són família, a primera vista ens fixem en els seus trets físics com el color dels ulls, el de la pell, entre d’altres. Per saber-ho amb més certesa, s’estudia l’ADN, en concret, unes zones específiques que s’anomenen STR’s.

En aquest treball ens hem centrat en reconèixer algunes relacions de parentiu. El problema central gira al voltant d’afirmar o negar si un senyor concret és el pare d’un noi. Utilitzant conceptes de genètica i probabilitat hem elaborat un mètode que ens ho permet reconèixer.

**Abstract:** In order to determine if two people are relatives, at first glance we look at their physical features such as the eye color, the skin color, among others. In order to be certain, we study the DNA, more concretely, particular areas called STR’s.

In this project we have focused on recognizing some specific relationships. The central problem is meant to affirm or deny whether a particular man is the father of a boy. Using concepts of genetics and probability we have developed a method that allows us to recognize it.

# Índex

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introducció</b>                          | <b>3</b>  |
| <b>2</b> | <b>Teoria</b>                               | <b>3</b>  |
| 2.1      | Genètica. Definicions . . . . .             | 3         |
| 2.1.1    | Vocabulari bàsic. Lleis de Mendel . . . . . | 3         |
| 2.1.2    | Conceptes importants . . . . .              | 4         |
| 2.2      | Probabilitat . . . . .                      | 5         |
| 2.2.1    | Exemple senzill . . . . .                   | 6         |
| <b>3</b> | <b>Problema</b>                             | <b>7</b>  |
| 3.1      | Context del problema . . . . .              | 7         |
| 3.2      | Simplificacions adoptades . . . . .         | 7         |
| 3.3      | Ressolució . . . . .                        | 7         |
| 3.4      | Resposta . . . . .                          | 10        |
| <b>4</b> | <b>Ampliació 1 del problema</b>             | <b>10</b> |
| <b>5</b> | <b>Ampliació 2 del problema</b>             | <b>11</b> |
| <b>6</b> | <b>Possibles refinaments</b>                | <b>13</b> |
| <b>7</b> | <b>Conclusions</b>                          | <b>13</b> |

## Enunciat

*L'ADN d'una persona resulta de combinar de l'ADN dels seus pares. Aquesta combinació té un component aleatori, però també s'ajusta a certs patrons que haurien de permetre reconèixer amb bastant de confiança les relacions de parentiu més properes. Com es pot fer això?*

## 1 Introducció

La genètica és una de les branques més noves de la biologia, en concret, s'inicia al segle XX amb les lleis de Mendel. Més tard, els científics Watson, Crick i Franklin, descobreixen que la informació genètica es troba emmagatzemada a una molècula que posteriorment s'anomenaria ADN. Avui en dia, sabem que l'ADN conté les instruccions genètiques per al desenvolupament, el funcionament, el creixement i la reproducció de tots els organismes coneguts i molts virus. Mitjançant l'estudi de l'ADN podem detectar malalties hereditàries, relacions de parentiu, etc.

En aquest treball aprofundirem en el tema de les relacions de parentiu. Hem volgut veure com es poden aplicar conceptes matemàtics per detectar les relacions de parentiu més properes. Les analitzarem de manera probabilística per tal de trobar un mètode efectiu que es pugui aplicar en diverses situacions.

## 2 Teoria

### 2.1 Genètica. Definicions

#### 2.1.1 Vocabulari bàsic. Lleis de Mendel

**Definició 2.1.1.** Un *gen* és una part del cromosoma que conté informació per a un caràcter. Les diferents variants que pot tenir un gen són els *al·lels*.

**Definició 2.1.2.** El conjunt de gens que té un organisme és el que anomenem *genotip*.

**Definició 2.1.3.** Diem que un organisme és *homozigot* quan té un parell idèntic d'al·lels per un gen concret. I es diu que un organisme és *heterozigot* quan aquest té dos al·lels diferents per un gen concret.

**Teorema 2.1.1.** *Lleis de Mendel:*

- **Primera Llei de Mendel** o Llei de la uniformitat dels híbrids de la primera generació filial: Quan s'encreuen dues varietats pures tots els híbrids de la primera generació ( $F_1$ ) són iguals.

- **Segona Llei de Mendel** o *Llei de la segregació*: Els dos factors hereditaris que informen sobre un caràcter queden diferenciats durant tota la vida de l'individu i es separen i reparteixen en el moment de la formació dels gàmetes per tornar-se a unir durant la fecundació.
- **Tercera Llei de Mendel** o *Llei de la distribució independent de caràcters*: En la formació dels gàmetes, els al·lells de cadascun dels gens se segreguen independentment dels al·lells de qualsevol altre gen.

### 2.1.2 Conceptes importants

**Definició 2.1.4.** *STR* de l'anglès *short tandem repeat* són seqüències curtes d'ADN que es repeteixen de manera consecutiva. Aquestes seqüències poden no tenir una funció específica, però donen lloc a una elevada variabilitat entre els ADN de les persones que pot ser útil per propòsits d'identificació.

La variació del nombre de repeticions de la seqüència (STR) crea diferents al·lells. És a dir, un individu que sigui 7-11 per un STR significarà que en una posició específica del cromosoma té 7 repeticions d'aquesta seqüència i 11 repeticions d'aquesta mateixa en el cromosoma homòleg.

A més a més, els STRs tenen una herència mendeliana simple, això vol dir que un al·lel és heretat del pare i l'altre de la mare, per aquest motiu són de gran utilitat en l'anàlisi de casos de paternitat.

Basat en diferents estudis sobre l'ADN, s'ha determinat que hi ha com a mínim 13 STRs que es troben en l'ADN de tots els humans. L'anàlisi d'aquests és suficient per poder determinar les relacions de parentiu de manera fiable.

**Definició 2.1.5.** Per cadascun d'aquests STRs, mitjançant diferents estudis, s'ha determinat la **frequència** de cada al·lel (nombre de repeticions) en la població total.

Per calcular la probabilitat de tenir un genotip concret s'han d'utilitzar les freqüències de cada al·lel. Distingim dos tipus de genotips

- En el cas que el nombre de repeticions d'un al·lel sigui homozigot, és a dir, tingui les mateixes repeticions als dos cromosomes, per calcular la probabilitat de tenir aquest nombre de repeticions a la població, es calcula de la següent manera, on  $X = (x,x)$  :  $P(X) = f_x^2$ , ja que tant la mare com el pare li passen el mateix nombre de repeticions.
- En canvi, si l'al·lel és heterozigot, és a dir, un cromosoma i el seu homòleg tenen un nombre diferent de repeticions del STR. Aquest es calcula de la següent forma, on  $Y = (x,y)$  :  $P(Y) = f_x \cdot f_y + f_y \cdot f_x = 2 \cdot f_x \cdot f_y$ . És a dir, primerament es calcula tenint en compte que el

pare li ha passat el nombre de repeticions  $x$  i la mare el  $y$ , i després el cas contrari en què el pare passa el nombre de repeticions  $y$  i la mare el  $x$ .

## 2.2 Probabilitat

Per poder dur a terme el nostre treball hem hagut d'adquirir diferents conceptes de probabilitat, que han estat els següents:

**Teorema 2.2.1. Teroema de Bayes:** *Siguin  $A$  i  $B$  dos esdeveniments. Calculem la probabilitat condicionada d'aquests a partir de:*

$$P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B | A) \cdot P(A)}{P(B)} \quad (1)$$

La probabilitat d'intersecció de dos esdeveniments es calcula de forma diferent segons si aquests són dependents o independents:

- Dependents: Siguin  $A$  i  $B$  dos esdeveniments, aleshores:

$$P(A \cap B) = P(A | B) \cdot P(B) = P(B | A) \cdot P(A) \quad (2)$$

- Independents: Siguin  $A$  i  $B$  dos esdeveniments, aleshores:

$$P(A \cap B) = P(A) \cdot P(B) \quad (3)$$

**Definició 2.2.1. L'odds** és la mesura del grau d'incertesa d'un esdeveniment. Es calcula com el quocient de la probabilitat de què passi l'esdeveniment i la probabilitat de què aquest esdeveniment no passi. Indica quant més probable és que passi l'esdeveniment que el seu contrari (és a dir, que no passi).

Distingim dos tipus de càlculs de l'odds d'un esdeveniment:

- **Odds a priori:**  $\text{Odds}(A) = \frac{P(A)}{P(A^c)}$ , on  $A^c$  és l'esdeveniment complementari d' $A$ .
- **Odds a posteriori:**  $\text{Odds}(A | E) = \frac{P(A|E)}{P(A^c|E)}$ . Aquest quocient mesura quant més probable és que passi un esdeveniment, a que no passi, ambdós esdeveniments condicionats per un altre, que en aquest cas és  $E$ .

**Definició 2.2.2. L'odds ratio** ens permet avaluar el pes de les evidències. Es calcula com el quocient entre l'odds a posteriori i l'odds a priori. Tracta la relació entre dos esdeveniments comparant l'odds d' $A$  sabent que passa  $E$  amb l'odds d' $A$ .

$$\text{odds ratio} = \frac{\text{odds a posteriori}}{\text{odds a priori}} = \frac{P(A | E)/P(A^c | E)}{P(A)/P(A^c)}$$

**Definició 2.2.3.** La *raó de versemblança* és una manera d'interpretar l'odds ratio, s'obté aplicant les fórmules de Bayes en l'odds ratio de la manera següent:

**Demostració:** Volem veure que  $LR = \frac{Odds(A|E)}{Odds(A)} = \frac{P(E|A)}{P(E|A^c)}$ .  
Partim de la definició de odds, i tenim:

$$odds(A) = \frac{P(A)}{P(A^c)}; \quad odds(A|E) = \frac{P(A|E)}{P(A^c|E)}$$

tal que  $P(A) = \frac{P(A|E)P(E)}{P(E|A)}$  i  $P(A^c) = \frac{P(A^c|E)P(E)}{P(E|A^c)}$ . Aleshores, si substituïm a la fórmula del Odds tenim:

$$odds(A) = \frac{P(A)}{P(A^c)} = \frac{P(A|E)P(E)P(E|A^c)}{P(E|A)P(A^c|E)P(E)}$$

Aleshores si substituïm, obtenim que:

$$LR = \frac{odds(A|E)}{odds(A)} = \frac{P(A|E)P(E|A)P(A^c|E)}{P(A^c|E)P(A|E)P(E|A^c)} = \frac{P(E|A)}{P(E|A^c)}$$

I per tant, veiem el que voleim demostrar:

$$\text{raó de versemblança} = LR = \frac{odds(A|E)}{odds(A)} = \frac{P(E|A)}{P(E|A^c)}$$

### 2.2.1 Exemple senzill

Tenim un dau amb 6 cares. Volem saber quant de més probable és que surti un 5 sabent que el nombre que ha sortit és senar, en comparació a sense saber-ho.

Aquesta pregunta ens demana clarament la odds ratio de que surti un 5. Declarem les variables dels esdeveniments:

$$\begin{aligned} A &= \text{treure un 5 al dau} \\ A^c &= \text{no treure un 5 al dau} \\ E &= \text{ser senar} \end{aligned}$$

Per resoldre el problema calculem les probabilitats de cada esdeveniment:

$$P(A) = 1/6 \quad P(A^c) = 5/6 \quad P(E) = 1/2$$

$$P(A|E) = 1/3 \quad P(A^c|E) = 2/3$$

Un cop hem calculat les probabilitats de cada esdeveniment i les probabilitats condicionades, ja podem calcular la odds:

$$odds(A) = \frac{P(A)}{P(A^c)} = \frac{1/6}{5/6} = 1/5; \quad odds(A|E) = \frac{P(A|E)}{P(A^c|E)} = \frac{1/3}{2/3} = 1/2$$

Finalment, un cop tenim les odds necessàries, calculem la odds ratio:

$$\frac{odds(A|E)}{odds(A)} = \frac{1/2}{1/5} = 5/2 = 2,5$$

Resposta: És 2,5 vegades més probable que surti un 5 sabent que el nombre que ha sortit és senar, que no sabent res.

### 3 Problema

#### 3.1 Context del problema

Volem reconèixer amb confiança el pare biològic d'una persona.

Les hipòtesis que hem adoptat són que coneixem els genomes de les 3 persones implicades (senyor, mare, fill)

Hem extret les dades d'aquesta taula :

| Sistema/ Mostra | Pare  | Fill  | Mare  | Sistema/Mostra | Pare  | Fill  | Mare  |
|-----------------|-------|-------|-------|----------------|-------|-------|-------|
| D8S1179         | 12-13 | 13-13 | 13    | D19S433        | 12-13 | 13-14 | 12-14 |
| D21S11          | 30-   | 30-   | 28-30 | vWA            | 18    | 15-18 | 15-19 |
|                 | 32.2  | 32.2  |       |                |       |       |       |
| D7S820          | 10    | 9-10  | 9-12  | TPOX           | 8     | 8-8   | 8-11  |
| CSF1PO          | 11-12 | 11-12 | 11    | D18S51         | 13-17 | 14-17 | 13-14 |
| D3S1358         | 17-18 | 15-17 | 14-15 | Amelogenina    | XY    | XY    | XX    |
| TH01            | 6-9   | 9-9.3 | 7-9.3 | D5S818         | 11-12 | 11-11 | 11-12 |
| D13S317         | 8-11  | 8-13  | 11-13 | FGA            | 22-23 | 21-22 | 20-21 |
| D16S539         | 11-13 | 11-12 | 12-13 | Penta E        | 11-13 | 11-13 | 11-12 |
| D2S1338         | 18-24 | 18-19 | 18-19 | Penta D        | 11-12 | 11-12 | 12-13 |

Figura 1: dades extretes de la font [7]

#### 3.2 Simplificacions adoptades

- Suposem que no hi ha mutació a l'ADN, és a dir, en certs casos és possible que l'ADN pateixi mutacions i això pot fer variar els al·lels per algun STR. Nosaltres suposem que en el cas del problema aquestes mutacions no es donen, ja que generalment això no passa i a més podria ser que fos gairebé impossible de resoldre.
- Suposarem que la probabilitat de que un home sigui pare en la població total d'homes és del 50%.

#### 3.3 Resolució

En aquest problema, ens centrarem en fer els càlculs per un STR concret: D13S317.

El nostre objectiu principal es calcular  $P(A|N \cap M \cap S)$ .

Hem de tenir en compte els següents esdeveniments:

|   |   |                                     |
|---|---|-------------------------------------|
| M | = | tenir el genoma de la mare (11,13)  |
| N | = | tenir el genoma del noi (8,13)      |
| S | = | tenir el genoma del senyor (8,11)   |
| A | = | que el senyor sigui el pare del noi |

A continuació, hem de fer els següents passos intermedis:  
 Calculem les probabilitats de tenir les repeticions de cada genoma a la població total:

$$P(M) = 2 \cdot f_{11} \cdot f_{13} = 2 \cdot 0'236 \cdot 0'118 = 0'056$$

$$P(S) = 2 \cdot f_8 \cdot f_{11} = 2 \cdot 0'121 \cdot 0'236 = 0'057$$

A continuació, calculem les probabilitats condicionades en dos sentits, que necessitem més endavant:

Primerament, calculem la probabilitat de que el fill tingui el genoma N sabent el genoma del seu pare i el de la seva mare, i després sabent només el de la mare:

$$P(N|M \cap S \cap A) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \text{ i } P(N|M) = 1/2 \cdot f_8 = 0,5 \cdot 0,121 = 0,0605$$

Seguidament, amb les **fórmules de Bayes** calculem la probabilitat de que el senyor tingui S genoma condicionat al genoma del noi i de la mare, i a més que el senyor és el pare:

$$\begin{aligned} P(S|M \cap N \cap A) &= \frac{P(N|M \cap S \cap A) \cdot P(M \cap S \cap A)}{P(M \cap N \cap A)} = \\ &= \frac{P(N|M \cap S \cap A) \cdot P(M) \cdot P(S) \cdot P(A)}{P(M \cap N) \cdot P(A)} = \frac{P(N|M \cap S \cap A) \cdot P(M) \cdot P(S)}{P(N|M) \cdot P(M)} = \\ &= \frac{P(N|M \cap S \cap A) \cdot P(S)}{P(N|M)} = \frac{0'25 \cdot 0'057}{0'5 \cdot f_8} = 0'236 \end{aligned}$$

Però la pregunta del nostre problema ens demana la probabilitat de que el senyor amb genoma S sigui el pare del noi.

Per trobar-la, procedim a calcular la raó de versemblances (LR):

$$LR = \frac{Odds(A|M \cap N \cap S)}{Odds(A)} = \frac{P(M \cap N \cap S|A)}{P(M \cap N \cap S|A^c)}$$

Començarem calculant la probabilitat del numerador:



$$\begin{aligned}
P(M \cap N \cap S | A) &= \frac{P(M \cap N \cap S \cap A)}{P(A)} = \frac{P(S | M \cap N \cap A) \cdot P(M \cap N \cap A)}{P(A)} \\
&= \frac{P(S | M \cap N \cap A) \cdot P(M \cap N) \cdot P(A)}{P(A)} = P(S | M \cap N \cap A) \cdot P(N | M) \cdot P(M) \\
&= 0'236 \cdot 0'0605 \cdot 0'056 = 7'95 \cdot 10^{-4}
\end{aligned}$$

Seguidament, calculem la probabilitat que es troba al denominador:

$$\begin{aligned}
P(M \cap N \cap S | A^c) &\simeq P(M \cap N \cap S) = P(N \cap M) \cdot P(S) \\
&= P(N | M) \cdot P(M) \cdot P(S) = 0'0605 \cdot 0'056 \cdot 0'057 = 1'93 \cdot 10^{-4}
\end{aligned}$$

Finalment, obtenim que:

$$LR = \frac{7'95 \cdot 10^{-4}}{1'93 \cdot 10^{-4}} = 4'11$$

Aquest resultat ens diu que l'odds condicionat (a posteriori) és 4 cops més gran que l'odds sense condicionar (a priori). Per tant és 4,11 vegades més probable que sigui pare sabent el genoma de la mare que sense saber res.

Observem que la fórmula de LR es pot simplificar molt, gràcies a Bayes, i així facilitarem els nostres càlculs per la resta de STR:

$$LR = \frac{P(S | M \cap N \cap A) \cdot P(N | M) \cdot P(M)}{P(N | M) \cdot P(M) \cdot P(S)} = \frac{P(S | M \cap N \cap A)}{P(S)}$$

Una altra simplificació que podem fer és la següent: substituïm  $P(S | M \cap N \cap A) = \frac{P(N | M \cap S \cap A) \cdot P(S)}{P(N | M)}$  i, aleshores obtenim:

$$LR = \frac{P(N | M \cap S \cap A) \cdot P(S)}{P(N | M) \cdot P(S)} = \frac{P(N | M \cap S \cap A)}{P(N | M)}$$

Els valors de les probabilitats dels nous numeradors i denominadors són més senzilles de calcular que les de la fórmula anterior. I obtenim els següents resultats:

| STR    | LR    | STR     | LR    |
|--------|-------|---------|-------|
| D18S51 | 3'876 | D3S1358 | 2'45  |
| D21S11 | 3'876 | D5S818  | 1'43  |
| FGA    | 3'333 | D7S820  | 1'70  |
| TH01   | 3'333 | D8S1179 | 1'87  |
| VWA    | 5'848 | D18S51  | 3'87  |
| CSF1PO | 1,40  | TPOX    | 2'123 |

Després d'haver fet aquests càlculs per 13 STR's diferents, és a dir, haver calculat la raó de versemblances (per cada STR diferent), calculem la LR total multiplicant les LR de cadascun dels 13 STR's estudiats, i obtenim:

$$LR_T = LR_1 \cdot LR_2 \cdots LR_{13} = 513472'42$$

Finalment, suposant  $P(A) = P(A^c) = 0,5$ , podem trobar la probabilitat de que el senyor sigui pare amb la següent fórmula:

$$PP = \frac{LR_T}{1 + LR_T} = P(A|S \cap M \cap N)$$

En el nostre cas obtenim :

$$PP = \frac{513472'42}{1 + 513472'42} = 0'999998$$

Per tant, podem concloure que la probabilitat de que sigui el pare és del 99'9998%.

### 3.4 Resposta

El nostre mètode ens diu si és o no el pare amb una fiabilitat del 99,9998%.

El senyor que havíem triat **sí** que és el pare del noi.

Per tant, el nostre mètode ens proporciona un resultat fiable.

## 4 Ampliació 1 del problema

En aquest cas considerarem que no coneixem el genoma de la mare. Per resoldre aquest problema seguirem els mateixos passos que hem seguit anteriorment. Per començar hem de considerar els següents esdeveniments:

- N = tenir el genoma del noi (8,13)
- S = tenir el genoma del senyor (8,11)
- A = que el senyor sigui el pare del noi

Primerament, calcularem les probabilitats de tenir aquest genoma:

$$P(S) = 0'057 \quad P(N) = 2 \cdot f_8 \cdot f_{13} = 2 \cdot 0'121 \cdot 0'118 = 0'029$$

A continuació, calcularem les probabilitats condicionades a tenir aquests genomes:

En primer lloc, calcularem  $P(N|S \cap A) = \frac{1}{2} \cdot f_{13} = \frac{1}{2} \cdot 0'118 = 0'059$ .

Seguidament, calculem la implicació contrària, és a dir,  $P(S | N \cap A)$ . Per calcular aquesta probabilitat utilitzarem el Teorema de Bayes, tal que:

$$\begin{aligned} P(S | N \cap A) &= \frac{P(N | S \cap A) \cdot P(S \cap A)}{P(N \cap A)} = \frac{P(N | S \cap A) \cdot P(S) \cdot P(A)}{P(N) \cdot P(A)} \\ &= \frac{P(N | S \cap A) \cdot P(S)}{P(N)} = \frac{0'059 \cdot 0'057}{0'029} = 0'118 \end{aligned}$$

Un cop hem calculat aquestes dues probabilitats, podem calcular la raó de versemblances:

$$LR = \frac{Odds(A | N \cap S)}{Odds(A)} = \frac{P(N \cap S | A)}{P(N \cap S | A^c)}$$

Primerament calculem el valor del numerador:

$$\begin{aligned} P(N \cap S | A) &= \frac{P(N \cap S \cap A)}{P(A)} = \frac{P(N | S \cap A) \cdot P(S \cap A)}{P(A)} = \\ &= P(N | S \cap A) \cdot P(S) = 0,059 \cdot 0,057 = 3,63 \cdot 10^{-3} \end{aligned}$$

Ara calculem el valor del denominador:

$$P(N \cap S | A^c) \simeq P(N \cap S) = P(N) \cdot P(S) = 0,029 \cdot 0,057 = 1,63 \cdot 10^{-3}$$

Per tant, obtenim que:

$$LR = \frac{3,63 \cdot 10^{-3}}{1,63 \cdot 10^{-3}} = 2,23$$

Finalment, repetirem el procés per cada STR de la taula sense tenir en compte el genoma de la mare. I un cop tinguem el  $LR_T$  procedirem a calcular la probabilitat de parentiu.

## 5 Ampliació 2 del problema

Un cop acabat el problema inicial ens vam preguntar com ho faríem amb una relació de parentiu més llunyana, per exemple, per saber la probabilitat de que un senyor sigui l'avi d'un noi.

En aquest cas, les hipòtesis que hem adoptat són que coneixem els genomes de les 2 persones implicades (senyor i noi). I, per tant, hem de tenir en compte els següents esdeveniments:

- N = tenir el genoma del noi
- S = tenir el genoma del senyor
- E = que el senyor és l'avi del noi

Per tal d'aconseguir-ho, haurem de fer els següents passos intermedis:  
 Calculem les probabilitats de tenir les repeticions de cada genoma a la població total:  $P(N)$  i  $P(S)$

A continuació, calculem les probabilitats condicionades en dos sentits, que necessitarem més endavant:

Primerament, calculem la probabilitat de que el noi tingui N genoma sabent el genoma del seu avi utilitzant teoria de genètica:  $P(N|S \cap E)$

Seguidament, amb les fórmules de Bayes calculem la probabilitat de que el senyor tingui S genoma condicionat al genoma del noi (sabent que és el seu nét):

$$P(S|N \cap E) = \frac{P(N|S \cap E)P(S \cap E)}{P(N \cap E)} = \frac{N|S \cap E)P(S)P(E)}{P(N)P(E)}$$

Però la pregunta del nostre problema ens demana la probabilitat de que el senyor amb genoma S sigui l'avi del nét:  $P(E|N \cap S)$

Per trobar-la, procedim a calcular la raó de versemblances (LR):

$$LR = \frac{Odds(E|S \cap NS)}{Odds(E)} = \frac{P(N \cap S|E)}{P(N \cap A|E^c)}$$

Començarem calculant la probabilitat del numerador:

$$\begin{aligned} P(N \cap S | E) &= \frac{P(N \cap S \cap E)}{P(E)} = \frac{P(N | S \cap E) \cdot P(S \cap E)}{P(E)} = \\ &= \frac{P(N | S \cap E) \cdot P(S) \cdot P(E)}{P(E)} = P(N | S \cap E) \cdot P(S) \end{aligned}$$

Seguidament, calculem la probabilitat que es troba al denominador:

$$P(N \cap S|E^c) \simeq P(N \cap S) = P(N) \cdot P(S)$$

Finalment, obtenim que:

$$LR = \frac{P(N|S \cap E)P(S)}{P(N)P(S)} = \frac{P(N|S \cap E)}{P(N)}$$

Fins aquí hem fet els càlculs per un STR. A continuació hauríem de fer els mateixos càlculs per la resta d'STR i trobar les  $LR_i$  per  $i = 1, \dots, 13$  per calcular la raó de versemblances total:

$$LR_T = LR_1 \cdot LR_2 \cdots LR_{13}$$

Finalment, podem trobar la probabilitat de que el senyor sigui l'avi amb la següent fórmula:

$$PP = \frac{LR_T}{1 + LR_T} = P(E|N \cap S)$$

I doncs, podem concloure que la probabilitat de que el senyor sigui l'avi és del  $PP \cdot 100\%$ .

## 6 Possibles refinaments

- Considerar que els STR poden mutar, ja que estan formats per un seguit de nucleòtids i es pot donar el cas en què aquests pateixin alguna mutació.
- No suposar que  $P(A) = 0,5$  al final del problema. Podríem tractar de buscar un valor que s'ajustés més a la probabilitat de ser pare en general.
- En el cas de l'ampliació 2 del problema, per estudiar la relació de paternitat entre un avi i un noi ens basem en els STR del cromosoma sexual, però com a refinament podríem tractar de tenir en compte al·lels qualsevol de l'ADN que no siguin concretament els sexuals.

## 7 Conclusions

Vam començar el treball amb l'objectiu inicial de trobar la probabilitat de què un senyor sigui el pare d'un noi, sabent els seus genomes i el de la mare. Per trobar aquesta probabilitat, hem passat per diverses fases, a mesura que anàvem perfeccionant el nostre mètode.

Inicialment vam proposar-nos analitzar gens aleatoris de cada persona i comparar-los. Però hem trobat que és molt més útil analitzar els STRs (com que s'hereten directament dels progenitors) i hi ha una gran varietat de possibles combinacions d'al·lels a la població. Això implica que sigui menys probable tenir els 13 STRs iguals, per tant, en el cas que 2 persones els tinguin iguals, serà més probable que siguin pare-fill (la probabilitat de què sigui una coincidència és gairebé nul·la).

També creiem en un inici que per arribar al resultat tindríem prou amb aplicar les fórmules de Bayes, però no va ser possible, ja que no es pot calcular  $P(A)$  (probabilitat de què sigui pare) i ens era impossible trobar el resultat de la probabilitat final de ser pare sense utilitzar aquesta dada. Vam trobar el concepte de la Odds ratio i el mètode de la raó de versemblança i va ser

un recurs necessari per trobar la solució del nostre problema.

Hem conclòs que sí que hem arribat al resultat esperat, ja que hem obtingut un 99,9998% de probabilitat de què sigui el pare i les dades inicials mostraven que sí que ho era. Per tant, el nostre mètode ha funcionat.

Hem comprovat què passaria si agafem un senyor que no és pare i efectivament, hem obtingut una probabilitat del 0% de què aquest sigui pare. Com a mínim, per 1 dels 13 STR ( $STR_i$ ), els 2 al·lells del fill són diferents dels dos del pare, per tant és impossible que el pare li passi qualsevol dels 2 al·lells al fill. Aleshores, per aquest STR,  $P(N|M \cap S \cap A) = 0$ , i doncs,  $LR_i = 0$  i finalment la raó de versemblances total és 0:  $LR_T = 0$  i  $PP = 0/1 = 0$ .

Per tant hem trobat un mètode que amb el genoma del fill, de la mare i el d'un possible pare (senyor), mitjançant càlculs de probabilitat i comprensió del funcionament de la genètica, ens proporciona la resposta a si **aquest senyor és o no el pare** amb un 99'9998% de fiabilitat.

També cal fer menció a les dues ampliacions del problema. En l'ampliació 1 hem vist que, tot i que disposem de menys dades inicials, el mètode utilitzat és anàleg al del problema i ens permet donar resposta a un problema diferent del que plantejem en un inici. Hem obtingut un resultat amb menys pes que si tenim en compte el genoma de la mare, per tant és coherent. Finalment, en l'ampliació 2, hem de tenir en compte quina zona de l'ADN hem estudiem per obtenir un resultat verídic, ja que amb STR qualsevol no hi ha certesa de què el nét hereti algun al·lel de l'avi. Concretament estudiem el cromosoma Y (línia paterna) i l'ADN-mitocondrial (línia materna).

## Referències

- [1] Arthur J. Eisenberg *Parentage statistics*
- [2] BioNinja. *DNA Profiling*
- [3] Cosmolinux. *Els experiments i les lleis de Mendel*
- [4] El proyecto biológico. *¿Qué es un polimorfismo de repeticiones cortas en tándem (STR)?*
- [5] Jaime Cerda, Claudio Vera, Gabriel Rada *Odds ratio: Theoretical and practical issues*
- [6] J. Forensic Sci. *Allele Frequencies for 15 Autosomal STR Loci on U.S. Caucasian, African American, and Hispanic Populations*
- [7] Maria Francisca Bestard Vachiano. *STRs en la investigació biològica de la paternitat: interpretació i valoració dels resultats*. Universitat de les Illes Balears.
- [8] Rosario Delgado. *Xarxes bayesianes: una metodologia per avaluar riscos*
- [9] Sánchez García M., Sobrón Fernández M.I. *Calculo de probabilidades de consistencia y paternidad en sistema geneticos: estudio de dos modelos*
- [10] Wikipedia. *Microsatélite*