

LA DISTRIBUCION BINOMIAL NEGATIVA EN EL ESTUDIO DE POBLACIONES DE INSECTOS

*Por Hernán Gómez López**

I. INTRODUCCION

A diferencia de los fenómenos físico-químicos que pueden estudiarse bajo modelos matemáticos muy precisos, los fenómenos biológicos están sujetos a un grado apreciable de incertidumbre debido a la complejidad de los factores que los gobiernan.

Por ello, los modelos estadísticos que explican sus manifestaciones y sus leyes son apenas burdas aproximaciones y simplificaciones de esa compleja realidad. En un modelo matemático sólo se incluyen aquellos factores considerados como "relevantes" y que a su vez permiten un manejo eficiente de la información.

Al estudiar la distribución de los seres vivos en el espacio deben tenerse muy en cuenta sus hábitos biológicos.

Usualmente, las poblaciones de organismos están regidas por leyes o distribuciones de "contagio", donde la probabilidad de encontrar un individuo aumenta si ya se ha encontrado otro en un espacio dado.

Para el caso de los insectos, esta tendencia natural de agrupación puede explicarse por diversas causas, entre las cuales se pueden mencionar:

1. Altas densidades de estados inmaduros (larvas o ninfas), como consecuencia de la oviposición en masa.
2. Respuesta a microlimas o a un factor particular como la temperatura, humedad, viento, luz o suelo.
3. Respuesta a fuentes de alimentos separadas en el espacio o preferencia alimenticia por ciertas especies de plantas.

Para estudiar las poblaciones de organismos considerando estos fenómenos de agregación o "contagio", se han postulado algunas distribuciones probabilísticas, entre las cuales sobresale la Distribución Binomial Negativa, que ha presentado un grado de ajuste aceptable en un gran número de casos.

En el presente trabajo se analizan algunas de sus propiedades y su importancia dentro del estudio de poblaciones de insectos.

* Profesor Asociado, Facultad de Agronomía.

II. LA FUNCION DE PROBABILIDAD Y SUS PROPIEDADES

La distribución Binomial Negativa puede generarse mediante diferentes formas (Robles y Carrillo L., 1971).

En interpretaciones biológicas se utiliza la siguiente expresión:

$$f(x) = \frac{\gamma(k+x)}{\gamma(x+1)\gamma(k)} \left(\frac{m}{k+m}\right)^k \cdot \left(\frac{k}{k+m}\right)^x$$

Donde $f(x)$ es la probabilidad de obtener X organismos en una unidad de muestreo.

$\gamma(k)$ es la función matemática Gamma que está dada por:

$$\gamma(k) = \int_0^{\infty} y^{k-1} e^{-y} dy \quad \text{para } k > 0$$

Se puede demostrar que si k es entero $\gamma(k+1) = k!$

Las propiedades matemáticas de la Binomial Negativa han sido descritas por diferentes autores (Anscombe 1949, Bliss 1953, Barko 1961).

La función de probabilidad está completamente especificada por sus dos parámetros m y k : m es la media del número de individuos por unidad de muestreo y el valor de k se ha considerado como una medida de agregación de los organismos (Bliss 1956, Waters 1959, Wilson y Gerard 1971).

A medida que k disminuye, el contagio o agregación aumenta. Si el valor de k aumenta, el contagio disminuye y la distribución tiende hacia la de Poisson (individuos distribuidos al azar).

La media y la varianza están dadas por:

$$E(x) = m \qquad \text{Var}(x) = m + \frac{m^2}{k}$$

La función generatriz de momentos presenta la siguiente expresión:

$$M_x(t) = b^k (1 - ae^t)^{-k}$$

$$\text{Donde } a = \frac{m}{k+m} \text{ y } b = \frac{k}{k+m}$$

Puede observarse que si k tiende a infinito, el valor de la varianza tiende a la media, como ocurre en la distribución de Poisson. En la práctica, si el valor de k es superior a 8, puede considerarse que la distribución es al azar.

III. ESTIMACION DE LOS PARAMETROS

Supongamos que en un muestreo simple aleatorio se ha contado el número de insectos de N unidades de muestreo y se han encontrado unidades con cero insectos, entonces:

1. El parámetro m se estima eficientemente con el promedio aritmético \bar{x} del número de insectos por observación.
2. Para estimar el valor de k se han desarrollado varios métodos (Anscombe 1949, Bliss 1953, Bliss 1956, Katti y Gurland 1962); se mencionan a continuación algunos de los más usuales:

a. Método de Momentos.

De la ecuación: $\text{Var}(x) = m + \frac{m^2}{k}$

Si sustituimos los valores calculados en la muestra \bar{x} y S^2 por m y $\text{Var}(x)$ respectivamente, se tiene:

$$S^2 = \bar{x} + \frac{\bar{x}^2}{k} \quad \text{de donde} \quad \hat{k} = \frac{\bar{x}^2}{S^2 - \bar{x}}$$

con lo cual obtenemos un estimativo para el valor de k .

b. Frecuencia de unidades con cero insectos.

En la expresión:

$$f(x) = \frac{\gamma(k+x)}{\gamma(x+1)\gamma(k)} \left(\frac{m}{k+m}\right)^x \cdot \left(\frac{k}{k+m}\right)^k$$

se obtiene la proporción esperada de ceros (probabilidad de obtener unidades con cero insectos).

$$f(0) = \left(\frac{k}{k+m}\right)^k$$

Si se iguala la proporción esperada de ceros con la proporción observada $\frac{n_0}{N}$ y además se sustituye m por \bar{x} , se obtiene:

$$\frac{n_0}{N} = \left(\frac{k}{k+\bar{x}}\right)^k$$

Ecuación que presenta solución única para k dados n_0 , N y \bar{x} . El valor de k puede obtenerse por aproximaciones sucesivas. Un valor de k para iniciar el proceso puede ser 2.

c. Método de Máxima Verosimilitud.

El valor estimado de k puede obtenerse mediante aproximaciones sucesivas en la siguiente expresión:

$$N [\ln(k + \bar{x}) - \ln(k)] = \frac{n_1 + n_2 + \dots + n_r}{k} + \frac{n_2 + \dots + n_r}{k+1} + \dots + \frac{n_r}{k+r-1}$$

donde n_i es la frecuencia observada de unidades con x_i insectos. $i = 0, 1, 2, \dots, r$. r es el máximo número de insectos encontrados en una unidad de muestreo.

$$N = \sum_{i=0}^r n_i \text{ es el número total de unidades observadas.}$$

$$\bar{x} = \frac{1}{N} \sum x_i n_i \text{ es la media del número de insectos por unidad de muestreo.}$$

Se puede iniciar el proceso con base en un valor de k obtenido por el método (a); si el lado izquierdo de la ecuación resulta superior al derecho, el valor de k asignado es demasiado grande y conviene disminuirlo para continuar el proceso y viceversa. El método iterativo se adelanta hasta encontrar el valor de k que satisfaga la ecuación anterior.

Como ilustración de este método se considera un ejemplo presentado por Bliss 1953.

Para estudiar la distribución del ácaro rojo en las hojas de manzano se tomaron 150 hojas al azar, provenientes de 6 árboles y se contó el número de individuos adultos por hoja. En la tabla siguiente se indican las frecuencias observadas.

No. de ácaros por hoja

x_i

0

1

2

3

4

5

6

7

8

No. de hojas observadas

n_i

70

38

17

10

9

3

2

1

0

$$N = 150$$

$$\bar{x} = 1,147 \quad S^2 = 2.273$$

Por el método (a) se obtiene $\hat{k} \cdot (a) = 1.15$ Con base en este valor se puede iniciar la estimación de k .

$$150 [\ln(1.15 + 1.147) - \ln(1.15)] = \frac{38 + 17 + \dots + 0}{1.15} + \frac{17 + 10 + \dots + 0 + \dots}{1.15 + 1}$$

$$+ \frac{1}{1.15 + 8 - 2} + 0$$

El valor inicial de $k = 1.15$ se ajusta en forma sucesiva hasta que se cumpla la igualdad. Mediante este proceso se llega a un valor de $\hat{k} = 1.025$.

Diferentes autores han analizado la eficiencia de estos métodos de estimación del valor k .

Anscombe 1950, encontró que la efectividad de estos métodos para el caso de muestras grandes varía de acuerdo a los valores de m y k . El método de momentos resulta más apropiado para valores de m bajos, o sea con bajas densidades de población. El método de las frecuencias de cero insectos se encontró más efectivo a medida que k se acerca a cero.

Bowman y Shenton 1966, encontraron sesgos en las estimaciones de estos parámetros bajo los diferentes métodos propuestos; los valores de sesgo varían de acuerdo a m y k y resultan positivos con respecto a la estimación de k .

Pieters et al 1977 utilizaron técnicas de simulación para estudiar la eficiencia de estos métodos en el caso de muestras de tamaño moderado, para valores de k entre 0.2 y 6.

El análisis de los resultados confirmó los sesgos en las estimaciones obtenidas por Bowman y Shenton y además demostró que ninguno de los métodos estudiados puede considerarse como óptimo.

El método de las frecuencias con cero insectos resultó inferior al de momentos y máxima verosimilitud. De éstos dos últimos, el método de momentos se recomienda por su sencillez y en especial, cuando $\hat{m} \leq \hat{k}^2$. Si al utilizar el método anterior se encuentra que $\hat{m} > \hat{k}^2$ entonces, se sugiere utilizar el método de máxima verosimilitud.

IV. IMPORTANCIA EN LA ESTIMACION DE INSECTOS

Se ha comprobado que la Binomial Negativa explica satisfactoriamente la distribución espacial de un buen número de especies de insectos. Como se explicó anteriormente, el parámetro m indica el número promedio de insectos y el parámetro k se ha considerado como una medida de contagio o agregación. Sin embargo, se ha demostrado (Southwood 1966), que el valor de k puede variar según el tamaño de la unidad de muestreo y el estado biológico del insecto, por lo cual, este índice debe tomarse como relativo y no como medida absoluta.

Lloyd 1967 propuso un índice de aglomeración media que permite dar una interpretación biológica al parámetro k . Este índice m^* (mean crowding) indica el número promedio (por individuo) de otros individuos en la misma unidad de muestreo. El autor encontró que

$$\frac{m^*}{m} = 1 + \frac{1}{k}$$

donde m es el promedio de organismos por unidad de muestreo.

De este modo, el recíproco de k es la proporción en que la aglomeración media excede a la densidad media. Así, si $k = 1/3$, entonces, la aglomeración media excede en tres veces a la densidad media.

A continuación se presentan algunos valores de k mencionados por la literatura:

Bliss y Fisher (1953), señalan un valor de $k = 1.455$ para el barrenador del maíz *Pyrausta nubilalis* Hubner, tomando la cepa como unidad de muestreo.

Waters y Henson (1959), encontraron un valor de $k = 0.214$ para las larvas de *Ryacionia frustana* Comstock en pinos, tomando como unidad de muestreo los verticilos.

Harcourth (1965) indica que el valor del parámetro k puede variar a medida que transcurre el desarrollo del insecto así, con larvas de *Trichoplusia ni* Riley en repollo, encontró un valor de $k = 1.28$ en el primer estadio del insecto y un valor de $k = 2.07$ en el quinto estadio.

Rojas (1964) indica un valor de $k = 3.16$ para gusanos alambre (familia Elateridae) en el suelo, utilizando unidades de muestreo de sección circular de 4 pulgadas de diámetro y 6 pulgadas de profundidad.

Southwood (1966) cita un valor de $k = 2.32$ para las larvas de *Pieris rapae* L. en repollo, tomando la planta comunidad de muestreo.

Safranyik y Raske (1970) encontraron un valor de $k = 1.04$ para larvas de *Monochamus oregonensis* Le Conte en pino, tomando como unidad de muestreo un pie cuadrado de área de corteza.

Alvarez (1971) estudiando la conchuela del frijol *Epilachna varivestis* Mulsant, bajo condiciones de invernadero, encontró que el parámetro k variaba según el estado de desarrollo de la larva obteniéndose valores cercanos a 1 en las últimas etapas de desarrollo.

Ford y Dimond (1973) encontraron un índice de agregación de $k = 0.55$ para larvas de *Pineus pinifolie* Fitch, considerando pequeñas ramas como unidades de muestreo.

Pieters, Gates et al (1977) citan los siguientes valores de k para diferentes insectos de algodón:

ESPECIE	VALOR DE k ESTIMADO
<i>Heliothis</i> sp. (huevos)	0.36
<i>Heliothis</i> sp. (larvas pequeñas)	0.88
<i>Heliothis</i> sp. (larvas grandes)	0.62
<i>Geocoris</i> sp.	2.71
<i>Orius</i> sp.	0.82
<i>Schymnus cinctus</i> Muls	0.56

Los valores anteriores se estimaron usando el método de máxima verosimilitud y tomando la planta como unidad de muestreo.

V. DISCUSION

Con base en los resultados encontrados por diferentes investigadores puede decirse que la Distribución Binomial Negativa es la que con mayor frecuencia ha explicado la disposición espacial de los insectos. Sin embargo, su aplicación al estudio de estos organismos presenta algunas restricciones relacionadas con la estimación de sus parámetros, las posibilidades de uso del análisis de la varianza y la falta de consistencia en las poblaciones con respecto al parámetro de contagio k .

Los diferentes métodos propuestos para la estimación del parámetro k , presentan algunas limitaciones de tipo estadístico como son incremento en varianzas y sesgo, dependiendo de los valores k , m y del tamaño de las muestras, por lo cual, ninguno de ellos puede considerarse como óptimo. El método de los momentos, por su sencillez puede resultar recomendable sobre todo cuando $\hat{m} \leq \hat{k}^2$.

Para aplicar el análisis de la varianza a los resultados de un muestreo o experimento, se requiere transformar la variable de tal modo que se cumplan los supuestos básicos de normalidad y varianza constante. En el caso de la Binomial Negativa, la varianza es una función de la media. Sobre el particular Beall (1942) propuso la transformación

$$z = f(x) = \text{sen } h^{-1} \sqrt{\frac{x}{k}}$$

que homogeniza las varianzas. Esta transformación fue complementada posteriormente por Anscombe (1948) mediante la función

$$y = \text{sen } h^{-1} \sqrt{\frac{x + 3/4}{k - 3/8}}$$

Un análisis numérico de esta expresión permitió comprobar que sólo resultaba efectiva para valores grandes de m (la media del número de insectos) y cuando $k > 2$. Para el caso de $k < 2$, propuso la expresión

$$y = \ln \left(x + \frac{k}{2} \right)$$

que también se hace más eficiente a medida que m aumenta.

Para el caso de los promedios de una Distribución Binomial Negativa, Gómez y Carrillo (1980) encontraron la transformación

$$y = 2\sqrt{nk} \text{ sen } h^{-1} \sqrt{\frac{z}{k}}$$

donde Z representa el promedio procedente de una muestra simple aleatoria de tamaño n .

Esta función estabiliza varianzas para cualquier valor de k haciéndose más eficiente a medida que n aumenta. Estudios numéricos comprobaron que el

ajuste a la Distribución Normal y la estabilización de varianzas era muy satisfactorio aún para valores pequeños de n (v. gr. $n = 10$).

La transformación propuesta puede ser útil para analizar esquemas de muestreo donde los promedios generan una serie de tiempo o en ensayos donde la unidad experimental está compuesta por varias subunidades que responden a la Binomial Negativa.

Por otro lado, se ha encontrado que el parámetro de contagio k varía dependiendo del tamaño de la unidad de muestreo utilizada, por lo cual, es recomendable unificar este criterio para estudios relacionados con la misma especie de organismo. Además, en algunos casos, se ha observado que el parámetro varía según el estado de desarrollo del individuo, por lo cual, su uso estaría restringido a condiciones de desarrollo similares.

De todos modos, entre las distribuciones de contagio, la Binomial Negativa es la que más ampliamente se ha ajustado para describir la disposición espacial de los insectos y sus propiedades ofrecen interesantes posibilidades para la descripción y análisis de la dinámica de estas poblaciones, teniendo en cuenta las limitaciones antes mencionadas.

VI. BIBLIOGRAFIA

- ALVAREZ, R.J. 1971. Estudio sobre la distribución espacial de la conchuela del fríjol *Epilachna varivestis* bajo condiciones de invernadero. Tesis M.C. Colegio de Postgraduados. Chapingo. 89 p.
- ANSCOMBE, F.J. 1948. The transformation of Poisson, Binomial and Negative Binomial data. *Biometrika*. 35: 246-254.
- — — — —. 1949. The statistical analysis of insects counts based on the Negative Binomial Distribution. *Biometrics*. 5: 165-173.
- BARTKO, J. 1961. The Negative Binomial Distribution. A review of properties and applications. The Department of Statistics and the Statistical Laboratory. Virginia Polytechnic Institute. Reprint. Series No. 190. 37 p.
- BEALL, G. 1942. The transformation of data from entomological field experiments so that the analysis of variance becomes applicable. *Biometrika*. 29: 243-262.
- BLISS, C.I. 1953. Fitting the Negative Binomial Distribution to biological data. *Biometrics* 9: 176-200.
- — — — —. 1956. The analysis of insect counts as Negative Binomial Distribution. *Proc. Tenth. Int. Congr. of Ent.* 2: 1015-1081.
- BOWMAN, K.C. and SHENTON, L.R. 1966. Biases of estimators of the Negative Binomial Distribution. Report. No. ORNL-4005. Union Carbide Corporation. Oak Ridge. Tenn.
- FORD, R.P. and DIMOND, J.B. 1973. Sampling populations of pine leaf chermid. *Pineus pinifoliae* Can. Ent. 105 (10): 1265-1274.

- GOMEZ, L.H. y CARRILLO L.A. 1980. Una transformación para uso de series de tiempo en poblaciones de insectos. *Agrociencia* No. 40: 107-120.
- HARCOURT, D.G. 1965. Spatial pattern of the cabbage looper *Trichoplusia ni* on crucifers. *Ann. Ent. Soc. Amer.* 58: 89-94.
- KATTI, S.K. and GURLAND, J. (1962). Efficiency of certain methods of estimation for the Negative Binomial and Neyman Type a distributions. *Biometrika* 49: 215-226.
- LLOYD, M. 1967. Mean crowding. *Jour. Anim. Ecol.* 36: 1-30.
- PIETERS, E.P. GATES, C.E. MATIS, J.H. and STERLING, W.L. 1977. Small sample comparison of different estimators of Negative Binomial. *Biometrika* 33: 718-723.
- ROBLES, R.R. y CARRILLO L.A. 1971. Generación de la distribución Binomial Negativa e interpretación biológica de sus parámetros. *Agrociencia*. 6: 11-128.
- ROJAS, B.A. 1964. La Binomial Negativa y la estimación de la intensidad de plagas en el suelo. *Fitotecnia Latinoamericana*. 1: 27-36.
- SAFRANYIK, L. and RASKIE, A.G. 1970. Secuencial sampling plan of larvae of *Monochamus* sp in lodgepole pine logs. *J. Econ. Ent.* 63 (6): 1903-1906.
- SOUTHWOOD, T.R. 1966. Ecological methods with particular reference to the study of insect populations. Methuen and Co. Ltda. Londres. p. 23-31.
- WATERS, W.E. 1959. A quantitative measure of aggregation in insects. *J. Econ. Ent.* 52 (6): 1180-1184.
- — — — — and HENSON, W.R. 1959. Some sampling attributes of the Negative Binomial Distribution with special reference to forest insects. *Forest. Sci.* 5 (4): 397-412.
- WILSON, L.F. y GERARD, D.J. 1971. A new procedure for rapidly estimating european pine sawfly (Hymenoptera Diprionidae) populations levels in young pine plantations. *Can. Ent.* 103: 1315-1322.

