

# La probabilidad de exclusión en los marcadores genéticos

Eliseo Martínez<sup>1</sup>, Héctor Varela<sup>1</sup>

## Abstract

En este artículo se entrega un algoritmo para el cálculo de la probabilidad de exclusión en un número  $k$  de marcadores genéticos polimorfos de un total de  $m$  marcadores utilizados para un problema de paternidad.

## 1. Resultados preliminares

Sea  $M = \{f_1, 2, \dots, m\}$  y definamos para cada  $k = 1, \dots, m$

$$!_k = \{f_1, \dots, i_j, \dots, i_k\} \subset M; i_j \in !_k \text{ si } j \in \{1, \dots, k\} \quad (1)$$

Es fácil notar que  $!_k$  es la colección de todas las muestras de  $f_1, 2, \dots, m$ . Luego cada elemento de  $!_k$  se puede expresar como

$$!_k = \{f_1, \dots, i_j, \dots, i_k\} \subset M; 1 \leq i_1 < i_2 < \dots < i_k \leq m \quad (2)$$

donde  $\{f_1, \dots, i_j, \dots, i_k\}$  de  $!_k$  según (2) es una muestra de  $k$  elementos de  $M$  que está ordenada. Por lo demás

$$|!_k| = \binom{m}{k}$$

Para  $m \geq k$ , el mayor elemento de  $M$ , la colección  $!_k$  admite una partición de la forma

$$!_k = !_k^{(1)} \cup \dots \cup !_k^{(n)} \quad (3)$$

donde

$$!_k^{(1)} = \{f_1, \dots, i_k\} \subset M; i_1 < i_2 < \dots < i_k; i_k \in m$$

$$!_k^{(n)} = \{f_1, \dots, i_k\} \subset M; i_1 < i_2 < \dots < i_k; i_k = m$$

Podemos observar que  $!_k^{(1)} = m - k + 1$ ;  $!_k^{(n)} = k$ .

Supongamos ahora que tenemos una colección de números  $\{x_1, x_2, \dots, x_n\}$  con  $0 < x_i < 1$ , para  $i = 1, \dots, n$ . Definamos para cada  $k \in \{1, \dots, n\}$  el número

$$P(k) = \prod_{i=1}^k (1 - x_i) \prod_{i=2}^k x_i \quad (4)$$

<sup>1</sup> Departamento de Matemáticas, Universidad de Antofagasta

entendiendo que  $\sum_{i=1}^n p_i = 1$ . Además definamos

$$P(0) = \prod_{i=1}^n p_i \quad (5)$$

**Proposición.**  $fP(k); k = 0; 1; \dots; n$  es una probabilidad sobre  $N_0 = f0; 1; \dots; n$ .

**Demostración.** Es claro que  $P(k) > 0$  para todo  $k = 0; 1; \dots; n$ . Debemos demostrar que  $\sum_{k=0}^n P(k) = 1$ , y esto es equivalente a demostrar que

$$\sum_{k=1}^n \sum_{i=1}^k p_i^{k-i} (1-p_i)^i = \sum_{i=1}^n p_i^{n-i} (1-p_i)^i \quad (6)$$

Lo haremos por inducción sobre  $n$ . Para  $n = 1$ , el resultado es trivial, puesto que

$$\sum_{i=1}^1 p_i^{1-i} (1-p_i)^i = 1-p_1$$

entendiendo que  $1-p_1 = f1g$  para  $N = f1g$ . Supongamos que el resultado es cierto para  $n$ . Debemos probar (6) para  $n + 1$ . Sea  $!_k$  la colección de muestras de tamaño  $k$  obtenidas de  $f1; \dots; n; n + 1g$ . Tenemos que demostrar

$$\sum_{k=1}^{n+1} \sum_{i=1}^k p_i^{k-i} (1-p_i)^i = \sum_{i=1}^{n+1} p_i^{n+1-i} (1-p_i)^i \quad (7)$$

El lado izquierdo de (7) se desarrolla como

$$\begin{aligned} \sum_{k=1}^{n+1} \sum_{i=1}^k p_i^{k-i} (1-p_i)^i &= \sum_{k=1}^n \sum_{i=1}^k p_i^{k-i} (1-p_i)^i + \sum_{i=1}^n p_i^{n+1-i} (1-p_i)^i \\ &= \sum_{k=1}^n \sum_{i=1}^k p_i^{k-i} (1-p_i)^i + \sum_{i=1}^n p_i^{n+1-i} (1-p_i)^i \end{aligned} \quad (8)$$

Ahora desarrollaremos el primer sumando de la última igualdad de (8), esto es

$$\sum_{k=1}^n \sum_{i=1}^k p_i^{k-i} (1-p_i)^i$$

La colección  $!_k$  la particionamos según (3) considerando  $m = n + 1$ . Entonces

$$\begin{aligned} &\sum_{k=1}^n \sum_{i=1}^k p_i^{k-i} (1-p_i)^i \\ &= \sum_{k=1}^n \sum_{i=1}^k p_i^{k-i} (1-p_i)^i + \sum_{k=1}^n \sum_{i=1}^k p_i^{k-i} (1-p_i)^i \end{aligned} \quad (9)$$

Vamos a analizar cada sumando del segundo miembro de (9). Consideremos el primer

sumando

$$\prod_{k=1}^n \frac{X_k}{2!} \prod_{i=1}^c \frac{Y_i}{2!}$$

Tenemos que,

$$\begin{aligned} \prod_{k=1}^n \frac{X_k}{2!} \prod_{i=1}^c \frac{Y_i}{2!} &= \prod_{k=1}^n \frac{X_k}{2!} \prod_{i=1}^c \frac{Y_i}{2!} \prod_{i=n+1}^c \frac{Y_i}{2!} \quad (10) \\ &= \prod_{i=n+1}^c \frac{Y_i}{2!} \prod_{k=1}^n \frac{X_k}{2!} \prod_{i=1}^c \frac{Y_i}{2!} \\ &= \prod_{i=n+1}^c \frac{Y_i}{2!} \prod_{i=1}^c \frac{Y_i}{2!} \end{aligned}$$

La última expresión se consigue por la hipótesis de inducción para n.

Trabajaremos ahora con el segundo sumando de (9), esto es

$$\prod_{k=1}^n \frac{X_k}{2!} \prod_{i=1}^c \frac{Y_i}{2!}$$

Tenemos el siguiente desarrollo,

$$\begin{aligned} \prod_{k=1}^n \frac{X_k}{2!} \prod_{i=1}^c \frac{Y_i}{2!} &= \prod_{i=1}^c \frac{Y_i}{2!} \prod_{k=1}^n \frac{X_k}{2!} \prod_{i=1}^c \frac{Y_i}{2!} \quad (11) \\ &= \prod_{i=1}^c \frac{Y_i}{2!} \prod_{k=2}^n \frac{X_k}{2!} \prod_{i=1}^c \frac{Y_i}{2!} \end{aligned}$$

y podemos ver que el primer sumando del miembro derecho de (11) es

$$\prod_{i=1}^c \frac{Y_i}{2!} \prod_{i=1}^c \frac{Y_i}{2!} \quad (12)$$

y el segundo sumando tiene el desarrollo

$$\begin{aligned} \prod_{k=2}^n \frac{X_k}{2!} \prod_{i=1}^c \frac{Y_i}{2!} &= \prod_{k=2}^n \frac{X_k}{2!} \prod_{i=1}^c \frac{Y_i}{2!} \prod_{i=n+1}^c \frac{Y_i}{2!} \quad (13) \\ &= \prod_{i=n+1}^c \frac{Y_i}{2!} \prod_{k=2}^n \frac{X_k}{2!} \prod_{i=1}^c \frac{Y_i}{2!} \end{aligned}$$

Queda entonces por calcular la expresión

$$\sum_{k=2}^n \sum_{i \in \{1, \dots, n+1\}} \sum_{i=1}^k \binom{k}{i} p_i^{i-1} (1-p_i)^{k-i}$$

Observemos que la suma interior es sobre el conjunto de índices

$$i \in \{1, \dots, n+1\}; i_k = i_1 < i_2 < \dots < i_k; i_k = n + 1$$

y puesto que la productoria no considera a  $n + 1$ , entonces en la suma exterior, que va desde  $k$  hasta  $n$ , falta el término  $\sum_{i=1}^n \binom{n}{i} p_i^{i-1} (1-p_i)^{n-i}$  para satisfacer la hipótesis de inducción, de manera que

$$\sum_{k=2}^n \sum_{i \in \{1, \dots, n+1\}} \sum_{i=1}^k \binom{k}{i} p_i^{i-1} (1-p_i)^{k-i} = \sum_{i=1}^n \sum_{i=1}^i \binom{i}{i} p_i^{i-1} (1-p_i)^{i-i} + \sum_{i=1}^n \sum_{i=1}^i \binom{i}{i} p_i^{i-1} (1-p_i)^{i-i} \quad (14)$$

Resumiendo todos nuestros resultados obtenidos en (8), (10), (12), (13) y (14), tenemos que

$$\sum_{k=1}^n \sum_{i \in \{1, \dots, n+1\}} \sum_{i=1}^k \binom{k}{i} p_i^{i-1} (1-p_i)^{k-i} = \sum_{i=1}^n \binom{n}{i} p_i^{i-1} (1-p_i)^{n-i} + \sum_{i=1}^n \binom{n}{i} p_i^{i-1} (1-p_i)^{n-i} + \sum_{i=1}^n \binom{n}{i} p_i^{i-1} (1-p_i)^{n-i} + \sum_{i=1}^n \binom{n}{i} p_i^{i-1} (1-p_i)^{n-i}$$

de tal manera que

$$\sum_{k=1}^n \sum_{i \in \{1, \dots, n+1\}} \sum_{i=1}^k \binom{k}{i} p_i^{i-1} (1-p_i)^{k-i} = \sum_{i=1}^n \binom{n}{i} p_i^{i-1} (1-p_i)^{n-i}$$

y de esta manera queda demostrado (6) ■

Observemos que la probabilidad  $P(k)$  definida sobre  $N_0 = \{0, 1, \dots, n\}$  mediante (4) y (5), es la probabilidad binomial si  $p_i = p$ , para todo  $i = 1, \dots, n$ .

En particular se tiene que

$$P(1) = \sum_{i=1}^n \binom{n}{i} p_i^{i-1} (1-p_i)^{n-i} \quad (15)$$

## 2. La Probabilidad de exclusión

Sea el loci  $K = \{L_1; L_2; \dots; L_m\}$ , de modo que para cada locus  $L_i$  admite un polimor-



de que sea excluyente en al menos un locus es de

$$1 - P(0) = \sum_{i=0}^{K-1} P(i) \quad (17)$$

El resultado dado en (17) es el que comunmente se entregan en los informes forenses para mostrar la potencia de la prueba a través de los marcadores genéticos relativos al loci  $K = \{L_1; L_2; \dots; L_m\}$  y que se obtiene mediante la frecuencia alélica dadas en (16). Sin embargo, en las normativas de algunos países, como por ejemplo en Chile, se pide que la exclusión ocurra en al menos dos locus para desechar una acusación de paternidad. Luego, como  $P(0) + P(1)$  es la probabilidad de que haya a lo más una exclusión, el complemento  $1 - (P(0) + P(1))$ , denota la probabilidad de que haya al menos dos exclusiones, esto es por (15) se tiene que

$$1 - (P(0) + P(1)) = \sum_{i=0}^{K-1} P(i) + \sum_{i=1}^{K-1} [(1 - P(i)) \sum_{j \in i} P(j)] \quad (18)$$

Y es el valor dado (18) la probabilidad de exclusión que se debe considerar cuando se quiere medir la potencia de la prueba mediante marcadores genéticos bajo la condición de que exista a lo menos dos exclusiones para desechar una acusación.

Ahora si se quiere extender la exclusión a  $k$  marcadores, entonces la probabilidad de exclusión es sencillamente  $\sum_{i=1}^{k-1} P(i)$ .

En algunas legislaciones se excluye a una persona acusada con a lo menos tres marcadores, entonces la probabilidad de exclusión es

$$1 - (P(0) + P(1) + P(2)) = \sum_{i=0}^{K-1} P(i) + \sum_{i=1}^{K-1} [(1 - P(i)) \sum_{j \in i} P(j)] + \sum_{i < j} [(1 - P(i))(1 - P(j)) \sum_{l \in i, l \in j} P(l)] \quad (19)$$

### 3. Conclusión

Por lo general se considera como potencia de la prueba en los marcadores genéticos a la expresión dada en (17), no obstante que se descarta una presunta paternidad si hay exclusión en a lo menos dos (y en algunos casos tres) marcadores genéticos. En consecuencia, la correcta probabilidad de exclusión que se debe calcular es la dada en (18) o en (19), respectivamente.

### 4. Referencia

- [1] Martínez, E; Varela, H. *El teorema de Bayes y los marcadores genéticos*. Comunicación a sesión invitada: Estadística. COMCA 2003. Antofagasta, 2003.