

## ENFOQUE PROBABILISTICO PARA CALCULAR EL TAMAÑO DE POBLACION EFECTIVO. I. AUTOFECUNDACION EXCLUIDA

Jaime Sahagún Castellanos

### RESUMEN

El número efectivo ( $N_e$ ) es un concepto que fue creado con el fin de que cualquier población, independientemente de su estructura de apareamiento, pueda ser analizada en términos de una población estándar. El número efectivo es de importancia en diversos aspectos teóricos y aplicados. En tanto que la deriva genética y la depresión endogámica están inversamente relacionados con  $N_e$ , la relación es directa entre  $N_e$  y el potencial del avance genético a largo plazo; sin embargo, la derivación del número efectivo, como originalmente fue planteada, demanda de herramientas estadísticas poco comunes que, en general, dificultan su aplicación. En este estudio se describe con detalle un método más simple para calcular  $N_e$ , basado solamente en los principios probabilísticos básicos. La sencillez y el potencial de este método se ilustran con la determinación del número efectivo de una población diploide de  $N$  individuos monoicos que se reproducen por apareamiento aleatorio donde se excluye la autofecundación. Al igual que con el método original, se obtuvo el resultado  $N_e \doteq N + 0.5$ , mismo que se discute e ilustra la eficiencia del método descrito.

### PALABRAS CLAVE ADICIONALES

Población monoica diploide, deriva genética, avance genético, población ideal.

### SUMMARY

The concept of effective population number ( $N_e$ ) was created in order to analyze any population, independently of its mating structure, in terms of a

standard population. The  $N_e$  is thus important in various theoretical and applied topics. Whereas the random genetic drift and inbreeding depression are inversely related to  $N_e$ , the relationship between  $N_e$  and the potential for long-term genetic improvement with moderate selection pressure is direct. However, the derivation of  $N_e$ , as originally described, requires uncommon statistical tools which make difficult its application. In this study, an easier procedure to determine the  $N_e$ , based upon the basic probabilistic principles, is described in detail. The simplicity and potential of this procedure is illustrated by determining the  $N_e$  of a monoecious diploid population ( $N$  individuals) under random mating with self fertilization excluded. As with the original procedure, the obtained result was  $N_e \doteq N + 0.5$ , which illustrates the efficiency of the proposed procedure.

### ADDITIONAL INDEX WORDS

Monoecious diploid population, genetic drift, genetic progress, ideal population.

### INTRODUCCION

El número efectivo de una población de individuos es un concepto creado por Wright (1931) con el fin de que el análisis de cualquier población, independientemente de su estructura de apareamiento, pueda hacerse en los términos de una población estándar, la que ha servido como punto de referencia y de cuyo estudio han derivado gran cantidad de conocimientos básicos de la Genética de Poblaciones. El número efectivo es muy informativo con respecto a diferentes tópicos teóricos y aplicados. Por ejemplo, Robertson (1960) demostró que cuando se hace mejoramiento genético a largo plazo utilizando una

-----  
1 Profesor Investigador, Departamento de Fitotecnia, Universidad Autónoma Chapingo. C.P. 56230. Chapingo, Edo. de Méx.

presión de selección moderada, el potencial del progreso esperado resulta ser directamente proporcional al número efectivo de la población. Crow y Kimura (1970), por su parte, muestran que el nivel de endogamia, la probabilidad de deriva genética y la probabilidad de extinción de un gene son inversamente proporcionales al número efectivo.

La derivación del número efectivo como originalmente fue planteada por Wright (1931), basada en el concepto del coeficiente de la correlación entre gametos que se unen para formar los cigotes, es un proceso ciertamente complejo que requiere de la relativamente sofisticada, desde un punto de vista estadístico, teoría de los coeficientes de sendero. El método que aquí se describe ha sido desarrollado en términos que permiten manejarlo sin más herramientas matemáticas que las involucradas en los principios básicos de la teoría probabilística.

La determinación del número efectivo utilizando el enfoque probabilístico ha sido realizada por Kimura y Crow (1963) para los casos en que, en relación a la población ideal: 1) todos los individuos contribuyen igualmente a la formación de la siguiente generación; 2) un solo individuo genera la siguiente generación; 3) tamaño de población constante; 4) contribución constante de los progenitores y tamaño de población constante; 5) contribución constante de los progenitores y tamaño de población decreciente; 6) contribución constante de los progenitores y progenitores homocigóticos ( $F=1$ ); 7) autofecundación; y 8) apareamiento entre hermanos. Sin embargo, el caso en que en una población diploide de individuos monoicos se suprime la autofecundación, sólo ha sido estudiado desde el punto de vista de los coeficientes de sendero por Wright (1931).

Los objetivos del presente estudio son: 1) describir con detalle un procedimiento para determinar el número efectivo con base en principios básicos de la probabilidad y 2) derivar el número efectivo, con base en sólo los principios probabilísticos básicos, de una población diploide de individuos monoicos en donde la autofecundación se encuentra suprimida.

## DESCRIPCION DEL METODO

En el contexto del tópico del número efectivo, el concepto de "población ideal" juega un papel crucial. Siguiendo a Falconer (1981), supóngase una población diploide de tamaño infinito de individuos diploides y monoicos que se reproducen por apareamiento aleatorio; supóngase además que, por muestreo aleatorio de esta población, se forma un número muy grande de subpoblaciones de tamaño  $N$  en las que: 1) el apareamiento al azar se realiza estrictamente entre los individuos de cada población; 2) no ocurre sobreposición de generaciones; 3) el número de individuos de cada subpoblación es constante a través de las generaciones y 4) ausencia completa de mutación, selección y migración. En realidad, para los efectos de este estudio es suficiente considerar la población base y sólo una subpoblación con las características anotadas. Es claro que en una subpoblación de tamaño  $N$  que avanza generacionalmente de acuerdo con las características de la población ideal, al cabo de un cierto número de generaciones algunos individuos de la población tendrán ancestros en común que, en consecuencia, llegado el momento del apareamiento, la endogamia surge como una posibilidad.

Considerando que el coeficiente de endogamia de la población base ( $F_0$ ) es cero, entonces, en relación con un locus, todos los

genes que forman los genotipos de la población base deben ser no idénticos por descendencia. De esta manera, en una subpoblación de N individuos tomados al azar de la población original, en relación con un locus, todos los genes (2N) serán no idénticos por descendencia o bien, en términos de gametos, habrá un número, 2N, de tipos diferentes de gametos (diferentes en cuanto a que cada gameto de un tipo cualquiera es portador de un gene que no es idéntico por descendencia al gene que es portado por un gameto de otro tipo cualquiera). Dado que el apareamiento es aleatorio, la formación de un cigote (un individuo de la generación 1) se puede representar como la extracción al azar de un gameto (masculino) de un recipiente que contiene 2N gametos, uno de cada tipo, y la extracción al azar de un gameto (femenino) de otro recipiente que contiene 2N gametos, uno de cada tipo. Así, la probabilidad de que tal cigote resulte formado por dos gametos que porten genes idénticos por descendencia ( $F_1$ ) se determina con base en el razonamiento siguiente: si los genes del i-ésimo individuo de la generación 0 son  $A_{1i}$  y  $A_{2i}$  ( $i = 1, 2, \dots, N$ ), entonces, considerando que la extracción del gameto masculino se realiza en forma independiente de la extracción del gameto femenino,

$$\begin{aligned}
 F_1 &= P(\text{extraer dos gametos con genes idénticos por descendencia}) \\
 &= P(A_{11}A_{11} \cup A_{21}A_{21} \cup A_{12}A_{12} \cup \dots \\
 &\quad \cup A_{1N}A_{1N} \cup A_{2N}A_{2N}) \\
 &= P(A_{11}A_{11}) + P(A_{21}A_{21}) + P(A_{12}A_{12}) + \dots + \\
 &\quad P(A_{1N}A_{1N}) + P(A_{2N}A_{2N}) \\
 &= P(A_{11})P(A_{11}) + P(A_{21})P(A_{21}) + P(A_{12}) \\
 &\quad P(A_{12}) + \dots + P(A_{2N})P(A_{2N}) \\
 &= (1/2N)(1/2N) + (1/2N)(1/2N) + (1/2N) \\
 &\quad (1/2N) + \dots + (1/2N)(1/2N) \\
 &= 2N(1/2N)(1/2N) \\
 &= 1/2N \tag{1}
 \end{aligned}$$

Esta expresión indica que el coeficiente de endogamia de la primera generación es  $1/2N$ . En otras palabras, la expresión (1) indica que en la generación 1 una proporción de individuos ( $1/2N$ ) portarán genes idénticos por descendencia y que la proporción restante ( $1-1/2N$ ) estará constituida por individuos que no portan genes idénticos por descendencia. Es conveniente observar que los individuos de la generación 1 que portan dos alelos idénticos por descendencia sólo pudieron haber sido formados por autofecundación. Como resultado del apareamiento aleatorio de los individuos que constituyen la generación 1 se formarán cigotes de dos tipos: los que portan alelos que son copias de un mismo alelo de la generación 1 (tipo A) que, al igual que en la generación 1, ocurren en una proporción de  $1/2N$ , y los que portan alelos que no son copia de un mismo alelo de la generación 1 (tipo B). Claramente, la probabilidad de que un cigote del tipo A porte dos alelos idénticos por descendencia es 1, en tanto que la probabilidad de que un cigote del tipo B porte dos alelos idénticos por descendencia es  $F_1$ . Por lo tanto, la proporción de individuos de la generación 2 que portan alelos idénticos por descendencia ( $F_2$ ) es

$$F_2 = 1/2N + (1-1/2N) F_1.$$

Es interesante hacer notar que los individuos del tipo A sólo pudieron haber sido formados por autofecundación, en tanto que de los individuos del tipo B, unos son producto de la autofecundación y el resto tiene como origen el apareamiento de individuos diferentes.

Para la generación t, por un razonamiento análogo al anterior, se puede encontrar que la proporción de individuos que portan alelos idénticos por descendencia ( $F_t$ ) es

$$F_t = 1/2N + (1-1/2N) F_{t-1} \quad (2)$$

Falconer (1981), al tratar este tema, define la tasa de endogamia ( $\Delta_F$ ) en una población ideal en la forma  $\Delta_F = 1/2N$  y, consecuentemente, expresa el coeficiente de endogamia de la generación  $t$  en la forma

$$F_t = \Delta_F + (1 - \Delta_F) F_{t-1}$$

A partir de esta expresión se encuentra que:

$$\Delta_F = \frac{F_t - F_{t-1}}{1 - F_{t-1}}, \quad (3)$$

lo cual permite apreciar que la tasa de endogamia es el incremento en el coeficiente de endogamia obtenido cuando se avanza de la generación  $t-1$  a la generación  $t$ , expresado como proporción del complemento de  $F_{t-1}$ .

Cuando una población particular no se ajusta a las especificaciones de la población ideal pero tiene una tasa de endogamia conocida, el número efectivo de dicha población ( $N_e$ ) resulta ser el tamaño de la población ideal que daría lugar a la misma tasa de endogamia. En una población monoica diploide que se reproduce por apareamiento aleatorio en donde, por ejemplo, se suprime la autofecundación, la tasa de endogamia que se genera es, intuitivamente, más pequeña en relación a la que se generaría en una población ideal (en la que sí se produce el fenómeno de la autofecundación) del mismo tamaño; es decir, para que la población ideal genere una tasa de endogamia igual, su tamaño tendría que ser mayor que el de la población en donde se excluye la autofecundación. Este tamaño de la población ideal sería el número efectivo de la población en estudio. A

continuación se derivará, de una manera muy simple, un procedimiento que permite determinar de una manera exacta el número efectivo de ciertas poblaciones que presentan, como el caso analizado, diferencias con respecto a la población ideal.

En virtud de que en la población ideal se tiene que  $\Delta_F = 1/2N$ , resulta que  $N = 1/2 \Delta_F$ . En general, si una población no ideal cualquiera de cualquier tamaño genera una tasa de endogamia  $\Delta_F$ , su número efectivo será expresable como

$$N_e = 1/2 \Delta_F. \quad (4)$$

Esta expresión puede derivarse y reescribirse de una forma alternativa interesante. Siguiendo a Kimura y Crow (1963), si en una población ideal  $P_t$  representa la probabilidad de que un individuo al azar de la generación  $t$  porte dos alelos que provienen de un mismo individuo de una generación anterior, el coeficiente de endogamia  $F_t$  puede ser expresado en términos de  $P_t$ . Un individuo al azar de la generación  $t$  puede ser portador de dos alelos idénticos por descendencia cuando dichos alelos provengan de un mismo individuo de la generación  $t-1$ , o bien cuando provengan de dos individuos emparentados de la generación  $t-1$ . Ahora bien, la probabilidad de que un individuo al azar de la generación  $t$  porte dos alelos de un mismo progenitor y de dos progenitores diferentes son, por definición,  $P_t$  y  $(1-P_t)$ , respectivamente. Dado un individuo de la generación  $t$  que recibió dos alelos de un mismo progenitor, la probabilidad de que éstos sean copias de un mismo gene del progenitor es  $1/2$  y la probabilidad de que sean copias, una de cada uno, de los dos alelos del progenitor es, consecuentemente, también  $1/2$ . Claramente, si los dos alelos son copias de un mismo alelo, la probabilidad de que sean idénticos

por descendencia es 1 y si, en cambio, son copias, una de cada uno de los dos alelos de un mismo progenitor, la probabilidad de que sean idénticos por descendencia es  $F_{t-1}$ . Si, en cambio, los alelos del individuo al azar de la generación  $t$  son copias, una de cada uno de dos alelos portados por individuos diferentes de la generación  $t-1$ , la probabilidad de que sean idénticos por descendencia es  $F_{t-1}$ , ya que por apareamiento aleatorio, la probabilidad de que cualesquiera dos alelos al azar se unan, es la misma que la de los dos alelos de un mismo individuo. Resumiendo la información anterior, se tiene que

$$F_t = P_t \left( (1/2)(1) + (1/2)F_{t-1} \right) + (1-P_t)F_{t-1}$$

y de aquí que

$$F_t = (1/2) P_t (1 - F_{t-1}) + F_{t-1}$$

de donde

$$(1/2)P_t = \frac{F_t - F_{t-1}}{1 - F_{t-1}} \quad (5)$$

Ahora, de acuerdo con las expresiones (3) y (5) resulta que  $\Delta_F = (P_t)/2$  y de aquí, en congruencia con (4), resulta que

$$N_e = 1/P_t; \quad (6)$$

es decir, el número efectivo, en términos numéricos, resulta ser expresable simplemente como el recíproco de la probabilidad de que dos gametos al azar provengan de un mismo individuo de una generación anterior.

### NUMERO EFECTIVO DE UNA POBLACION SIN AUTOFECUNDACION

En la generación  $t$  de una población monoica y diploide de  $N$  individuos, 1, 2, 3,

...,  $N$ , cuya única diferencia con respecto a la población ideal consiste en que no se realiza la autofecundación, el número total de apareamientos posibles (cruzas simples), incluyendo directos y recíprocos, es  $N(N-1)$ . De los individuos resultantes de estas cruzas, generación  $t+1$ , ninguno puede ser portador de dos alelos que provengan de un mismo progenitor de la generación  $t$ . En la generación  $t+2$ , en cambio, la presencia de individuos que porten dos alelos de un mismo progenitor de la generación  $t$ , sí es factible. El número de cruzas posibles (cruzas dobles) que se pueden hacer con los individuos de la generación  $t+1$  es  $N(N-1)N(N-1) - N(N-1)$ . Estas cruzas dobles, para los efectos del problema, pueden ser clasificadas, con relación a los progenitores originales (generación  $t$ ), en tres tipos:

1. Cruzas dobles formadas con cuatro progenitores de la población original diferentes;
2. Cruzas dobles formadas con tres progenitores de la población original, diferentes, y
3. Cruzas dobles formadas con dos progenitores de la población original diferentes.

Para calcular la probabilidad de que un individuo al azar de la generación  $t+2$  porte dos alelos provenientes de un mismo progenitor de la población original, es necesario determinar el número de cruzas dobles de cada tipo y la probabilidad con que cada tipo de cruce genera un individuo que porte dos alelos provenientes de un mismo progenitor original.

Es obvio que ninguna cruce doble del tipo 1 puede generar individuos que porten dos alelos provenientes de un mismo progenitor.

El número de cruza dobles de este tipo es  $N(N-1)(N-2)(N-3)$ .

Con respecto al número de cruza dobles posibles del tipo 2, dada una cruza simple, digamos  $ixj$  ( $i \neq j$ ), la otra cruza simple puede involucrar al progenitor  $i$  o al  $j$  y a uno de los  $N-2$  progenitores originales restantes. Así, si el tercer progenitor original es  $k$  ( $k \neq i$  y  $k \neq j$ ), las cruza dobles que se pueden realizar son cuatro:  $(ixj) \times (ixk)$ ,  $(ixj) \times (kxi)$ ,  $(ixk) \times (ixj)$  y  $(kxi) \times (ixj)$ . Como el progenitor  $i$  puede ser uno cualquiera de los  $N$  progenitores originales, el  $j$  podrá ser uno cualquiera de los  $N-1$  restantes y el  $k$  uno cualquiera de los  $N-2$  restantes, entonces el número de cruza dobles posibles del tipo 2 es  $4N(N-1)(N-2)$ . Ahora bien, dada una cruza doble cualquiera del tipo 2, la probabilidad de que se genere un individuo con dos alelos que provengan de un mismo progenitor original es  $1/4$ . Esto es así porque para una cruza, digamos  $(ixj) \times (ixk)$ , el progenitor original que se repite,  $i$ , necesariamente transmite un alelo a cada cruza simple, de manera que al formarse la cruza doble, cada cruza simple transmitirá el alelo proveniente de  $i$  con una probabilidad de  $1/2$ . Como se trata de eventos independientes, la probabilidad de que ambas cruza simples transmitan un alelo de  $i$  a la cruza doble es, por lo tanto,  $1/4$ .

Finalmente, dada una cruza doble del tipo 3, digamos  $(ixj) \times (jxi)$ ,  $i \neq j$ , la probabilidad de que un individuo resultante de esta cruza reciba dos alelos provenientes de un mismo progenitor de la generación  $t$  es  $1/2$ , el doble en relación a la de una cruza del tipo 2. Esto es así, porque los dos alelos provenientes de un mismo progenitor pueden provenir, con una misma probabilidad ( $1/2$ ), del progenitor original  $i$  o del progenitor original  $j$ . Por

otra parte, es obvio que el número de cruza dobles posibles del tipo 3 es  $N(N-1)$ .

Reuniendo la información anterior referente al número posible de cruza de cada tipo y su probabilidad correspondiente de reunir alelos de un progenitor común, resulta que

$$\begin{aligned}
 P_t &= \frac{N(N-1)(N-2)(N-3)(0) + 4N(N-1)(N-2)(1/4) + N(N-1)(1/2)}{N(N-1)N(N-1) - N(N-1)} \\
 &= \frac{N(N-1)(N-2+1/2)}{N(N-1)(N(N-1)-1)} \\
 &= \frac{N-1.5}{N(N-1)-1} \\
 &= \frac{N-1.5}{(N-1.5)(N+0.5) - 0.25}
 \end{aligned}$$

Por lo tanto, de acuerdo con la expresión (6),

$$\begin{aligned}
 N_e &= \frac{(N-1.5)(N+0.5) - 0.25}{N-1.5} \\
 &= N + 0.5 - \frac{0.25}{N-1.5} \tag{7}
 \end{aligned}$$

y si  $N$  es muy grande

$$N_e \doteq N + 0.5 \tag{8}$$

## DISCUSION

Al practicar selección, el fitomejorador actúa como un promotor del incremento de las frecuencias de los genes favorables; sin embargo, en forma simultánea puede ocurrir una pérdida de variabilidad genética, debida al fenómeno conocido como deriva genética, que trae consigo su cuota de reducción en el potencial del avance genético. Otra consecuencia indeseable que origina la selección, de particular importancia en las especies alógamas, se refiere a la depresión endogámica causada por el apareamiento entre individuos emparentados.

Tanto la deriva genética como la depresión endogámica mantienen una relación inversa con el número efectivo; a números efectivos menores corresponderán mayores probabilidades de pérdida (o fijación) de genes y mayores tasas de depresión endogámica (Crow y Kimura, 1970). Más precisamente, la deriva genética, que se manifiesta mediante una reducción en la variabilidad genética, y la depresión endogámica se encuentran inversamente relacionados con el número efectivo expresado en términos de varianza (el tamaño de la población ideal que dará lugar a la misma cantidad de variación aleatoria en las frecuencias génicas que la población en cuestión) y al número efectivo en términos de endogamia (el tamaño de la población en cuestión), respectivamente (Kimura y Crow, 1963). Debido a que en la práctica no tienen mucha utilidad las diferencias entre los valores de ambos conceptos de número efectivo en una población de diez o más individuos bajo apareamiento aleatorio, el uso de uno solo resulta adecuado (Wells *et al.*, 1987).

Por supuesto, el sólo conocer el tipo de relaciones que la probabilidad de deriva

genética, la tasa de endogamia y el potencial del avance genético tienen con el número efectivo, no es suficiente. El fitomejorador demanda de la determinación de los números efectivos críticos que concilien el potencial del avance genético con los recursos requeridos para el desarrollo de programas de mejoramiento basados con poblaciones de los tamaños que correspondan a esos números efectivos. Esta es una tarea en la que hay mucho por hacer (Gardner, 1977). No obstante, el contar con un procedimiento que permita determinar en forma expedita el número efectivo de los diferentes tipos de población es un avance substantivo en esa dirección.

Como se ha dicho, cuando en una población diploide de  $N$  individuos monoicos que se reproducen por apareamiento aleatorio se suprime la autofecundación, Wright (1931) encontró, mediante técnicas estadísticas relativamente complejas, que el número efectivo de esa población es, aproximadamente,  $N_e \doteq N+0.5$ , que es el mismo encontrado en este estudio mediante un método de una gran simplicidad, basado sólo en el conocimiento de los principios probabilísticos básicos. Esto ilustra el gran potencial del método aquí descrito para la determinación del número efectivo en las diferentes poblaciones en que se pueden basar ciertos programas de mejoramiento genético.

Por lo que se refiere al número efectivo encontrado para la población descrita,  $N_e \doteq N+0.5$ , parecería que por el hecho de suprimir la autofecundación uno debería esperar una considerable reducción de la endogamia y, en consecuencia, un número efectivo mayor que el obtenido. La exclusión de la autofecundación, sin embargo, no es un factor de mucha relevancia en la disminución de la tasa de endogamia en la población; aunque,

efectivamente, las progenies de los individuos autofecundados tienen una mayor proporción de individuos portadores de dos alelos idénticos por descendencia, este incremento en la tasa de endogamia no es permanente. En una población con apareamiento aleatorio, los individuos producto de la autofecundación muy probablemente se aparearán con individuos no (o muy poco) emparentados, de tal suerte que su coeficiente de endogamia, mayor que el del promedio de los individuos restantes, no se transmite a su descendencia. Por lo tanto, la exclusión de la autofecundación no es un factor importante en la disminución de la tasa endogámica.

### BIBLIOGRAFIA

- Crow, J. F. and M. Kimura. 1970.** An Introduction to Population Genetics Theory. Harper and Row Publishers, New York. 591 p.
- Falconer, D. S. 1981.** Introduction to Quantitative Genetics. Longman, New York. 340 p.
- Gardner, C. O. 1977.** Quantitative genetics and population study in maize and sorghum. In: Proceedings of the International Conference on Quantitative Genetics. E. Pollak, O. Kempthorne, and T. B. Bailey, Jr. (eds.). Iowa State Univ. Press, Ames, Io. pp. 29-37.
- Kimura, M. and J. F. Crow. 1963.** The measurement of effective population number. Evolution 17:279-288.
- Robertson, A. 1960.** A theory of limits in artificial selection. Proc. of the Royal Soc. of London 153:234-249.
- Wells, W. C., M. L. Rosse, and M. R. Guzy. 1987.** Effects of selection parameters on effective population size for mass selection. Crop Sci. 27:1146-1149.
- Wright, S. 1931.** Evolution in mendelian populations. Genetics 16:97-159.