¿Cuántas especies hay?

Los estimadores no paramétricos de Chao

Tania
Escalante Espinosa

El número exacto de especies que habitan la Tierra es aún un tema de gran debate, y la respuesta real a la pregunta del título va más allá de una simple suposición. En la segunda mitad del siglo XX algunos investigadores, inspirados en los estudios de Anne Chao de la National Tsing Hua University en Taiwán, han intentado resolver esta y otras preguntas estadísticamente.

Uno de los problemas clásicos en estudios ecológicos –y que obviamente pueden aplicarse en demografía–, es estimar el número de clases diferentes que hay en una población; por ejemplo: el número de especies de mamíferos en un estado del país, el número de individuos de cierta edad en una población, etcétera. El problema resulta trivial cuando extraemos un individuo de la población que tiene la misma probabilidad de pertenecer a cualquiera de todas las posibles clases, y el problema se reduce a una inferencia que incluye un solo parámetro,[1] donde se pueden emplear procedimientos tradicionales como el de máxima probabilidad (*maximum likelihood*).

Algunos investigadores[2] propusieron que la mayoría de los indicadores que se usaban en esos momentos para estimar el número de especies (riqueza) de una comunidad, a partir de una muestra, eran inapropiados. Uno de los métodos que se comenzó a utilizar con más frecuencia para conocer la riqueza de especies total de una

comunidad fueron las curvas de acumulación de especies. Estas curvas muestran el número de especies acumuladas conforme se va aumentando el esfuerzo de recolecta en un sitio, de tal manera que la riqueza aumentará hasta que llegue un momento en el cual por más que se recolecte, el número de especies alcanzará un máximo y se estabilizará en una asíntota (Figura 1). Pero incluso en estas curvas podrían obtenerse asíntotas antes de que muchas especies hubieran sido registradas, sobre todo por efecto de la estacionalidad, la diversidad beta (el grado de reemplazo de especies a través de gradientes ambientales)[3] y la abundancia relativa de las especies. Esto último constituyó un hallazgo importante, ya que no todos los individuos tienen la misma probabilidad de pertenecer a una especie determinada, puesto que hay especies comunes y especies muy raras.

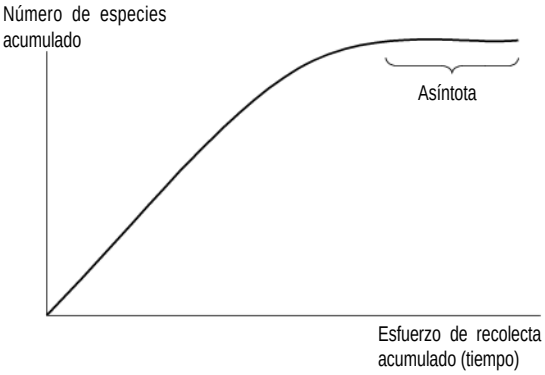


FIGURA 1. Curva de acumulación de especies. El número de especies registradas en una zona aumenta conforme aumenta el trabajo de campo, hasta un máximo donde se piensa que ya se han registrado todas las especies (asíntota).

En general, los métodos para estimar la riqueza de especies y la estructura de una población pueden dividirse en dos grupos: los métodos paramétricos y los no paramétricos. Los métodos paramétricos se llaman así porque parten de supuestos acerca de la población (por ejemplo: que la muestra sea aleatoria, que la probabilidad de cada clase sea la misma, que las medidas sean independientes), y por lo tanto requieren que los datos se distribuyan de cierta forma (por ejemplo, con una distribución normal). Entre los modelos paramétricos usados para estimar la riqueza específica están las funciones de acumulación, como la logarítmica, exponencial y la ecuación de Clench. Los modelos paramétricos que miden la estructura son, entre otros, la serie geométrica, la serie logarítmica, la distribución log-normal y el modelo de vara quebrada.[4] Por otro lado, los modelos no paramétricos han sido llamados también libres de distribución (*distribution-free)* porque los datos no asumen un tipo de distribución particular ni una serie de supuestos *a priori* que los ajusten a un modelo determinado. El cálculo de los modelos no paramétricos es más sencillo y rápido, son más fáciles de entender y explicar, y son relativamente efectivos. Los principales modelos no paramétricos que se han empleado para la estimación de la riqueza son *jacknife* de 1ro. y 2do. orden, *bootstrap* y el desarrollado por Anne Chao, el Chao2. En cuanto a los no paramétricos de estructura, están el Chao1 y el estadístico Q.[4]

Desde 1984,[1] Anne Chao intentó resolver el problema de la estimación del número de clases, haciéndolo equivalente a la estimación del número total de individuos en estudios de captura-recaptura. Su modelo consideraba un método basado en *bootstrap* para construir intervalos de confianza. Posteriormente probó modelos de heterogeneidad.[5] En 1992, junto con Lee,[6] usó un modelo llamado cobertura de muestra, el cual se define como la suma de las probabilidades de las clases observadas, y es muy útil en poblaciones heterogéneas. En 1993, Chao[7] intentó desarrollar reglas para pruebas en *software*, basadas en procedimientos de recaptura.

De todos los modelos que ha desarrollado Chao, el más utilizado desde entonces ha sido el propuesto en 1984.[1] Es quizá a partir del trabajo clásico de Colwell y Coddington[8] cuando el uso de los estimadores de Chao se generalizó. Entonces aumentó considerablemente el número de trabajos que intentan extrapolar de lo conocido hacia lo que aún no conocemos. Colwell y Coddington propusieron dos variantes con sus respectivos nuevos nombres para el estimador de Chao:[1] Chao1 para el estimador basado en abundancias y Chao2 para el estimador basado en incidencia.

A pesar de la facilidad de la aplicación de los estimadores de Chao, su uso se volvió problemático al tener una gran cantidad de muestras. En 1997 en un programa de cómputo llamado EstimateS, se implementaron los algoritmos de Chao1 y Chao2. Actualmente EstimateS[9] cuenta con otros estimadores de riqueza, algunos basados en la cobertura, calcula índices de diversidad, curvas de rarefacción, e incluso puede graficar curvas de acumulación de especies, entre otras bondades. EstimateS es gratuito a través de internet (http://viceroy.eeb.uconn.edu/EstimateS) y ha permitido que el empleo de los estimadores no paramétricos se generalice a nivel mundial.

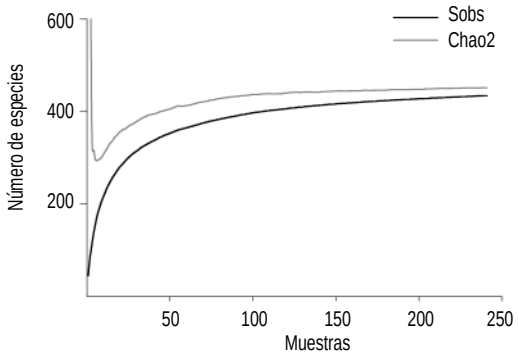


FIGURA 2. Gráfica del número de especies observadas y estimadas para los mamíferos terrestres de México en 241 muestras. En ella se aprecia que el número estimado es muy semejante al observado (extremo derecho de ambas curvas).

CHAO 1

Se ha llamado Chao1 al estimador de Chao[1] basado en la abundancia. Esto quiere decir que los datos que requiere se refieren a la abundancia de individuos que pertenecen a una determinada clase en una muestra. Una muestra es cualquier lista de especies en un sitio, localidad, cuadrante, país, unidad de tiempo, trampa, etcétera.

Como sabemos, hay muchas especies que sólo están representadas por pocos individuos en una muestra (especies raras), comparadas con las especies comunes, que pueden estar representadas por numerosos individuos. El estimador de Chao1 se basa en la presencia de las primeras. Es decir, requerimos saber cuántas especies están representadas por sólo un individuo en la muestra (*singletons*), y cuántas especies están representadas por exactamente dos individuos (*doubletons*): $S_{est} = S_{obs} + F^2/2G$, donde: S_{est} es el número de clases (en este caso, número de especies) que deseamos conocer, S_{obs} es el número de especies observado en una muestra, F es el número de *singletons* y G es el número de *doubletons*.

En EstimateS se ha integrado además una fórmula corregida para este modelo, la cual se aplica cuando el número de *doubletons* es cero: $S_{est} = S_{obs} + ((F^2/2G+1) - (FG/2(G+1)^2))$.

CHAO 2

Chao2 es el estimador basado en la incidencia. Esto quiere decir que necesita datos de presencia-ausencia de una especie en una muestra dada, es decir, sólo si está la especie y cuántas veces está esa especie en el conjunto de muestras: $S_{est} = S_{obs} + (L^2/2M)$, donde: L es el número de especies que ocurren sólo en una muestra (especies "únicas"), y M es el número de especies que ocurren en exactamente dos muestras (especies "dobles" o "duplicadas"). Por ejemplo, si tenemos un conjunto de cuadrículas, necesitamos saber cuántas especies están en una cuadrícula y cuántas especies están en dos.

La fórmula corregida en EstimateS, que se aplica cuando el número de dobles es cero, es: $S_{est} = S_{obs} + ((L^2/2M+1) - (LM/2(M+1)^2))$.

Para usar ambos estimadores en EstimateS se necesitan los datos en forma de matriz, donde los renglones y las columnas pueden representar indistintamente las muestras y las especies; se requiere establecer el orden una vez iniciado el programa. También EstimateS permite el cálculo de la desviación estándar de los dos estimadores.

Una vez que se realizan varias aleatorizaciones (se recomienda 50, pero pueden ser 100 o más), con o sin reemplazo, y cuando se ha empleado el número total de muestras, se obtiene el valor final del estimador y se pueden graficar los resultados. El número de muestras se presenta en el eje de las x, y el número de especies en la variable dependiente. Así, se pueden comparar la S_{est} y la S_{obs}. Pero la gráfica final se interpreta de manera distinta a la convencional: cuando se tiene el número total de muestras, existe cierta separación entre la curva de la S_{est} y de la S_{obs}. Esa separación estaría indicando cuántas especies faltan por registrar en esa comunidad. Entre más separadas estén, podríamos esperar que el número total de especies que contenga el lugar sea mayor que el que actualmente conocemos.

APLICACIONES

Los estimadores no paramétricos se han empleado para diversas aplicaciones, entre ellas, estudios con macroinvertebrados bentónicos, conejos, himenópteros, arañas, en selvas, entre otros. Los investigadores pueden hacer de estos estimadores una herramienta muy útil para saber si se requiere realizar un segundo estudio en una zona, aun cuando se obtenga una asíntota en la curva de acumulación de especies, e incluso pueden ser importantes en términos de costos, para reducir o aumentar la intensidad de muestreo.

Escalante y colaboradores[10] han usado estos métodos para saber cuántas especies más de mamíferos terrestres podemos encontrar en nuestro territorio. En la figura 2 se muestra una gráfica obtenida en EstimateS, donde se presenta la curva del estimador y el número de especies observadas.

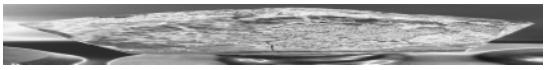
Al extremo derecho de la gráfica, con el número de muestras completo, es evidente que ambas curvas son muy próximas. Después de analizar los resultados encontramos que, en realidad, ya casi tenemos el inventario completo. Esto no quiere decir que ya conocemos todos los mamíferos de México, pero podemos establecer sitios estratégicos a escalas menores donde los inventarios no se hayan completado, y la S_{est} sea muy diferente de la S_{obs}. El uso de los estimadores no paramétricos aumentará y, sin duda, las contribuciones de Anne Chao en este campo constituyen un parteaguas en los estudios de extrapolación de la riqueza de especies.

El estudio de las comunidades naturales, incluidas las humanas, requiere de un conocimiento de su dinámica, lo cual raramente se logra, pero los conocimientos actuales, aún incompletos, pueden ser de gran ayuda al extrapolarlos por medio de estos métodos.

N O T A S

1 Chao, A., "Nonparametric estimation of the number of classes in a population", *Scandinavian Journal of Statistics,* núm. 11, 1984, pp. 256-270.

2 Heck, K. L. Jr., G. van Belle y D. Simberloff, "Explicit calculation of the rarefaction diverstity measurement and the determination of sufficient sample size", *Ecology,* núm. 56, 1975, pp. 1459-1461.



3 Whittaker, R. H., "Evolution and measurement of species diversity" ,*Taxon,* núm. 21, 1972, pp. 213-251.

4 Moreno, C. E., *Métodos para medir la biodiversidad*, M & T – Manuales y Tesis SEA, vol. 1, 2001, Zaragoza.

5 Chao, A., "Estimating the population size for capture-recapture data with unequal catchability" ,*Biometrics,* núm. 43, pp. 783-791.

6 Chao, A. y S. M. Lee, "Estimating the number of classes via sample coverage", *Jounal of American Statistical Association,* núm. 417, 1992, pp. 210-217.

7 Chao, A. M. C. Ma y M. C. K. Yang, "Stopping rules and estimation for recapture debugging with unequal failure rates" ,*Biometrika*, 80(1), 1993, pp. 193-201.

[8] Colwell, R. K. y J. A. Coddington, "Estimating terrestrial biodiversity through extrapolation" *, Philosophical Transaction of the Royal Society of London, Series B,* núm. 345, 1994, pp. 101-118.

[9] Colwell, R. K., *EstimateS: Statistical estimation of species richness and shared species from samples, Version 6.01b*, User's guide and application, 2000, http://viceroy.eeb.uconn.edu/estimates.

[10] Escalante, T., D., Espinosa y J. J. Morrone, "Patrones de distribución geográfica de los mamíferos terrestres de México", *Acta Zoológica Mexicana (n. s.),* núm. 87, 2002, pp. 47-65.

Tania Escalante Espinosa, Museo de Zoología, Departamento de Biología Evolutiva, Facultad de Ciencias, UNAM, tania_escalante@correo.unam.mx

© **Enrique Soto,** Museo Nacional, Praga, 2003.