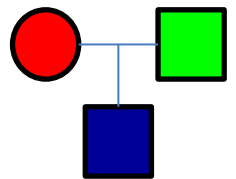




Comparaciones AM-PM

Repaso al Teorema de Bayes. Aproximaciones para cálculo de probabilidades a posterior: *one to one*, *PM-driven* y *AM-driven*

Lourdes Prieto



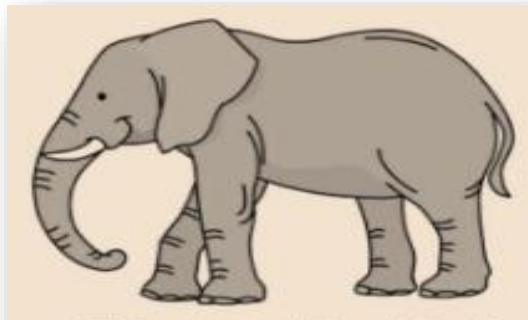
LR y Probabilidades a posteriori

$$LR = \frac{P(E|H_1)}{P(E|H_2)}$$

$$P(\text{orb a posteriori}) = P(H_1|E)$$

$$P(\text{orb a posteriori}) = P(H_2|E)$$

En las probabilidades condicionadas, el orden de los términos es crucial, órdenes diferentes resultan en valores diferentes



$P(E|H) = P(\text{tener 4 patas} \mid \text{soy un elefante})$



$P(H|E) = P(\text{ser un elefante} \mid \text{tengo 4 patas})$

La probabilidad de obtener un resultado genético si una hipótesis es cierta ($P(E|H)$) es diferente a la probabilidad de la hipótesis teniendo en cuenta los resultados genéticos ($P(H|E)$)

Repaso: Teorema de Bayes

- Un match NO es una identificación. Hay que integrar los resultados de ADN con los resultados obtenidos en otras disciplinas forenses
- El teorema de Bayes nos permite combinar resultados genéticos con otra evidencia no genética
- Ejemplos de evidencia no genética: dónde se encontraron los cuerpos, características antropológicas, ropa y otra evidencia asociada, documentos, etc.
- Toda esta evidencia no genética nos puede ayudar a identificar a los fallecidos.
- La evidencia no genética se puede tener en cuenta mediante la estima del valor de una probabilidad de identificación llamada probabilidad *a priori*.

Teorema de Bayes para 2 hipótesis

En forma de apuesta (*odds form*), dos hipótesis (mutuamente excluyentes) – La prob. *a priori* se relaciona con el LR de esta forma:

$$\text{Posterior odds} = \text{Prior odds} \times \text{LR}$$

Datos genéticos

Datos no genéticos

$$\text{Posterior odds} = P(H_1|E)/P(H_2|E) \quad \text{Prior odds} = P(H_1)/P(H_2)$$



En forma de probabilidad (*probability form*), dos hipótesis (mutuamente excluyentes) – La prob. *a priori* se relaciona con el LR de esta forma:

$$P(H_1|E) = \frac{P(E|H_1)P(H_1)}{P(E|H_1)P(H_1) + P(E|H_2)P(H_2)}$$

Datos genéticos

Datos no genéticos

$$P(H_1|E) + P(H_2|E) = 1$$

Las probabilidades *a priori* tienen una gran influencia en la probabilidad final de identificación (prob. *a posteriori*)

Marker	Indiv. 1	Indiv. 2
D3	15-17	15-17
VWA	15-18	15-18
D16	10-10	10-13
D8	14-14	8-14
SE33	19-30.2	19-24.2
D10	12-13	12-15
D2	20-21	18-21
D13	11-11	11-12
FGA	22-25	22-23
THO1	7-9.3	8-9.3
D19	13-14	13-14
D18	16-20	17-20
D21	20-30	30-32.2

¿Estos individuos pueden ser un dúo padre-hijo?

$$\text{Posterior odds} = \text{Prior odds} \times LR$$

Información no genética:

Indiv. 1 tiene 3 años

Indiv. 2 tiene 1 año

Prior prob. = 0

¿Cómo se puede estimar la probabilidad *a priori*?

EN ESCENARIOS CERRADOS

El número de personas desaparecidas (MP) se puede cuantificar (ej., accidente aéreo*), de forma que la probabilidad *a priori* se puede basar en el número total de personas desaparecidas

Prob. *a priori* = $1 / MP$, si no hay información adicional y tratamos a todas las MP igual.

También $1 / (MP + 1)$ si queremos tener en cuenta que la MP que evaluamos no está en este escenario

*La lista de pasajeros puede no ser exacta.

EN ESCENARIOS ABIERTOS

El número de MP es sólo una estima, con menos certeza que en el caso de escenarios cerrados, pero en muchos casos se puede estimar, aunque sea aproximadamente

Ejemplo: Ataque terrorista WTC: se determinó que algo menos de 3000 personas murieron. La prob. *a priori* elegida fue $1/3000$

¿Cómo se puede estimar la prob. *a priori*?

En escenarios de guerra o genocidios, los restos humanos no se encuentran en un sólo sitio, sino que están en múltiples localizaciones.

Algunos labs usan “datos geográficos” para establecer la prob. *a priori* (ej., n° de individuos desaparecidos en un área específica y en un período de tiempo específico). Esto no refleja la posibilidad de que los cuerpos se trasladen de una región a otra.

La probabilidad *a priori* se puede re-definir teniendo en cuenta otros meta datos:

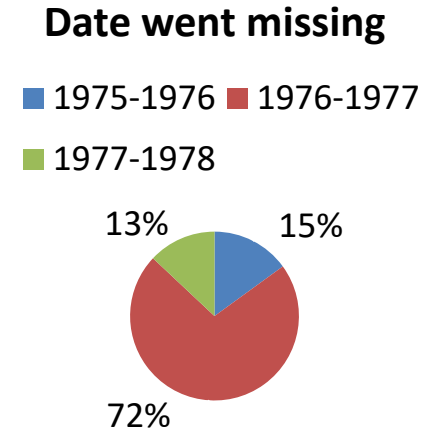
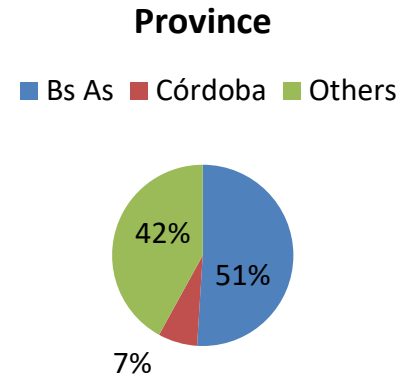
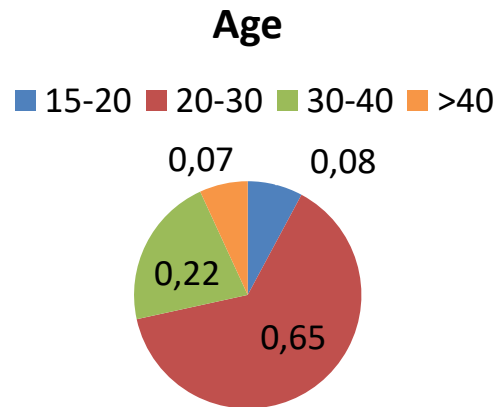
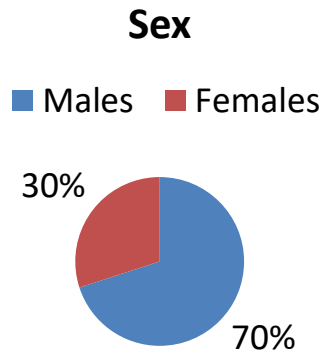
- Características antropológicas (género, edad, prótesis, etc.)
- Otra info: ropas, documentos, etc.

Esto puede aumentar la probabilidad *a priori*

La probabilidad *a priori también* se puede re-definir a medida que las identificaciones progresan y el número de desaparecidos decrece (dando por ciertas las identificaciones realizadas)

Ejemplo: Argentina

Escenario con 10.000 MP: *prior* = 1 / 10.000



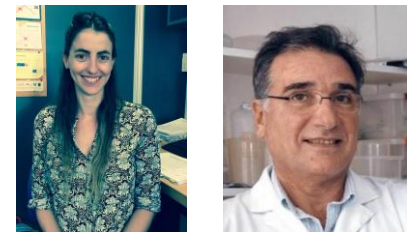
PM (ejemplo)

- Cuerpo hallado en Córdoba (0.07)
- Mujer (0.3)
- Edad 20-30 (0.65)
- ¿Fecha desaparición? (1)

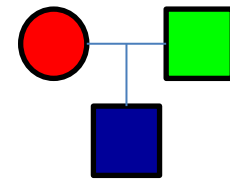


Metadatos combinados

- $10000 \times 0.07 \times 0.3 \times 0.65 \times 1 = 136$
- $\text{Prior} = 1 / 136$



Cómo se define la probabilidad *a priori* en Familias



Se puede definir en la ventana “Add unidentified person(s) (PM)”

#Missing persons
3 Use #PM Number of unidentified remains: 3

ID	Gender	DNA data
V1	Male	Marker 1: 2, 3;
V2	Male	Marker 1: 1, 2;
V3	Male	Marker 1: 4, 4;

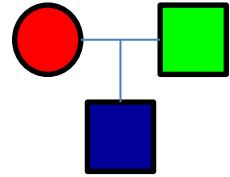
- Por defecto (use #PM), Familias tiene en cuenta el número de muestras no identificadas (3 en el ejemplo) y añade la posibilidad de que la persona desaparecida no se haya encontrado: $1 / 4 = 0.25$

#Missing persons
4 Use #PM Number of unidentified remains: 3

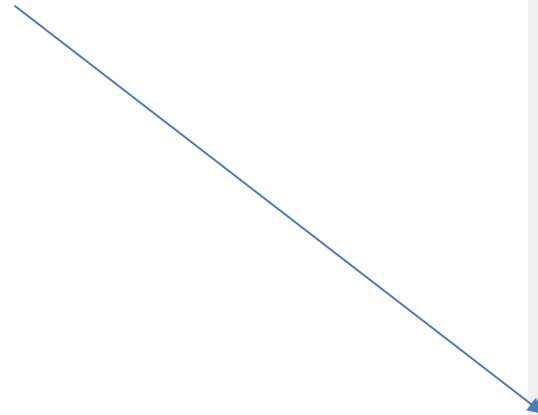
ID	Gender	DNA data
V1	Male	Marker 1: 2, 3;
V2	Male	Marker 1: 1, 2;
V3	Male	Marker 1: 4, 4;

- También se puede desactivar la casilla “use #PM” y añadir el número deseado. Por ejemplo, si en este caso se sabe hay 4 víctimas (aunque sólo se encontrara a 3), podemos añadir “4” en la casilla. Así, el prior será $1/5 = 0.20$, para cada víctima.

Cómo se define la probabilidad *a priori* en Familias



Y también en la ventana "Results"



Search

Search

Quick scan

Sort

Apply threshold

Display

Result options

Posterior model

AM driven ▼

#Missing persons

5 ?

Probabilidad *a posteriori*: UMBRALES

- Cada lab debe decidir su umbral mínimo para considerar que un cuerpo puede estar identificado. Ejemplos:
 - 99.90%
 - 99.95%
 - 99.99%
- El umbral elegido es política del lab, basado en el error que el lab está dispuesto a asumir. Se ha de encontrar un equilibrio entre maximizar el número de identificaciones reales y minimizar las falsas identificaciones
- Muchos labs usan el mismo umbral para re-asociar e identificar.

Cómo influye la prob. *a priori* en la prob. *a posteriori*

	LR				
MP	100	1000	10,000	100,000	1,000,000
10	91.7431%	99.108%	99.9101%	99.991%	99.9991%
100	50.2513%	90.9918%	99.0197%	99.9011%	99.9901%
500	16.6945%	66.7111%	95.2472%	99.5035%	99.9501%
1000	9.0992%	50.025%	90.9174%	99.0109%	99.9002%
2000	4.7642%	33.3444%	83.3403%	98.0402%	99.8005%
5000	1.9612%	16.6694%	66.6711%	95.239%	99.5026%

100 MP, posterior 99.90%: se debe alcanzar un valor de LR de al menos 100,000

Teorema de Bayes para j hipótesis

- Dos hipótesis

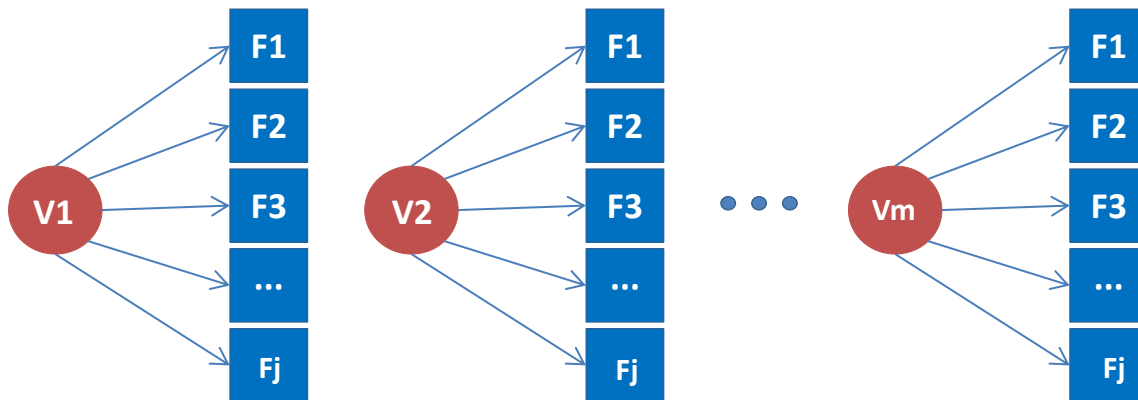


$H_1 = V_1$ pertenece a F_1

$H_2 = V_1$ no está relacionado con F_1

$$P(H_1|E) = \frac{P(E|H_1) P(H_1)}{P(E|H_1) P(H_1) + P(E|H_2) P(H_2)}$$

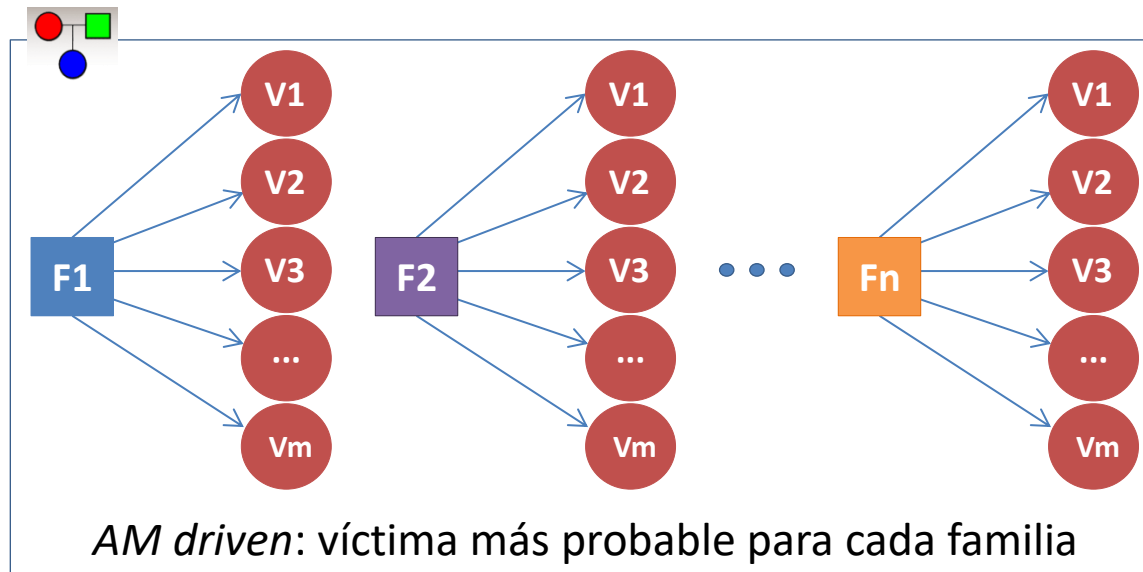
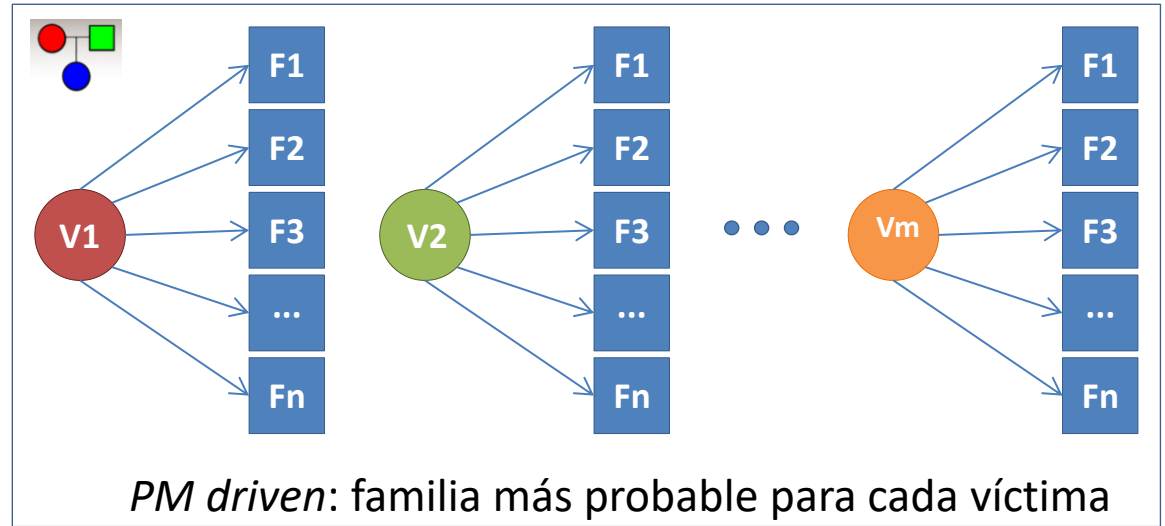
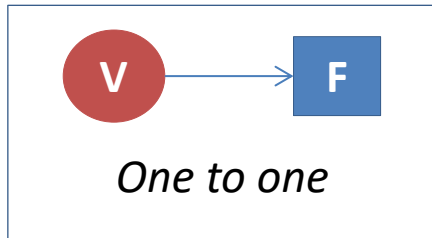
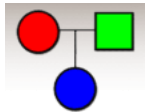
- j hipótesis



$$P(H_i|E) = \frac{P(E|H_i) P(H_i)}{\sum_{j=1}^k P(E|H_j) P(H_j)}$$

Hay diferentes abordajes para calcular posteriors con j hipótesis. Depende del número de hipótesis que queramos considerar y de cómo definamos esas hipótesis

Familias - Casos Masivos: 3 abordajes para estimar *posteriors*



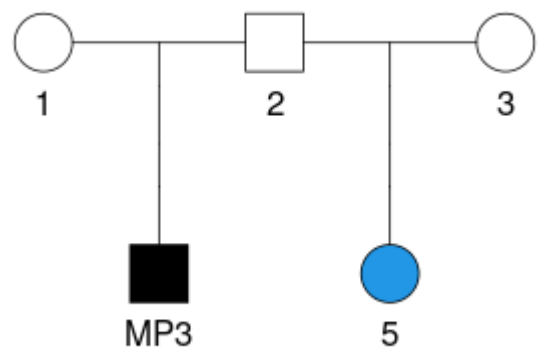
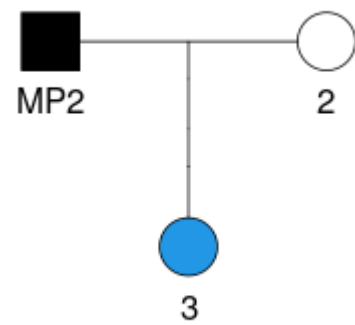
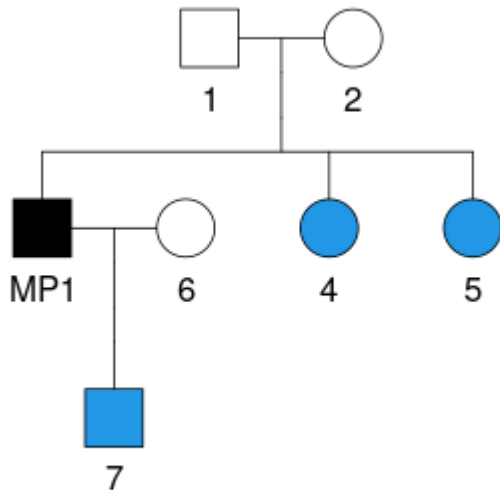
Probabilidades *a posteriori*: ejemplo motivador

V1

V2

V3

Escenario con 3 víctimas y 3 MP



Suposición:
V1 = MP1
V2 = MP2
V3 = MP3

Resultados de las comparaciones AM-PM

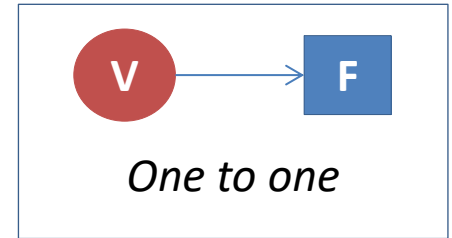
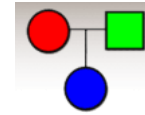
Family	Victim	LR
F1	V1	1 000 000
F2	V1	0
F3	V1	0
F1	V2	0
F2	V2	100 000
F3	V2	0
F1	V3	0
F2	V3	0
F3	V3	500

Situación ideal
Sólo un match por F y por V

Family	Victim	LR
F1	V1	1 000 000
F2	V1	0
F3	V1	100
F1	V2	0
F2	V2	100 000
F3	V2	10
F1	V3	0
F2	V3	10
F3	V3	500

Situación real sencilla
Más de un match por F y/o V

Abordaje *One to one*



Sólo se tiene en cuenta **uno** de los *matches* para cada víctima

No hay problema si los *matches* falsos se pueden distinguir de los *matches* reales (ej., a través de información no genética)

Family	Victim	LR
F1	V1	1 000 000
F2	V1	0
F3	V1	100
F1	V2	0
F2	V2	100 000
F3	V2	10
F1	V3	0
F2	V3	10
F3	V3	500

$H_1 = V1$ es MP1

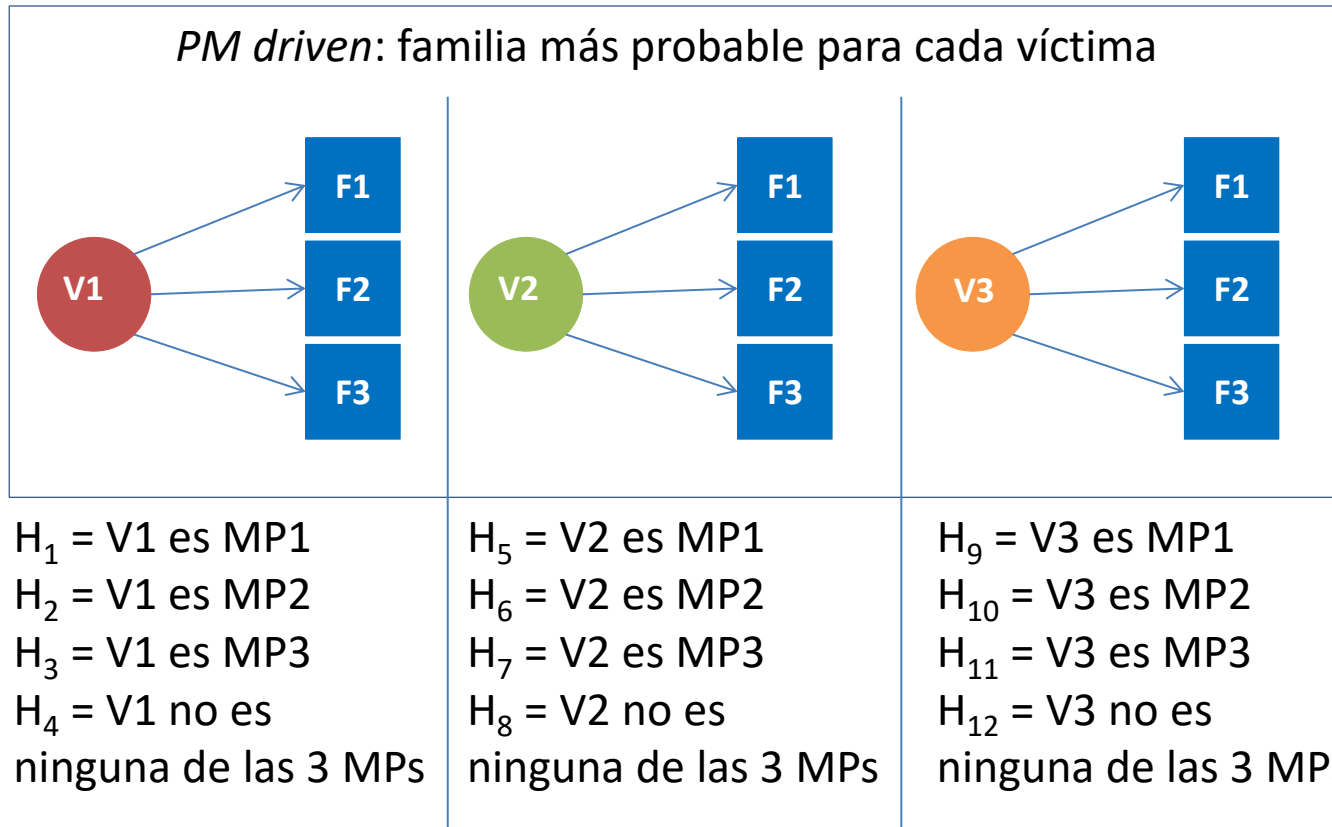
$H_2 = V1$ no está relacionada con MP1

Si la prob. a posteriori de que V1 sea MP1 ($P(H_1|E)$) es, por ejemplo, del con 99%, V1 no está relacionada con MP1 $P(H_2|E)$ con 1% de probabilidad *a posteriori*

$$P(H_1|E) = 0.99$$

$$P(H_2|E) = 0.01$$

PM-driven



Abordaje PM-driven

	Family	Victim	LR
--	--------	--------	----

H ₁	F1	V1	1 000 000
H ₂	F2	V1	0
H ₃	F3	V1	100
H ₄	*	V1	1

H ₅	F1	V2	0
H ₆	F2	V2	100 000
H ₇	F3	V2	10
H ₈	*	V2	1
H ₉	F1	V3	0
H ₁₀	F2	V3	10
H ₁₁	F3	V3	500
H ₁₂	*	V3	1

Se tienen en cuenta todos los *matches* para cada víctima

H₁ = V1 es MP1

H₂ = V1 es MP2

H₃ = V1 es MP3

H₄ = V1 no es ninguna de las 3 MPs

Las probabilidades a posteriori de los *matches* para cada V suman 1

$$P(H_1|E) = 0.99989$$

$$P(H_2|E) = 0$$

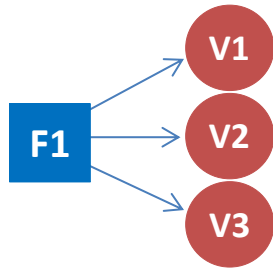
$$P(H_3|E) = 0.0000999$$

$$P(H_4|E) = 0.0000009998$$

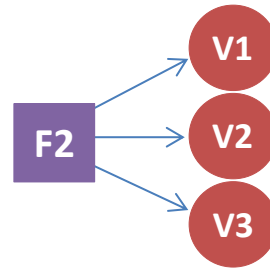
} = 1

AM-driven

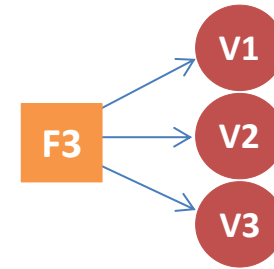
AM driven: víctima más probable para cada familia



$H_1 = \text{MP1 es V1}$
 $H_2 = \text{MP1 es V2}$
 $H_3 = \text{MP1 es V3}$
 $H_4 = \text{MP1 no es ninguna de las 3 Vs}$



$H_5 = \text{MP2 es V1}$
 $H_6 = \text{MP2 es V2}$
 $H_7 = \text{MP2 es V3}$
 $H_8 = \text{MP2 no es ninguna de las 3 Vs}$



$H_9 = \text{MP3 es V1}$
 $H_{10} = \text{MP3 es V2}$
 $H_{11} = \text{MP3 es V3}$
 $H_{12} = \text{MP3 no es ninguna de las 3 Vs}$

AM-driven

	Family	Victim	LR
H ₁	F1	V1	1 000 000
	F2	V1	0
	F3	V1	100
H ₂	F1	V2	0
	F2	V2	100 000
	F3	V2	10
H ₃	F1	V3	0
	F2	V3	10
	F3	V3	500
H ₄	F1	*	1
	F2	*	1
	F3	*	1

Se tienen en cuenta todos los *matches* para cada familia

H₁ = MP1 es V1

H₂ = MP1 es V2

H₃ = MP1 es V3

H₄ = MP1 no es ninguna de las 3 Vs

Las probabilidades a posteriori de los *matches* para cada F suman 1

$$P(H_1|E) = 0.999999$$

$$P(H_2|E) = 0$$

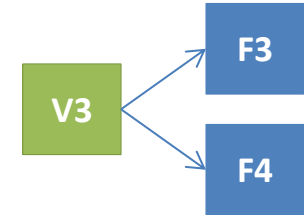
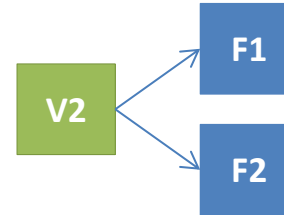
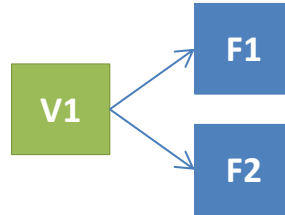
$$P(H_3|E) = 0$$

$$P(H_4|E) = 0.000001$$

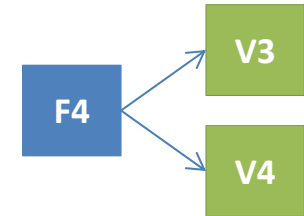
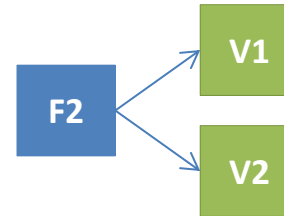
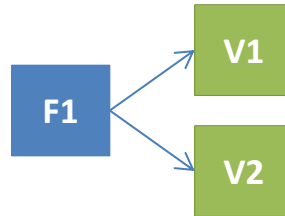
} = 1

Problemas de estos abordajes

Una Víctima asocia con más de 1 Familia



Una Familia asocia con más de 1 Víctima



Posteriors pueden ser irreales porque se reparten hasta sumar 1

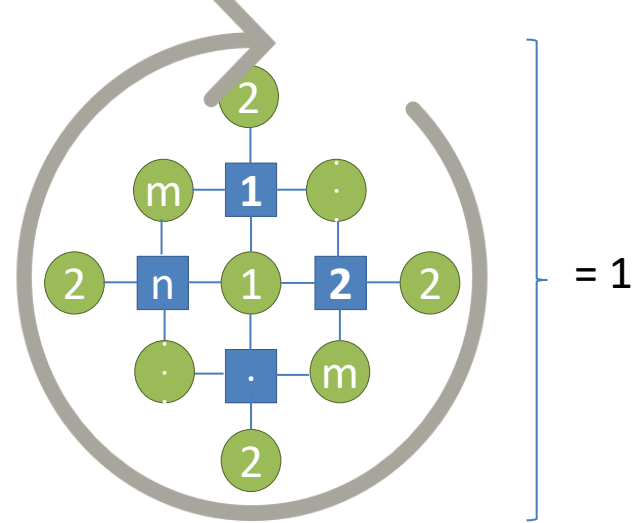
Famil...	Uni...	Prior	Posterior	LR
F01	V1	0.090909	0.999986	8.0616655e+022
F01	V2	0.090909	1.44248e-005	1.1628946e+018
F02	V1	0.090909	0.999986	8.0616655e+022
F02	V2	0.090909	1.44248e-005	1.1628946e+018
F03	V3	0.090909	>0.999999	2.971482e+008
F04	V3	0.090909	0.666667	1.2503518e+010
F04	V4	0.090909	0.333333	6.2517588e+009

Habitual en escenarios con víctimas emparentadas

One to one

Family	Victim	LR
F1	V1	1 000 000
F2	V1	0

Posteriors = 1



Joint approach

PM driven

	Family	Victim	LR
H ₁	F1	V1	1 000 000
H ₂	F2	V1	0
H ₃	F3	V1	100
H ₄	*	V1	1

= 1

	Family	Victim	LR
H ₅	F1	V2	0
H ₆	F2	V2	100 000
H ₇	F3	V2	10
H ₈	*	V2	1

= 1

...

AM driven

	Family	Victim	LR
H ₁	F1	V1	1 000 000
H ₂	F1	V2	0
H ₃	F1	V3	0
H ₄	F1	*	1

= 1

	Family	Victim	LR
H ₁	F2	V1	0
H ₂	F2	V2	1 000 000
H ₃	F2	V3	10
H ₄	F2	*	1

= 1

...