

PADN

Parentiu i ADN

Taller de modelització
Anna Aumatell, Alícia Chimeno, Irene de Merlo, Laia Lluís

June 12, 2022

INTRODUCCIÓ: 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 bastanta confiança les relacions de parentiu més properes**. Com es pot fer això?



Problema de paternitat: reconèixer amb confiança el pare biològic d'una persona.

TEORIA (STRs)

STR:

- Seqüències curtes d'ADN, que es repeteixen de manera consecutiva.
- Donen lloc a una elevada variabilitat entre els ADN de les persones que pot ser útil per propòsits d'identificació.
- Tenen una herència mendeliana simple, és a dir, un al·lel és heretat del pare i l'altre de la mare.

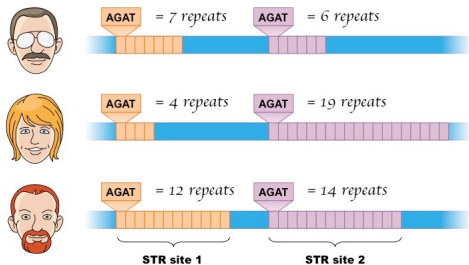


Figure: [2]

TEORIA (freqüència)

Freqüència:

- S'analitzen 13 STRs.
- Per cada STR hi ha una freqüència o probabilitat associada.

Per calcular la probabilitat de tenir una determinada freqüència d'un al·lel s'ha de tenir en compte si és homozigot $X=(x,x)$ o heterozigot $Y=(x,y)$:

$$\text{Homozigot} \rightarrow P(X) = f_x^2$$

$$\text{Heterozigot} \rightarrow P(Y) = f_x * f_y + f_y * f_x = 2 * f_x * f_y$$

TEORIA (probabilitat)

Teorema de Bayes:

$$P(X | Y) = \frac{P(X \cap Y)}{P(Y)} = \frac{P(Y | X) \cdot P(X)}{P(Y)}$$

Simplificació de la intersecció:

$$P(X_1 \cap X_2 \dots \cap X_n) = P(X_1) \cdot \dots \cdot P(X_n)$$

sent X_i esdeveniments independents.

TEORIA (Odds ratio)

Odds d'un esdeveniment (odds a priori):

$$Odds(A) = \frac{P(A)}{P(A^c)}$$

Odds d'un esdeveniment (odds a posteriori):

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

Odds ratio:

$$\text{Odds ratio} = \frac{\text{odds a posteriori}}{\text{odds a priori}}$$

TEORIA (Raó de versemblances)

Concepció

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

Raó de versemblances:

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

Interpretació:

$$\left\{ \begin{array}{l} = 1 \rightarrow \text{l'evidència } E \text{ no dona suport a } A \text{ ni al seu contrari.} \\ > 1 \rightarrow \text{l'evidència } E \text{ dona suport a } A. \\ < 1 \rightarrow \text{l'evidència } E \text{ dona suport al contrari d}'A \text{ (} A^c \text{).} \end{array} \right. \quad (1)$$

EXEMPLE SIMPLE

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 amb no sabent-ho.

Aquesta pregunta ens demana clarament la odds ratio de que surti un 5.

Declarem les variables dels esdeveniments:

A = treure un 5 al dau

A^c = no treure un 5 al dau

E = ser senar

Resolució:

Primerament calculem les probabilitats de cada esdeveniment:

$$P(A) = 1/6$$

$$P(A^c) = 5/6$$

$$P(E) = 1/2$$

EXEMPLE SIMPLE

$$P(A|E) = 1/3$$

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

$$\text{Calculem } odds(A) = \frac{P(A)}{P(A^c)} = \frac{1/6}{5/6} = 1/5$$

$$\text{Calculem } odds(A|E) = \frac{P(A|E)}{P(A^c|E)} = \frac{1/3}{2/3} = 1/2$$

Finalment, ara que 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.

CONTEXT DEL PROBLEMA

Volem reconèixer amb confiança el pare biològic d'una persona. Les hipòtesis seràn 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
THO1	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

Figure: dades extretes de la font [1]

PROBLEMA (part 1)

Per un STR concret : D13S317.

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

Objectiu: $P(A|N \cap M \cap S)$

PROBLEMA

Càlcul de la probabilitat de tenir aquest genoma:

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

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

Calculem les probabilitats condicionades que necessitem més endavant:

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

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

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

PROBLEMA (part 2)

Calculem la raó de versemblances (LR):

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

Numerador:

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

PROBLEMA

Denominador:

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

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

Aquest resultat ens diu que l'odds condicionat (a posteriori) és 4,11 vegades més gran que l'odds sense condicionar (a priori).

PROBLEMA

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%.

AMPLIACIÓ PROBLEMA

Resolem el problema sense tenir en compte l'esdeveniment M.

STR concret : D13S317. Esdeveniments:

- N = tenir el genoma del noi (8,13) $P(N) = 0'029$
- S = tenir el genoma del senyor (8,11) $P(S) = 0'057$
- A = que el senyor sigui el pare del noi

Analogament, a l'apartat anterior:

$$P(N|S \cap A) = \frac{1}{2} \cdot f_{13} = \frac{1}{2} \cdot 0'118 = 0'059$$

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

AMPLIACIÓ PROBLEMA

Calculem la raó de versemblances:

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

Numerador:

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

Denominador:

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

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

CONCLUSIONS I INTERPRETACIÓ DELS RESULTATS

- Recapitulació del treball
- Efectivament hem arribat al resultat esperat, ja que el genoma del senyor haviem "triat" que fos el mateix que el del pare. Per tant, el nostre mètode ha funcionat.

POSSIBLES REFINAMENTS

- Considerar que els STR poden mutar.
- No suposar que $P(A) = 0,5$.
- Resoldre el mateix problema però amb relacions de parentiu menys properes.

DEMOSTRACIÓ RAÓ DE VERSEMBLANÇA

$$\text{Raó de versemblança} = \text{LR} = \frac{\text{odds}(A|E)}{\text{odds}(A)}$$

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

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

$$\text{com que } P(A) = \frac{P(A|E)P(E)}{P(E|A)} \text{ i } P(A^c) = \frac{P(A^c|E)P(E)}{P(E|A^c)}$$

$$\text{Veiem: } \frac{P(A)}{P(A^c)} = \frac{P(A|E)P(E)P(E|A^c)}{P(E|A)P(A^c|E)P(E)}$$

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

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

DEMOSTRACIÓ ÚLTIMA FÓRMULA

Com hem vist anteriorment la probabilitat de paternitat la calculavem a partir de:

$$PP = \frac{LR}{1 + LR}$$

Suposem que $P(A) = P(A^c) = 0.5$ i $E = M \cap N \cap S$. Llavors podem fer:

$$\begin{aligned} \frac{LR}{1 + LR} &= \frac{\frac{P(E|A) \cdot P(A)}{P(E|A^c) \cdot P(A^c)}}{1 + \frac{P(E|A) \cdot P(A)}{P(E|A^c) \cdot P(A^c)}} = \frac{\frac{P(E \cap A)}{P(E \cap A^c)}}{1 + \frac{P(E \cap A)}{P(E \cap A^c)}} = \frac{P(E \cap A)}{P(E \cap A) + P(E \cap A^c)} = \frac{P(E \cap A)}{P(E)} = P(A|E) \\ &\Rightarrow \frac{LR}{1 + LR} = P(A|E) \end{aligned}$$

D'aquesta manera trobem la probabilitat de que sigui el pare condicionat a E.



Maria Francisca Bestard Vachiano. *STRs en la investigació biològica de la paternitat: interpretació i valoració dels resultats*. Universitat de les Illes Balears.



BioNinja. *DNA Profiling*



Rosario Delgado. *Xarxes bayesianes: una metodologia per avaluar riscos*