

VARIANZA Y COVARIANZA DEL PARENTESCO EN UN LOCUS

García-Cortés¹, L.A., Legarra², A., Chevalet³, C. y Toro⁴, M. A.

¹Departamento de Mejora Genética Animal, INIA, Madrid, Spain. garcia@inia.es

²INRA, UR 631 SAGA, Castanet-Tolosan, France

³INRA, UMR 444 LGC, Castanet-Tolosan, France

⁴Departamento de Producción Animal, Universidad Politécnica de Madrid, Madrid, Spain

INTRODUCCIÓN

El grado de parentesco entre dos individuos mide la proporción del genoma que es idéntico por descendencia y se infiere a partir de la información genealógica. El parentesco es una probabilidad invariable para una determinada relación genealógica. Sin embargo, existe variabilidad en su valor realizado como consecuencia del muestreo mendeliano y el ligamiento. Por ejemplo, dependiendo de la suerte durante el muestreo, dos hermanos completos pueden compartir los dos alelos, uno o incluso ninguno, mientras que un padre y un hijo, en ausencia de consanguinidad, siempre comparten uno y solo uno. Ya se han publicado fórmulas para calcular los momentos de algunos tipos de parentescos (Hill y Weir, 2011; Hill & Weir, 2012) pero no una fórmula general.

Para un locus, el parentesco entre dos individuos es la probabilidad de que dos alelos escogidos al azar, uno por individuo, sean idénticos por descendencia. El coeficiente de fraternidad se define como la probabilidad de que los genotipos de dos individuos (ambos alelos) sean idénticos por descendencia en un cierto locus. El parentesco y la fraternidad realizados son los valores equivalentes observados en cada individuo.

En este trabajo se presenta un método sencillo para calcular varianzas y covarianzas entre parentescos y fraternidades realizadas.

MATERIAL Y MÉTODOS

En esta sección mostramos como obtener las varianzas y covarianzas buscadas a partir de los estados de identidad (Harris, 1964; Gillois, 1964) que se suelen calcular a partir de los coeficientes de parentesco generalizados (Karigl, 1981). Estos estados son bien conocidos y se presentan en la figura 1.

Cada estado indica las relaciones de identidad por descendencia de las cuatro copias después de la segregación acontecida en los padres de ambos individuos, es decir, es una descripción exhaustiva de todos los posibles estados de identidad realizada. Los estados S_i se presentan en cada pareja con una probabilidad Δ_i . Estas probabilidades pueden obtenerse a partir de la relación genealógica de ambos individuos (Karigl, 1981). Por ejemplo, dos hermanos completos no consanguíneos sólo pueden estar en tres estados: $\Delta_7 = 0.25$, $\Delta_8 = 0.5$ y $\Delta_9 = 0.25$, en cambio, todas las parejas padre-hijo están en el estado S_8 con una probabilidad $\Delta_8 = 1$. Ambos parentescos son igual a **0.5** pero por motivos diferentes y con variabilidad realizada diferente.

En general, las 9 probabilidades que caracterizan una relación son parámetros de una distribución categórica (caso particular de la distribución multinomial con una sola observación) y el estado de identidad realizado es un valor muestreado de esta distribución. Si en la notación usamos corchetes de Iverson, los momentos de esta distribución categórica son:

$$E\left(\left[i = k \right] \right) = \Delta_k$$

$$Var\left(\left[i = k \right] \right) = \Delta_k \left(1 - \Delta_k \right) \quad (1)$$

$$\text{Cov}([i = k], [j = m]) = -\Delta_k \Delta_m$$

donde “i=k” significa que el evento “i” está en el estado “k”. En la notación de Iverson $[i = k]$ es 1 si el evento “i” está en el estado “k” y 0 en caso contrario. El parentesco de dos individuos, ϕ_{ij} , puede expresarse a partir de los coeficientes de identidad con la siguiente expresión.

$$\phi_{ij} = \Delta_1 + \frac{1}{2}(\Delta_3 + \Delta_5 + \Delta_7) + \frac{1}{4}\Delta_8$$

La varianza del parentesco se calcula tomando varianzas en la expresión anterior y usando la fórmula (1). Tras un poco de álgebra (García-Cortés et al., 2013) se puede demostrar que

$$\text{var}(\phi_{ij}) = \sum w_k^2 \Delta_k - \sum w_k \Delta_k \sum w_k \Delta_k$$

donde w_i es cada elemento del vector $(1 \ 0 \ 0,5 \ 0 \ 0,5 \ 0 \ 0,5 \ 0,25 \ 0)$. Esta expresión tiene una interpretación muy sencilla usando los parentescos generalizados de Karigl (1981)

$$\text{var}(\phi_{ij}) = \phi_{i,j} - \phi_i \phi_j$$

donde $\phi_{i,j}$ es el parentesco “de doble pareja”. Los detalles acerca de su significado e interpretación pueden consultarse en Karigl (1981).

Esta última fórmula se puede extender para la covarianza entre dos coeficientes de parentesco de esta forma

$$\text{cov}(\phi_{ij}, \phi_{kl}) = \phi_{i,j,kl} - \phi_i \phi_j \phi_k \phi_l$$

También puede extenderse para el cálculo de la varianza de la fraternidad sencillamente modificando los valores de w_i , que en este caso serían $(1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0)$.

RESULTADOS y DISCUSIÓN

A modo de ejemplo, calcularemos la relación entre de Dash 2º y Moll 3º, probablemente los dos perros más relevantes en la fundación de la raza Setter inglés (Laverack, 1872). Su parentesco es el producto de sucesivos cruces entre hermanos completos que, al estar finalmente cruzados con una línea colateral, presenta los 9 coeficientes de identidad distintos de cero. Esta genealogía puede encontrarse en Laverack (1872) o García-Cortés et al (2013).

De Dash 2º y Moll 3º, se calculó la varianza del parentesco usando $w = (1 \ 0 \ 0,5 \ 0 \ 0,5 \ 0 \ 0,5 \ 0,25 \ 0)$, también se calculó la varianza de la fraternidad usando $w = (1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0)$, la de la consanguinidad de Dash 2º usando $w = (1 \ 1 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0)$ y la varianza de la consanguinidad de Moll 3º usando $w = (1 \ 1 \ 0 \ 0 \ 1 \ 1 \ 0 \ 0 \ 0)$. Los resultados se contrastaron con los obtenidos usando una simulación por goteo de genes.

En la Tabla 1 se presentan los cuatro coeficientes (parentesco, fraternidad y las dos consanguinidades) y sus correspondientes varianzas realizadas sobre un locus. Los

resultados son similares a los obtenidos a partir de un algoritmo de goteo de genes convencional.

Es algo bien conocido que la consanguinidad de un individuo es igual al parentesco entre los padres, pero al considerar valores realizados, sólo las esperanzas coinciden, siendo la varianza de la consanguinidad mayor que la varianza del parentesco. Por ejemplo, la varianza del parentesco entre Cora 2º y Sting, padres de Dash 2º, fue de 0.0922, mucho menor que el valor 0.2451 presentado en las tabla 1. Esto se debe a que la consanguinidad de Dash 2º podría considerarse el parentesco entre los padres “realizado” y por tanto acumulando más variación.

El procedimiento presentado puede extenderse a modelos con más de un locus usando una distribución multinomial en la ecuación 1, es decir, dividiendo las varianzas por el número efectivo de loci (Goddard, 2009).

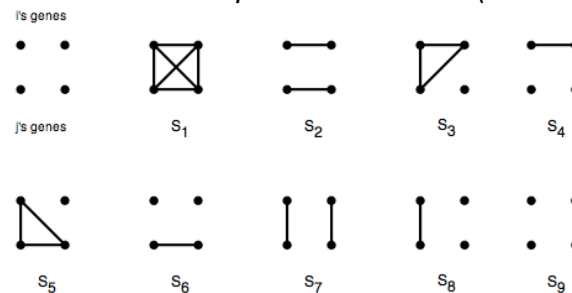
REFERENCIAS BIBLIOGRÁFICAS

- García-Cortés, L.A., Legarra, A., Chevalet, C. & Toro⁴, M. A. 2013. Plos One (en prensa)
- Gillois, M. 1964. Thesis, Faculté des Sciences de Paris
- Goddard, M. 2009. Genetica 136:245-257
- Harris, D.L. 1964. Genetics 50:1319-1348
- Hill, W.G. & Weir, B.S. 2011. Genet Res 93: 47-64
- Hill, W.G. & Weir B.S. 2012. Genet Res 94:267-74
- Karigl, G. 1981. Ann Hum Genet 45:299-305
- Laverack, E. (1872) Longmans, Green and co., London

Tabla 1. Relaciones exactas entre Dash 2º y Moll 3º y sus correspondientes estimaciones usando goteo de genes (100000 muestras).

	Método exacto		Goteo de genes	
	Media	Varianza	Media	Varianza
Parentesco	0.5371	0.0810	0.5367	0.0811
Fraternidad	0.4194	0.2435	0.4169	0.2431
Consanguinidad Dash 2º	0.4297	0.2451	0.4309	0.2452
Consanguinidad Moll 3º	0.4844	0.2498	0.4845	0.2497

Figura 1. Los nueve estados de identidad por descendencia (las líneas marcan identidad).



ABSTRACT

The relationship between pairs of individuals is an important topic in many areas of population and quantitative genetics. It is usually measured as the proportion of the genome identical by descent shared by the pair and it can be inferred from pedigree information. But there is a variance in actual relationships as a consequence of Mendelian sampling, whose general formula has not been developed. The goal of this work is to develop this general formula for the one-locus situation, . We provide simple expressions for the variances and covariances of all actual relationships in an arbitrary complex pedigree. The proposed method relies on the use of the nine identity. Formulas have been checked by computer simulation.

Keywords: Identity coefficients, coancestry, kinship, fraternity