

## ANÁLISIS MATEMÁTICO DE LOS TIEMPOS DE ESPERA BAJO AUTOFECUNDACION EN FUNCION DE LA SELECCION PARA SISTEMAS MULTIALELICOS

N. BARTOLONI (1)

Recibido: 5-10-87

Aceptado: 17-2-88

### RESUMEN

Se ha analizado una variable producto de la aplicación de la teoría de las cadenas finitas de Markov a sistemas genéticos: el tiempo de espera hasta entrada en estado absorbente y algunos de sus parámetros. Se han analizado la media y la variancia de dicha variable en función de valores selectivos genotípicos y alélicos. Se ha empleado la metodología tradicional de descripción de las funciones indicando para cada caso: intercepción, asíntotas, puntos de inflexión y extremos relativos de cada curva. Se han propuesto cinco tipos de herencia para el carácter aptitud sin distinción de viabilidad y fertilidad, suponiendo que los valores selectivos genotípicos dependen de los alelos que componen el genotipo. Este trabajo está dirigido a los genetistas de poblaciones que analizan problemas de naturaleza teórica. Muchas de las consideraciones son de índole estrictamente matemática y no genética.

### MATHEMATICAL ANALYSIS OF THE WAIT TIMES UNDER SELFING WITH SELECTION FOR MULTIALLELIC SYSTEMS

### SUMMARY

It has been analysed the wait time until entrance in an absorbing state and some of its parameters through the application of the Markov chains theory to genetic systems. The parameters analysed were the mean and the variance in relation to genotypic and allelic selective values. It has been employed the usual methodology for the description of the functions indicating in any case, interception, asintotes, points of inflection and relative extreme points to any curve. Five types of inheritance for fitness have been proposed without distinction between viability and fertility supposing that the genotypic values depend upon the allelic ones. This paper is submitted to population geneticists who analyse theoretical aspects. Most of the considerations have a strictly mathematical essence but not a genetical one.

### INTRODUCCION

La estructura de una población genética en el futuro puede predecirse en términos probabilísticos haciendo uso de la teoría. El proceso de la transmisión de la información genética tiene una naturaleza descriptible en

términos de ciertos modelos matemáticos tales como las cadenas markovianas. En el presente trabajo se efectúa un análisis de las relaciones existentes entre ciertos parámetros propios del modelo y la selección genotípica y alélica, en un sistema unigénico multialélico bajo autofecundación.

(1) Cátedra de Genética, Facultad de Agronomía. UBA.

Avda. San Martín 4453. (1417) Buenos Aires -Argentina-

METODOS

Se han considerado valores selectivos genotípicos (los cuales se denotan con los símbolos  $w_{ii}$  para los homocigotas  $A_i A_i$ , y  $w_{ij}$  para los heterocigotas  $A_i A_j$ , siendo  $i \neq j$ ) y alélicos (los cuales se denotan con el símbolo  $w_i$  siendo  $w_i$  el valor selectivo del alelo  $A_i$ ). En diploides se tienen  $n$  genotipos homocigotas distintos y  $n(n-1)/2$  genotipos heterocigotas distintos, siendo  $n$  el número de alelos del gene  $A$ . Se ha utilizado el método de las cadenas finitas de Markov para la obtención de las expresiones explícitas de los parámetros (media y variancia) de la variable "número de generaciones de espera hasta alcanzar un estado absorbente (EA)" en función de los valores selectivos genotípicos y alélicos y, luego se han propuesto cinco casos posibles de herencia del carácter aptitud considerando como fijos los valores selectivos de los homocigotas  $w_{ii} = 2w_i$ . Los valores selectivos variables del heterocigota según el caso serían:

- a)  $w_{ij} = 3w_j - w_i$ ; b)  $w_{ij} = 2w_j$ ;
- c)  $w_{ij} = w_i + w_j$ ; d)  $w_{ij} = 2w_i$ ;
- e)  $w_{ij} = 3w_i - w_j$ .

Dentro de los valores selectivos alélicos se ha considerado la variable  $k_{ij} = w_{ij}/w_i$  de modo tal que se considerará que  $k$  varía sólo en función de la relación existente entre los valores selectivos de los alelos. De este modo, los cinco casos, serán ahora:

- a)  $w'_{ij} = 3k-1$ ; b)  $w'_{ij} = 2k$ ;
- c)  $w'_{ij} = 1+k$ ; d)  $w'_{ij} = 2$ ;
- e)  $w'_{ij} = 3-k$  siendo  $k \geq 1/3$  en el caso a) y  $0 \leq k \leq 3$  en el caso e).

La metodología antes indicada implica la existencia de una matriz de probabilidades de transición (MPT) con las siguientes características:

$(MPT) = \begin{pmatrix} I & O \\ R & Q \end{pmatrix}$  donde la submatriz  $I$  sería de orden  $n$ , la submatriz  $O$  es una matriz nula de orden  $n(n-1)/2 \times n$ ,  $R$  es la m. de p. de t. de estados transitorios (ET) a estados absorbentes (EA) y es de orden  $n \times n(n-1)/2$ , y por último,  $Q$  es la m. de p. de t. entre ET y es una matriz diagonal de orden  $n(n-1)/2$ .

Como ya se adelantó, se han obtenido las expresiones de la media (T) de la variancia (T2) de los tiempos de espera hasta llegar a EA para lo cual se ha utilizado la metodología indicada por Bosso et al. (1969). Dichas expresiones indican la relación existente entre la media y la variancia de los tiempos de espera y los valores selectivos.

Se ha efectuado un análisis matemático de dichas funciones indicando para cada curva la intercepción con el eje de ordenadas, las asíntotas verticales y horizontales, los puntos de inflexión y los extremos relativos dentro del dominio de definición de las variables con sentido biológico real, siguiendo la metodología tradicional (Britton et al., 1978). Para la obtención de los parámetros buscados, la matriz de partida es la submatriz  $Q$ . La submatriz  $Q$  está compuesta, a su vez, por  $n-1$  submatrices diagonales con la siguiente cantidad de elementos:

Submatriz	Nº de elementos	Nº acumulado de elementos
$Q_1$	$n-1$	$n-1$
$Q_2$	$n-2$	$2n-3$
$Q_3$	$n-3$	$3n-6$
.	.	.
.	.	.
$Q_{n-1}$	$n-(n-1)-1$	$n(n-1)-n(n-1)/2 = n(n-1)/2$

De modo que hay un total de  $n(n-1)/2$  elementos en la submatriz  $Q$ . Los elementos son:

$$q_{11} = 2W_{12} w_{12} / c_{12};$$

$$q_{22} = 2W_{13} w_{13} / c_{13}; \dots;$$

$$q_{n-1;n-1} = 2W_{1n} w_{1n} / c_{1n};$$

$$q_{n,n} = 2W_{23} w_{23} / c_{23}; \dots;$$

$$q_{2n-3;2n-3} = 2W_{2n} w_{2n} / c_{2n};$$

$$q_{2n-2;2n-2} = 2W_{34} w_{34} / c_{34}; \dots;$$

$$q_{3n-6;3n-6} = 2W_{3n} w_{3n} / c_{3n}; \dots;$$

$$q_{\frac{n(n-1)}{2}; \frac{n(n-1)}{2}} = 2W_{n-1;n} w_{n-1;n} / c_{n-1;n}$$

donde:  $c_{12} = W_{11} w_{11}^2 + 2W_{12} w_{12} + W_{22} w_{22}^2$

$$c_{13} = W_{11} w_{11}^2 + 2W_{13} w_{13} + W_{33} w_{33}^2$$

...

$$c_{1n} = W_{11} w_{11}^2 + 2W_{1n} w_{1n} + W_{nn} w_{nn}^2$$

...

$$c_{n-1;n} = W_{n-1;n-1} w_{n-1;n-1}^2 + 2W_{n-1;n} w_{n-1;n} + W_{nn} w_{nn}^2$$

La submatriz R está compuesta de n-1 submatrices cuyos elementos son:

R<sub>1</sub> (de orden (n-1) X n):

$$r_{11} = W_{11} w_{11}^2 / c_{12}; r_{21} = W_{11} w_{11}^2 / c_{13}; \dots;$$

$$; r_{n-1;1} = W_{11} w_{11}^2 / c_{1n}$$

$$r_{12} = W_{22} w_{22}^2 / c_{12}; r_{23} = W_{33} w_{33}^2 / c_{13}; \dots;$$

$$; r_{n-1;n} = W_{nn} w_{nn}^2 / c_{1n}$$

siendo el resto de los elementos, ceros.

R<sub>2</sub> (de orden (n-2) X n):

$$r_{12} = W_{22} w_{22}^2 / c_{23}; r_{22} = W_{22} w_{22}^2 / c_{24}; \dots;$$

$$; r_{n-2;2} = W_{22} w_{22}^2 / c_{2n}$$

$$r_{13} = W_{33} w_{33}^2 / c_{23}; r_{24} = W_{44} w_{44}^2 / c_{24}; \dots;$$

$$; r_{n-2;n} = W_{nn} w_{nn}^2 / c_{2n}$$

siendo el resto de los elementos, ceros.

.....

R<sub>n-1</sub> (de orden 1 X n):

$$r_{1;n-1} = W_{n-1;n} w_{n-1;n}^2 / c_{n-1;n};$$

$$; r_{1;n} = W_{nn} w_{nn}^2 / c_{n-1;n}$$

siendo el resto de los elementos, ceros.

El vector T es un vector columna cuyos elementos T<sub>ij</sub> tienen la expresión

$$T_{ij} = 1 + \frac{2W_{i;i+j} w_{i;i+j}}{W_{ii} w_{ii}^2 + W_{i+j;i+j} w_{i+j}^2} = 1 + t_{i;i+j}$$

El vector está compuesto por un total de n(n-1)/2 elementos distribuidos en n-1 subvectores cuyos elementos son:

Subvector	Elementos
T <sub>1</sub>	T <sub>11</sub> , T <sub>12</sub> , ..., T <sub>1;n-1</sub>
T <sub>2</sub>	T <sub>21</sub> , T <sub>22</sub> , ..., T <sub>2;n-2</sub>
..	...
T <sub>n-1</sub>	T <sub>n(n-1)/2}; 1</sub>

Se dará a la expresión:

$$\frac{2W_{1;i+j}W_{1;i+j}}{W_{1;i}^2 + W_{1;j;i+j}^2}$$

el símbolo  $t_{1;i+j}$

El vector  $T_2$  tiene una configuración similar al vector  $T$ . Sus elementos

tienen la expresión:

$$T_{2ij} = t_{1;i+j} + t_{1;i+j}^2$$

Dado que las conclusiones son similares para todos los genotipos heterocigotas desde los cuales el sistema puede partir, se darán los resultados y las conclusiones generales en función de un parámetro indefinido  $t$  ( $T=1+t$ ).

RESULTADOS

En función de los valores selectivos anteriormente mencionados se tienen las siguientes expresiones para  $T$ :

$$T_a = 1 + \frac{k(3k-1)}{1+k^3}; T_b = 1 + \frac{2k^2}{1+k^3}; T_c = 1 + \frac{k}{1+k^3}; T_d = 1 + \frac{2k}{1+k^3}; T_e = 1 + \frac{k(3-k)}{1+k^3};$$

el dominio de la variable  $k$  impone que  $k \geq 0$  para todos los casos, que en el caso  $T_a, k \geq (1/3)$  y que en  $T_e, k \leq 3$ . Estamos hablando del dominio de la variable para el cual la variable tiene sentido biológico. Las expresiones para  $T_2$  se obtienen reemplazando en la expresión  $t + t^2$ , los valores respectivos de  $t$  dados para la media  $T$ . Los resultados para  $T$  están en el Cuadro N° 1. Los resultados para  $T_2$  están en el Cuadro N° 2.

Cuadro N° 1: Valores de las asíntotas, puntos de inflexión, intercepción y extremos relativos para  $T = f(k)$ .

T	Asíntotas		Puntos de inflexión (k)	Extremos relativos		Intercepción (T)
	Horiz. (T)	Vert. (k)		Máximo (k)	Mínimo (k)	
$T_a$	1	-1	0,631	1,478	0,165	1
$T_b$	1	-1	0,526	1,259	0,000	1
$T_c$	1	---	0,347	1,000	-1,000	1
$T_d$	1	-1	1,260	0,793	---	1
$T_e$	1	-1	1,166 y 8,988	0,677	6,041	1

P.D.: en todos los casos, debe considerarse que hay valores de  $k$  que salen del rango de valores establecidos.

Cuadro N° 2: Valores de las asíntotas, puntos de inflexión, intercepción y extremos relativos para  $T_2 = f(k)$ .

$T_2$	Asíntotas		Puntos de inflexión (k)	Extremos relativos		Intercepción (T)
	Horiz. ( $T_2$ )	Vert. (k)		Máximo (k)	Mínimo (k)	
$T_{2a}$	0	-1	0,832 y 2,042	1,478	-6,341	0
$T_{2b}$	0	-1	0,650 y 1,814	1,259	0,000	0
$T_{2c}$	0	-1	0,391 y 1,606	1,000	-1,000	0
$T_{2d}$	0	-1	0,344 y 1,183	0,793	-0,246	0
$T_{2e}$	0	-1	0,272 y 1,029	0,677	6,041	0

P.D.: en todos los casos, debe considerarse que hay valores de k que salen del rango de valores establecidos.

CONSIDERACIONES

Promedios

Todos los promedios tienen como asíntota horizontal  $T=1$  lo que significa que, siendo  $dT/dk > 0$  para todo k superior a 1,478, a medida que  $k \rightarrow \infty$  el valor de T se aproxima indefinidamente a 1.  $k \rightarrow \infty$  implica que la superioridad de un alelo sobre otro toma un valor arbitrariamente grande.

Las asíntotas verticales corresponden a valores de  $k < 0$  lo cual, como queda dicho, carece de sentido biológico. Sólo en el caso  $T_c$ , la función tiene como asíntota horizontal  $T=1$  también para  $k < 0$ , puesto que pertenece al caso general de funciones de la forma:

$$a_1 + \frac{a_2 x}{a_3 x^2 + a_4 x + a_5}$$

Los máximos relativos van disminuyendo desde  $T_a$  hasta  $T_e$ , es decir, a medida que  $W_{ij}^a$  aumenta (para  $k < 1$ ) o disminuye (para  $k > 1$ ). En  $k=1$ , todas las funciones tienen el mismo valor ( $T = 2$ ).

Los mínimos relativos que existen tienen valores sin sentido biológico (salvo  $T_b$ ) puesto que se encuentran fuera del rango de valores establecidos de  $k.t_d$ ; no tiene mínimos relativos puesto que la función tiende a 1 cuando  $k \rightarrow \infty$  y a  $-\infty$  cuando  $k \rightarrow -1+0$  (o sea, por derecha). La función no es analizada para  $k < -1$ .

Los puntos de inflexión disminuyen desde  $T_a$  hasta  $T_e$  de acuerdo a la disminución progresiva de los máximos en ese sentido, lo que ilustra acerca del parecido entre las funciones: son funciones de la forma,

$$1 + \frac{a_1 k^2 + a_2 k}{1+k^3}$$

que difieren entre sí en los valores de los parámetros  $a_i$ .

Variaciones

Todas las funciones tienen asíntota horizontal  $T_2 = 0$  ( $T_2 \rightarrow 0$  cuando  $k \rightarrow +\infty$ ) y asíntota vertical  $k = -1$ , lo que no interesa desde el punto de



$k= 1,166$  (dominancia igual a  $0,332$ ), con dominancia completa para los casos **b** (dominancia igual a  $-0,474$ ) y **d** (dominancia igual a  $0,26$ ), y con ventaja de uno de los homocigotas en el caso **c** (dominancia igual a  $0$ ).

#### Variaciones

En todos los casos, las variaciones tienden a cero a medida que la ventaja selectiva de  $w_1$  sobre  $w_2$  se hace arbitrariamente grande (salvo en  $T_{2e}$  donde el máximo valor de  $k$  admitido es  $3$ ).

Existe una manifiesta correlación entre medias y variaciones, de modo que se esperan los mayores valores de variación con los mismos valores de dominancia que para los promedios. En cambio, las inversiones de tendencia tienen lugar con los siguientes valores de dominancia: para el caso **a**  $-0,336$  (subdominancia) y  $1,984$  (sobredominancia) para el caso **b**,  $-0,35$  y  $0,814$ ; para el caso **c**,  $0$ ; para **d**,  $0,656$  y  $-0,183$  y; por último, para el caso **e**,  $1,456$  (sobredominancia) y,  $-0,058$  (subdominancia).

#### BIBLIOGRAFIA

- 1) BOSSO, J.A.; O.M. SORRARAIN and E.A. FAVRET. 1969. Application of finite absorbent Markov chains to sib mating populations with selection. *Biometrics*, 22(1):17-26.
- 2) BRITTON, J.R.; R. BEN KRIEGH y L.W. RUTLAND. 1976. *Matemáticas universitarias*. Ed. CECSA México. 2 Volúmenes. pp. 381-426 (Vol. 1).