TAMAÑO DE MUESTRA PARA LA ESTIMACIÓN DEL BANCO DE SEMILLAS DEL SUELO

J. DORADO*, L. AMBROSIO** y J.P. DEL MONTE*

*Dpto. de Producción Vegetal: Botánica y Protección Vegetal, **Dpto. de Economía y Ciencias Sociales Agrarias, E.T.S.I. Agrónomos, Universidad Politécnica, Ciudad Universitaria, 28040 Madrid.

Resumen: El estudio del banco de semillas de malas hierbas en el suelo plantea el problema del tamaño muestral necesario para estimar, de forma satisfactoria, el número de semillas por unidad de superficie. La distribución espacial de las semillas de las especies arvenses se puede ajustar fundamentalmente a dos leyes de distribución estadística: Poisson y Binomial Negativa. A partir de estas leyes, y sin necesidad de normalización, es posible estimar el número de muestras a tomar en función del número medio de semillas por muestra y del grado de precisión y nivel de confianza deseados.

INTRODUCCIÓN

El estudio y conocimiento del banco de semillas permite estimar la infestación potencial de malas hierbas, y se estima evaluando el número de las mismas en una muestra del suelo. Una cuestión importante a resolver es la determinación del tamaño de muestra (**n**), la cual ha sido tratada, entre otros, por Goyeau y Fablet (1982), Barralis *et al.* (1986) y Zanin *et al.* (1989). Estos autores se interesan en la distribución espacial de las semillas de las especies que estudian, como cuestión previa a la determinación del tamaño de muestra; sin embargo, determinan dicho tamaño normalizando previamente los datos sin considerar dicha distribución.

En el presente trabajo se pretende resolver el problema de la determinación del tamaño de muestra, tomando en consideración la distribución espacial de las semillas del suelo, en función de un margen de error y un nivel de confianza establecidos. Resulta razonable pensar que, al ser el comportamiento de las especies arvenses diferente, también lo será su modelo de distribución espacial. Estas diferencias están producidas por condiciones, tanto intrínsecas como extrínsecas, de la propia especie. Dentro de las primeras están: la capacidad reproductiva (especies con semillas pequeñas son más prolíficas en general que especies de semillas grandes), la facilidad de dispersión en el espacio (presencia de vilano para su dispersión por el viento, pelos o ganchos para adherirse a los animales, etc.) y la duración de la latencia y longevidad de las semillas en el suelo. Dentro de las condiciones extrínsecas se encuentran: las condiciones del medio (clima y recursos disponibles) y los factores antrópicos (laboreo, rotación de cultivos, control de malas hierbas, etc.).

Goyeau y Fablet (1982) y Zanin *et al.* (1989), coinciden en señalar que las especies cuya presencia es menor a una semilla por muestra, siguen un modelo de distribución aleatoria (Poisson) y aquellas cuya presencia es mayor a una semilla por muestra, poseen cierto grado de agregación (según Zanin *et al.*,1989, la distribución Binomial Negativa explica la mayoria de los casos); estas dos leyes estadísticas sirven también de modelo de distribución espacial en otros organismos vivos (Bliss y Fisher, 1953). Por otro lado, Goyeau y Fablet (1982) encuentran que, especies fuertemente representadas se ajustan a una distribución Normal. Todos los autores coinciden en que el análisis del potencial de semillas del suelo exige un muestreo siempre importante, tanto mayor cuanto menor sea la densidad de semillas enterradas en el suelo o mayor sea el grado de agregación de las mismas.

MATERIAL Y MÉTODOS

Se ha realizado el estudio del banco de semillas del suelo en la finca experimental "La Higueruela", perteneciente al Centro de Ciencias Medio-Ambientales (C.S.I.C.), dentro del término municipal de Santa Olalla (Toledo). El muestreo se ha efectuado en la primavera de 1993 y 1994, en 18 parcelas de 324 m^2 (9x36 m), tomando un total de 36 muestras por parcela, con una sonda de 4,6 cm de diámetro y 25 cm de profundidad, según una trama de 3x3 m. Las muestras se han procesado de forma individual, tamizando bajo chorro de agua la suspensión de suelo, a la que se añade hexametafosfato sódico como dispersante de las arcillas, con tamices de mallas de 2, 0,5 y 0,2 mm. Posteriormente se separan, identifican y enumeran las semillas aparentemente viables contenidas en la fracción recogida, utilizando microscopio estereoscópico.

En el análisis estadístico se plantean tres supuestos: ajuste de los datos a una distribución de Poisson (a), a una distribución Binomial Negativa (b) y a una distribución Normal (c). Para las dos primeras distribuciones se ha utilizado un test χ^2 de ajuste, Steel y Torrie (1985). En el caso de la distribución Normal se ha empleado el test de Shapiro y Wilks, Peña (1991).

RESULTADOS Y DISCUSIÓN

La norma general dentro de los suelos cultivados es encontrar una población de malas hierbas compuesta por: pocas especies dominantes que se encuentran presentes en alto número, pocas especies presentes con nivel moderado y gran variedad de especies presentes con un nivel bajo, Vengris (1953) citado por Wilson *et al.* (1985). En las 1.296 muestras tomadas durante los dos años, se han identificado 33 especies de malas hierbas, de las cuales 2 aparecen de forma dominante (número medio de semillas por unidad de muestreo = $\mathbf{m} > 15$), 4 presentan un nivel moderado ($1 < \mathbf{m} < 5$) y las otras 27 aparecen con un nivel bajo ($\mathbf{m} < 1$). Los datos de las dos especies mayoritarias (*Amaranthus albus* L. y *Polygonum aviculare* L.) y las cuatro intermedias (*Chenopodium album* L., *Heliotropium europaeum* L., *Papaver rhoeas* L. y *Portulaca oleracea* L.) se ajustan a la distribución Binomial Negativa. Dentro de las minoritarias, los datos de 13 especies se ajustan a la distribución de Poisson (*Amaranthus blitoides* L., *Anagallis arvensis* L., *Capsella bursa-pastoris* Moench., *Caucalis daucoides* L., *Convolvulus arvensis* L., *Draba verna* L., *Scandix pecten-veneris* L., *Silene gallica* L., *Spergula arvensis* L., *Spergularia rubra* L. J. et C. Presl., *Raphanus raphanistrum* L., *Veronica hederifolia* L. y *Veronica triphyllos* L.), no pudiendo comprobar los datos del resto de las especies por tener presencia muy baja ($\mathbf{m} \approx 0$).

En principio, parece existir una correspondencia entre **m** y el tipo de distribución: $\mathbf{m} < 1$ se corresponde con distribuciones de Poisson, y $\mathbf{m} > 1$ con Binomial Negativa, tal y como señalan Goyeau y Fablet (1982). No obstante, la distribución espacial está ligada fundamentalmente al comportamiento ecológico de cada especie, aunque los factores extrínsecos a la misma, pueden condicionar el valor de **m** y, de este modo, su modelo de distribución. Coincidiendo con los resultados de Zanin *et al.* (1989), no se encuentra ninguna especie cuyos datos se ajusten a una distribución Normal.

Una vez conocida la ley de distribución de la variable respuesta, se plantea determinar el tamaño de muestra **n** necesario para estimar **m**, con un margen de error relativo máximo **r**, y un nivel de confianza γ, esto es, estimar **n** de modo que se verifique:

$$P[|\bar{x}-m| \leq \delta] = \gamma \text{ , con } 0 < \gamma < 1 \qquad [1]$$

donde $\bar{x}$, la media muestral, es el estimador de **m**, y $\boldsymbol{\delta}$ = rm, con $0 < r < 1$

__Supuesto a__: ajuste a una distribución de Poisson. Si los datos se ajustan a una distribución de Poisson, la ecuación [1] toma la forma:

$$\sum_{z=nm(1-r)}^{z=nm(1+r)} e^{-nm} * (nm)^z / z! = \gamma \qquad [2]$$

donde **z** es un número entero. Fijados **r** y $\boldsymbol{\gamma}$, no es posible encontrar una solución simple del tamaño de muestra **n**, en función de **m** que satisfaga la ecuación [2]. Sin embargo, dicha ecuación es útil para encontrar soluciones particulares de **n**. En la tabla **1a** se muestran soluciones para una serie de valores de **m**, que abarca la práctica totalidad de las poblaciones naturales hasta ahora documentadas sobre banco de semillas del suelo.

Por el Teorema Central del Límite, cuando **n** ó **m** son grandes, la distribución de **n**$\bar{x}$ tiende a una ley Normal, de modo que la expresión [2] puede escribirse como:

$$2\,F[r(nm)^{1/2}] - 1 = \gamma \qquad [3]$$

en la que **F** es la función de distribución de una ley N(0,1): en la tabla **1a** se puede observar como para los valores de **m** y **n** considerados, la probabilidad definida en [1] y calculada según [2] y [3], es aproximadamente igual. La equivalencia entre estas ecuaciones permite utilizar la ecuación [3] en lugar de la [2]. A partir de la ecuación [3] es más fácil establecer una relación entre **n** y **m** para obtener una solución general al problema planteado, siempre en función de **m**:

$$n = U^2_{(1-\gamma)/2} / r^2 m \qquad [4]$$

donde $\mathbf{U_{(1-\gamma)/2}}$ es el cuantil de orden $(1-\gamma)/2$ en la distribución N(0,1).

Con tamaño de muestra **n** dado por [4], por satisfacer la ecuación [1], se puede afirmar, con un nivel de confianza $\boldsymbol{\gamma}$, que el error relativo en la estimación de **m**, $|\bar{x}-m|/m$, es inferior a **r**. Por otra parte, dada una muestra de tamaño **n** definida por [4], en la que se ha observado un media muestral $\bar{x}$, se puede afirmar, con un nivel de confianza $\boldsymbol{\gamma}$, que el verdadero valor de **m** está comprendido entre $\bar{x}/(1+r)$ y $\bar{x}/(1-r)$: este es el intervalo de confianza para **m** con un nivel de confianza $\boldsymbol{\gamma}$.

__Supuesto b__: ajuste a una distribución Binomial Negativa. Si los datos se ajustan a una distribución Binomial Negativa, la ecuación [1] toma la forma:

$$\sum_{z=nm(1-r)}^{z=nm(1+r)} (z+ny-1)! / z!(ny-1)! * p^{ny} * (1-p)^z = \gamma \qquad [5]$$

donde **y** y **p** son los parámetros de los que depende dicha distribución (**p** se estima mediante: $\bar{x}/s^2$; **y** se estima mediante: $\bar{x}^2/(s^2-\bar{x})$; con $\mathbf{s^2}$ = cuasivarianza muestral). Al igual que en el **supuesto a**, fijados **r** y $\boldsymbol{\gamma}$, no es posible encontrar una solución **n** en función de **m** (y de los parámetros **y** y **p**). Sin embargo, si es posible encontrar soluciones particulares de **n**. En la tabla **1b** se muestran varias de estas soluciones para distintos valores de **m**, **y** y **p**.

Por el Teorema Central del Límite, para valores de **n** grande, la distribución de **n**$\bar{x}$ converge a una Normal y la ecuación [5] se puede expresar como:

$$2\,F[nmr/(nv_{\bar{x}})^{1/2}] - 1 = \gamma \qquad [6]$$

siendo $\mathbf{v_{\bar{x}}}$ la varianza poblacional: en la tabla **1b** se puede observar como para los valores de **m**, **n** y $\mathbf{v_{\bar{x}}}$ considerados, la probabilidad definida en [1] y calculada según [5] y [6], prácticamente coincide. La equivalencia entre estas ecuaciones permite utilizar la expresión [6] en lugar de la [5]. A partir de la expresión [6] se tiene:

$$n = 1/p * U^2_{(1-\gamma)/2} / r^2 m \qquad [7]$$

Como en el supuesto a, para tamaños de muestra **n** dados por [7], se puede afirmar con un nivel de confianza $\boldsymbol{\gamma}$, que el error relativo en la estimación de **m** es inferior a **r**; asímismo, con **n** definida por [7], el verdadero valor de **m** está comprendido entre $\bar{x}/(1+r)$ y $\bar{x}/(1-r)$: este es el intervalo de confianza para **m** con un nivel de confianza $\boldsymbol{\gamma}$.

__Supuesto c__: ajuste a una distribución Normal. En este caso, y con varianza desconocida, el número de muestras a tomar viene dado por la expresión:

$$n = t^2 s^2 / \delta^2 \qquad [8]$$

donde t es el cuantil de orden $(1-\gamma)/2$ de la distribución t de Student, con n-1 grados de libertad (aproximadamente igual que $U_{(1-\gamma)/2}$ para n grande).

En la tabla 1 se recogen las soluciones particulares para especies que se distribuyen según: a) la ley de Poisson en función de m, con un nivel de confianza γ=0,95 y una precisión r=20%; b) la ley Binomial Negativa en función de m, y y p, con r=20%, y el número de muestras necesarias para γ=0,95.

CONCLUSIONES

Corroborando los resultados de GOYEAU y FABLET (1982) y ZANIN *et al.* (1989), es posible establecer modelos estadísticos para la distribución espacial de las semillas de las malas hierbas. Para las especies consideradas, estos modelos estadísticos son: la distribución de Poisson o la distribución Binomial Negativa.

La consideración de estos modelos estadísticos permite la deducción de expresiones para la determinación del tamaño de muestra n, más simples y eficientes que las propuestas en la revisión bibliográfica (basadas en la normalización previa de los datos).

Para un límite máximo del error relativo del 20% y un nivel de confianza del 95%, en la estimación del número medio de semillas por unidad de muestreo (sonda de 4,6 cm de diámetro y 25 cm de profundidad), se adjuntan tablas que pueden ser de utilidad para determinar el tamaño de muestra n, en la estimación del banco de semillas del suelo de zonas de cultivo.

REFERENCIAS BIBLIOGRÁFICAS

BLISS, C.I.; FISHER, R.A. (1953). Fitting the negative binomial distribution to biological data, *Biometrics* 9: 176-200.

BARRALIS, G.; CHADOEUF, R.; GOUET, J.P. (1986). Essai de détermination de la taille de l'échantillon pour l'étude du potentiel semencier d'un sol. *Weed Research* 26: 291-297.

GOYEAU, H.; FABLET, G. (1982). Etude du stock de semences de mauvaises herbes dans le sol: le problème de l'échantillonnage. *Agronomie* 2(6): 545-552.

PEÑA, D. (1991). *Estadística. Modelos y métodos. Fundamentos.* Alianza Universidad Textos. Madrid

STEEL, R.G.D.; TORRIE, J.H. (1985). *Bioestadística. Principios y procedimientos.* De. McGraw-Hill. México. 622 pp.

WILSON, R.G.; KERR, E.D.; NELSON, L.A. (1985). Potential for using weed seed content in the soil to predict future weed problems. *Weed Science* 33: 171-175.

ZANIN, G.; BERTI, A.; ZUIN, M.C. (1989). Estimation du stock semencier d'un sol labouré ou en semis direct. *Weed Research* 29: 407-417.

__Summary: Sample size to estimate the seed bank in the soil__. The study of weed seed bank in the soil states the question of sample size necessary to estimate the number of seeds per area in a satisfactory way. The seed spatial distribution of arvensis species can be adjusted to Poisson and negative binomial distributions mainly. Based on these, and without normalization, it is possible to estimate the number of samples taken in function of number of seeds average by sample, precession needed and confidence level wanted.

Tabla 1. Soluciones particulares del tamaño de muestra **n** para especies que se distribuyen según: **a)** Ley de Poisson, en función del número medio de semillas por unidad de muestreo (**m**), con un nivel de confianza $\gamma=0{,}95$ y una precisión $r=20\%$; **b)** ley Binomial Negativa, en función de **m** y de los parámetros de la distribución **y** y **p**, con $r= 20\%$, y una comparación con los valores de **n** obtenidos en una aproximación a la ley Normal ($\gamma=0{,}95$).

a)

Poisson			
			Aprox.Normal
m	n	γ[2]	γ[3]
0,1	900	0,9492	0,9422
0,2	450	0,9492	0,9422
0,3	300	0,9492	0,9422
0,4	225	0,9492	0,9422
0,5	180	0,9492	0,9422
1	90	0,9492	0,9422
1,5	60	0,9492	0,9422
2	45	0,9492	0,9422
2,5	36	0,9492	0,9422
3	30	0,9492	0,9422
3,5	27	0,9497	0,9481
4	22	0,9517	0,9394
4,5	20	0,9492	0,9422
5	18	0,9492	0,9422
6	15	0,9492	0,9422
7	13	0,9479	0,9436
8	11	0,9517	0,9394
9	10	0,9492	0,9422
10	9	0,9492	0,9422

m = n° semillas / muestra

n = tamaño muestral

Poisson

z=nm(1+r)

γ[2] $\sum e^{-nm} * (nm)^z / z!$

z=nm(1-r)

γ[3] $2F[r(nm)^{1/2}]-$

Binomial Negativa

s^2 = varianza muestral

p = se estima por $\bar{x}/s^2$

y = se estima por $\bar{x}^2/s^2-\bar{x}$

z=nm(1+r)

$\gamma[5]=\sum (z+ny-1)!/z!(ny-1)!*p^{ny}*(1-p)^z$

z=nm(1-r)

γ[6] $2F[nmr/(nv_x)^{1/2}]-1$

n(0,95)= n según ec.[7] y γ=0,95

b)

Binomial Negativa							
						Aprox.	Normal
m	v_x	p	y	n	γ[5]	γ[6]	n (0,95)
0,	0,	0,63	0,83	120	0,7995	0,7793	305
0,	1	0,50	0,50	150	0,7962	0,7793	384
0,	2	0,25	0,17	210	0,7085	0,6945	768
1	1,	0,67	2,00	50	0,7762	0,7518	143
1	2	0,50	1,00	75	0,7962	0,7793	192
1	4	0,25	0,33	105	0,7085	0,6945	384
1,	2	0,75	4,50	26	0,7630	0,7206	85
1,	3	0,50	1,50	50	0,7962	0,7793	128
1,	6	0,25	0,50	70	0,7085	0,6945	256
2	3	0,67	4,00	25	0,7762	0,7518	72
2	4	0,50	2,00	40	0,8077	0,7941	96
2	8	0,25	0,67	54	0,7237	0,7013	192
2,	4	0,63	4,17	24	0,7995	0,7793	61
2,	5	0,50	2,50	30	0,7962	0,7793	77
2,	10	0,25	0,83	42	0,7085	0,6945	154
3	4	0,75	9,00	10	0,6974	0,6572	43
3	6	0,50	3,00	25	0,7962	0,7793	64
3	12	0,25	1,00	35	0,7085	0,6945	128
4	6	0,67	8,00	10	0,7292	0,6983	36
4	8	0,50	4,00	20	0,7942	0,7941	48
4	16	0,25	1,33	27	0,7237	0,7013	96
5	7	0,71	12,50	8	0,7465	0,7150	27
5	10	0,50	5,00	15	0,7962	0,7793	38
5	20	0,25	1,67	21	0,7085	0,6945	77
6	8	0,75	18,00	6	0,7224	0,7013	21
6	12	0,50	6,00	10	0,7482	0,7267	32
6	24	0,25	2,00	18	0,7217	0,7013	64
7	10	0,70	16,33	6	0,7290	0,7218	20
7	14	0,50	7,00	10	0,7815	0,7633	27
7	28	0,25	2,33	15	0,7085	0,6945	55
8	12	0,67	16,00	6	0,7858	0,7421	18
8	16	0,50	8,00	10	0,8077	0,7941	24
8	32	0,25	2,67	12	0,6829	0,6728	48
9	12	0,75	27,00	4	0,7224	0,7013	14
9	18	0,50	9,00	8	0,7749	0,7699	21
9	36	0,25	3,00	11	0,6998	0,6803	43
10	14	0,71	25,00	4	0,7455	0,7150	14
10	20	0,50	10,00	5	0,7083	0,6827	19
10	40	0,25	3,33	9	0,6731	0,6572	38