

Cálculo del valor de recombinación que maximiza el LOD Score

Antonio Carvajal-Rodríguez

Departamento de Bioquímica, Genética e Inmunología. Universidad de Vigo, 36310 Vigo, Spain
acraaj@uvigo.es

El LOD score o puntuación LOD (Morton 1955) es un estimador que nos permite detectar ligamiento y estimar la frecuencia de recombinación más probable entre genes en la especie humana. Se calcula como el logaritmo de la razón entre la probabilidad de la descendencia dada la existencia de ligamiento con frecuencia de recombinación R respecto al caso sin ligamiento.

$$Z = \log_{10} \frac{\text{Probabilidad de los genotipos de una familia dada una cierta recombinación } R}{\text{Probabilidad de los mismos genotipos suponiendo segregación independiente}}$$

El análisis por LOD requiere un modelo genético preciso respecto al modo de herencia, penetrancia, frecuencias génicas, etc., lo cual lo hace difícilmente aplicable a enfermedades complejas.

Dada su escala logarítmica es factible sumar los diferentes valores de Z obtenidos para diferentes familias (asumiendo el mismo modelo genético). Nos interesa el valor del parámetro R que maximiza Z . Usualmente se aceptan como evidencia de ligamiento valores de $Z \geq 3$. No es raro ver en problemas y ejemplos e incluso en textos de genética general (por ej. Griffiths p.160, 7ª ed. en Español), cómo para resolver casos sencillos, donde los potenciales recombinantes se identifican perfectamente, se prueban diferentes valores de R y se selecciona aquél que ofrece el máximo valor de Z .

Sin embargo, el tanteo numérico, si bien necesario en casos donde la información genética no es completa, es innecesario cuando disponemos de información completa sobre los genotipos. En este caso la estima del valor de recombinación que maximiza la puntuación LOD se reduce al conteo de recombinantes. Voy a mostrar (ver también Haldane & Smith 1947; Cavalli-Sforza, 1981. p. 854 ed. en Español) porqué esto es cierto para casos de LOD en los que podemos obtener una expresión final del tipo

$$\log_{10} \frac{(1-R)^K R^{N-K}}{p^N} \quad (1)$$

donde N es el número total de descendientes estudiados, $K \leq N$ es el número de no recombinantes cuando asumimos una tasa de recombinación R y p es un valor de probabilidad.

Veremos que:

Dada una expresión del tipo de (1) el valor de R que la maximiza es justamente $(N-K)/N$, esto es, la frecuencia de casos en que hemos asumido un posible evento de recombinación respecto al total N de casos.

En primer lugar, dado un número cualquiera n de familias estudiadas si para cada una de ellas podemos obtener una expresión del tipo de (1) entonces la suma de todas ellas se puede reducir a una expresión del mismo tipo dado que

$$\sum_i^n \log_{10}(x_i) = \log_{10}\left(\prod_i^n x_i\right) \text{ y que } R^{k1} \times R^{k2} = R^{k1+k2}.$$

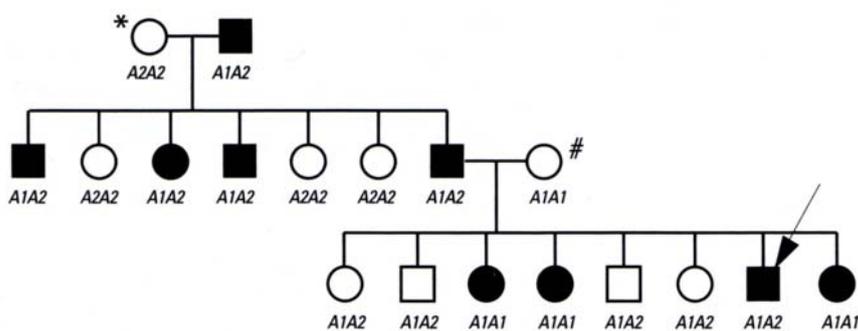
Partiendo entonces de una expresión del tipo de (1) queremos conocer el valor de R que maximiza la expresión. Para ello simplemente debemos calcular la derivada primera respecto a R igualada a 0 y estudiar si el valor de R obtenido es un máximo o un mínimo. Tras derivar (1), es fácil ver que

$$\frac{-KR + (1-R)(N-K)}{R(1-R)\ln 10} = 0 \Rightarrow R = \frac{N-K}{N} \quad (2)$$

Al sustituir el valor obtenido en (2) en la derivada segunda respecto a R la expresión obtenida es negativa para todo valor de $K > 0$ y $R < 0.5$. Por tanto el valor en (2) es un máximo.

Queda entonces demostrado que el valor de recombinación que maximiza el LOD es precisamente la frecuencia respecto al total del número de casos en los que hemos asumido un posible evento de recombinación.

Ejemplo



Familia 1: Lod score $Z = \log_{10} \frac{(1-r)^7 r^0}{(0,5)^7}$

Familia 2: Lod score $Z = \log_{10} \frac{(1-r)^7 r^1}{(0,5)^8}$

Hay un recombinante en 15 descendientes ($(N-K)/N = 1/15 = 0.066$). No es necesario calcular el valor de Z dando diferentes valores a r puesto que sabemos que es precisamente 0.066 el valor que maximiza Z . $Z(0.066) = 2.919$.