

# Del nicho ecológico al nicho de las enfermedades: el método de Máxima Entropía (MaxEnt)

MAURICIO CANALS LAMBARRI\*

PROGRAMA DE SALUD AMBIENTAL, ESCUELA DE SALUD PÚBLICA  
DEPARTAMENTO DE MEDICINA INTERNA  
FACULTAD DE MEDICINA, UNIVERSIDAD DE CHILE

## 1. Introducción

En Ecología, el concepto de nicho se concibió como el conjunto de variables ambientales que permitían la existencia y sobrevivencia de una especie (Grinnell, 1917), definición conocida como Grinnelliana debido a su autor. Aunque este concepto se amplió gracias a las contribuciones de Elton (1927) y Hutchinson (1957), incorporando las relaciones con otras especies y una visión multidimensional del nicho (nicho hipervolumen), el concepto original sigue teniendo mucho valor cuando se le quiere extender a la comprensión de la distribución geográfica de una especie.

La modelación ecológica del nicho o de distribución potencial de especies ha tenido un explosivo desarrollo (Hirzel & Le Lay, 2008; Hirzel et al., 2006; Mateo et al., 2011; Peterson et al., 2011). Esta modelación puede ser considerada como un conjunto de representaciones o mapas de la idoneidad de un espacio para la presencia de una especie en función de las variables ambientales. Existen diferentes tipos de modelos o técnicas de modelación. Hay algunas que son descriptivas y otras que necesitan datos de presencia y ausencia de la especie en diferentes puntos o celdas como árboles de clasificación, correspondencia canónica, redes neuronales y modelos lineales generalizados (GLM), entre otras (Mateo et al., 2011).

Surge de manera casi natural extender el concepto de nicho Grinnelliano a las enfermedades infecciosas. Después de todo estas son producidas o transmitidas por especies biológicas que tienen preferencias ambientales para su reproducción y sobrevivencia. Así, podríamos definir el *nicho de una enfermedad* como *el conjunto de variables ambientales que permiten o explican la presencia y persistencia de una enfermedad en un espacio determinado*. También tiene sentido extender el concepto a las enfermedades no infecciosas, principalmente a aquellas que tienen un claro componente ambiental como contaminantes y material particulado, entre otras. En epidemiología espacial, los mapas se han utilizado con dos objetivos principales: (i) la elaboración de mapas de riesgo estáticos para caracterizar la variación espacial del riesgo contemporáneo, y (ii) mapas espacio-temporales retrospectivos de la dinámi-

ca epidémica para comprender los factores que rigen los patrones espaciales y la tasa de propagación de las enfermedades, o para describir las características de las ondas viajeras en epidemias como el sarampión y la fiebre hemorrágica del dengue (Ostfeld et al., 2005). Existen numerosas aplicaciones recientes de mapas en la distribución de diversas enfermedades infecciosas, como el síndrome pulmonar por hantavirus, la leishmaniasis, la encefalitis de La Crosse, la oncocercosis, la esquistosomiasis, las enfermedades transmitidas por garrapatas (Ostfeld et al., 2005), la malaria (Parham & Michael, 2010) y el dengue (Bhatt et al., 2013).

Así, conceptualmente, los mapas de riesgo epidemiológico de enfermedades infecciosas se pueden clasificar en tres categorías: (i) mapas ecológicos, basados en la distribución de vectores y el reservorio, (ii) mapas ecoepidemiológicos (como los mapas de número reproductivo  $R_0$ ), basados en las variaciones geográficas de los parámetros de la población de la interfaz vector-reservorio-enfermedad, y (iii) mapas epidemiológicos, basados en las variaciones geográficas de la incidencia de la enfermedad (riesgo basado en la incidencia) (Ostfeld et al., 2005). Estos tres tipos de mapas pueden ser considerados como estimadores del nicho Grinnelliano potencial de una enfermedad, que a su vez describen las zonas de riesgo de estas enfermedades.

## 2. Método de Máxima Entropía, con raíces en la teoría de la información, la física y la estadística

Entre los modelos de nicho Grinnelliano más usados en ecología destaca la modelación de Máxima Entropía (MaxEnt), que sólo necesita datos de presencia de una especie en un lugar determinado, generando sus propias ausencias, lo que lo ha popularizado en su uso. Esto es importante ya que lo que habitualmente llamamos ausencia puede ser una ausencia real o bien, falta de búsqueda o de datos.

El método de MaxEnt consiste en encontrar la mejor distribución, o de máxima entropía (o máxima ignorancia), sujeta a las restricciones impuestas por las

\*mcanals@uchile.cl.

variables ambientales obtenidas de la distribución empírica (Baldwin, 2009; Phillips et al., 2004, 2006).

### Fundamento

Considerando el conjunto  $X = (x_1, x_2, \dots, x_m)$  de presencias de una especie o enfermedad y el conjunto  $F = (f_1, f_2, \dots, f_n)$  de variables ambientales, se define  $p(x)$  como la distribución de probabilidad de  $X$ . Si  $f_j$  asigna un valor fijo  $f_j(x)$  a cada  $x$  en  $X$  (por ejemplo, temperatura en Santiago). Entonces el valor esperado de  $f_j$  bajo  $p$  es:

$$\langle f_j \rangle = \sum_{x \in X} p(x) f_j(x) = p(f_j). \quad (1)$$

El valor empírico de  $f_j$  corresponde al promedio de la variable  $f_j$  en las presencias (por ejemplo, el promedio de temperaturas donde está una especie)

$$\tilde{p}(f_j) = \sum_{i=1}^m \frac{f_j(x_i)}{m}. \quad (2)$$

Entonces, se busca un estimador de  $p(x)$  que se aproxime a la distribución empírica sujeto a las restricciones impuestas por las variables ambientales ( $F$ ). Se propone entonces que un buen estimador será aquel que maximice la «incerteza» o entropía. Este concepto proviene de la física (Boltzmann, 1872) y de la teoría de la información (Shannon, 1948). Para Boltzmann, la energía perdida por un sistema en forma de energía no utilizable o entropía ( $S$ ) es, en mecánica estadística, proporcional al número de microestados que explican un macro estado. Así planteó su famosa ecuación (que se encuentra grabada en su tumba)  $S = k \ln W$ , donde  $k = 1,38 \times 10^{-23} \text{ J}^\circ\text{K}$  es la constante de Boltzmann y  $W$  es el número de estados equiprobables (distribución uniforme). Cuando no todos los estados son equiprobables, esta se expresa como  $S = -k \sum_i p_i \ln p_i$ . Obviamente  $S$  es máxima si la distribución de probabilidades es homogénea:  $S = -k \sum_1^W (1/W) \ln(1/W) = -k \ln(1/W) = k \ln W$ . Esta conceptualización lleva a interpretar la entropía como una medida del «desorden». Shannon, en cambio, trabajaba en incerteza en los códigos, notando que estas tienen propiedad logarítmica, es decir, sumar información reduce multiplicativamente la incerteza. Así la definió como  $H = -\sum_i p_i \ln p_i$ . Entonces la información contenida en un mensaje sería igual al cambio en la incerteza antes y después del mensaje: Información =  $\Delta H$ . Es evidente la analogía entre entropía e incerteza, y prontamente Shannon comenzó a llamarla así (Shannon, 1971).

Entonces, la idea es maximizar

$$H(p) = - \sum_{x \in X} p(x) \ln p(x), \quad (3)$$

sujeto a las restricciones

- i)  $\sum_x \ln p(x) = 1$  (debe ser una distribución de probabilidades), y
- ii)  $\sum_{x \in X} p(x) f_j(x) = \tilde{p}(x)$  (debe estar sujeto a las restricciones ambientales impuestas por la distribución empírica).

Introduciendo multiplicadores de Lagrange se puede demostrar que la solución a este problema es encontrar un estimador dado por una distribución de Gibbs. Introducimos multiplicadores de Lagrange  $\lambda_0$  para la normalización y  $\lambda_j$  para cada restricción. El Lagrangiano es:

$$\mathcal{L}(p, \lambda) = - \sum_x p(x) \ln p(x) - \lambda_0 \left( \sum_x p(x) - 1 \right) - \sum_j \lambda_j \left( \sum_x p(x) f_j(x) - \tilde{p}(x) \right).$$

Derivamos respecto a  $p(x)$  y hacemos cero:

$$\frac{\partial \mathcal{L}}{\partial p(x)} = -(\ln p(x) + 1) - \lambda_0 - \sum_j \lambda_j f_j(x) = 0.$$

Reordenando, tenemos

$$\ln p(x) = -1 - \lambda_0 - \sum_j \lambda_j f_j(x),$$

y

$$p(x) = \exp \left( -1 - \lambda_0 - \sum_j \lambda_j f_j(x) \right).$$

Definiendo  $C := \exp(-1 - \lambda_0)$ , entonces

$$p(x) = C \exp \left( - \sum_j \lambda_j f_j(x) \right).$$

Pero  $p(x)$  debe cumplir que  $\sum_x p(x) = 1$ , entonces

$$C \sum_x \exp \left( - \sum_j \lambda_j f_j(x) \right) = 1. \quad (4)$$

Definiendo la función de partición

$$Z(\lambda) := \sum_x \exp \left( - \sum_j \lambda_j f_j(x) \right),$$

entonces  $CZ(\lambda) = 1 \Rightarrow C = \frac{1}{Z(\lambda)}$ . Así

$$p(x) = \frac{1}{Z(\lambda)} \exp \left( - \sum_j \lambda_j f_j(x) \right).$$

Como la entropía es una función cóncava en  $p(x)$ , este punto es un máximo. A esta distribución la conocemos como la distribución de Gibbs de la mecánica estadística, cuyo caso clave es una sola función  $f(x) = E(x)$  (la energía del estado) con la restricción  $\sum_x p(x)E(x) = \langle E \rangle$ . Entonces la solución MaxEnt se vuelve  $p(x) = (1/Z) \exp(-\lambda E(x))$ . En física habitualmente se escribe  $\lambda = \beta = 1/(K_b T)$ , donde  $K_b$  es la constante de Boltzman y  $T$  la temperatura. Así la distribución de Gibbs es la distribución de máxima entropía, conocida la energía promedio.

### 3. Relación con máxima verosimilitud (MV) y la divergencia de Kullback-Leibler ( $D_{KL}$ )

En términos generales hemos encontrado una distribución  $p(x)$  (de máxima entropía) que estima una distribución empírica  $q(x)$ . Una manera de comparar distribuciones es a través de la divergencia  $D_{KL}$ :

$$D_{KL}(p \parallel q) = \sum_x p(x) \ln \frac{p(x)}{q(x)}.$$

Aplicando las propiedades del logaritmo obtenemos la importante relación

$$D_{KL}(p \parallel q) = -H(p) - \sum_x p(x) \ln q(x). \quad (5)$$

Si  $q(x)$  es uniforme, es decir, máxima ignorancia, entonces  $\ln q(x)$  es constante:  $D_{KL}(p \parallel q) = -H(p) + \text{constante}$ . Por lo tanto, maximizar la entropía es equivalente a minimizar  $D_{KL}$ . En el caso que la distribución no sea uniforme, esta relación es también válida, pudiendo recuperarse la distribución de Gibbs a través de la minimización de  $D_{KL}$  mediante multiplicadores de Lagrange para las restricciones.

Además, si para el conjunto de datos  $x_1, x_2, \dots, x_n$  definimos la distribución empírica  $\hat{p}(x) = (1/n) \sum_i \delta(x - x_i)$  (equivalente a la distribución de frecuencias) y consideramos el modelo (MaxEnt)  $p_\lambda(x) = (1/Z(\lambda)) \exp(-\sum_j \lambda_j f_j(x))$ , el logaritmo de la verosimilitud se puede escribir como  $\ell(\lambda) = \sum_{i=1}^n \ln p_\lambda(x_i)$ . Observamos que si un valor  $x$  aparece  $n_x$  veces en la muestra, entonces:

$$\sum_{i=1}^n \ln p_\lambda(x_i) = \sum_x n_x \ln p_\lambda(x).$$

Como la frecuencia empírica es  $\hat{p}(x) = n_x/n$ . Sustituyendo  $n_x = n\hat{p}(x)$  obtenemos:

$$\begin{aligned} \sum_x n_x \ln p_\lambda(x) &= \sum_x n\hat{p}(x) \ln p_\lambda(x) \\ &= n \sum_x \hat{p}(x) \ln p_\lambda(x), \end{aligned}$$

que es exactamente  $\ell(\lambda) = n \sum_x \hat{p}(x) \ln p_\lambda(x)$ . Ahora, usando

$$\sum_x \hat{p}(x) \ln p_\lambda(x) = -D_{KL}(\hat{p} \parallel p_\lambda) - H(\hat{p}),$$

tenemos que

$$\ell(\lambda) = -nD_{KL}(\hat{p} \parallel p_\lambda) + \text{constante}. \quad (6)$$

Esta relación nos muestra que maximizar la verosimilitud es equivalente a minimizar  $D_{KL}$ . Así, relacionándolo con MaxEnt, se puede establecer que encontrar la distribución de máxima entropía (Gibbs) es equivalente a minimizar  $D_{KL}$  y maximizar la verosimilitud. Así, en términos coloquiales, la distribución MaxEnt nos aleja lo menos posible de la ignorancia (uniforme), acercándonos lo mas posible a los datos con mínima  $D_{KL}$  y máxima verosimilitud (Bishop, 2006; Choi, 2020; Jaynes, 1957a, 1957b; Phillips et al., 2004, 2006).

Así, el método MaxEnt tiene un soporte estadístico muy sólido que permite de una manera simple y rápida la modelación del nicho expresada en mapas de distribución potencial. El método ha demostrado en la práctica excelentes predicciones de distribución de especies (Peterson et al., 2011).

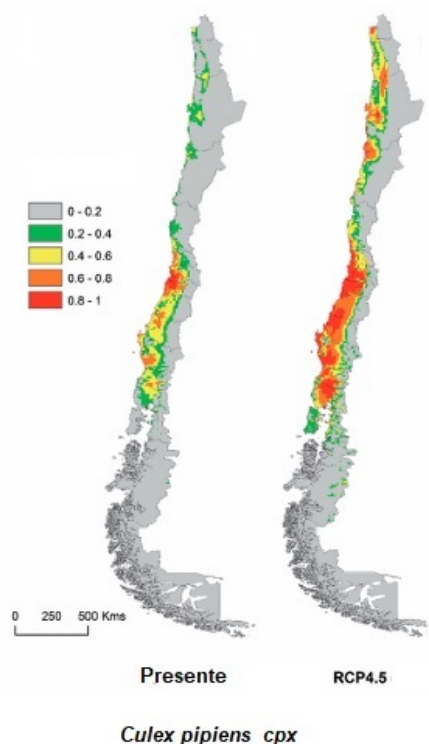
### 4. Ejemplos de aplicación en ecología y epidemiología

Veamos dos ejemplos que ilustran los conceptos enunciados y la utilidad del método, con componente ecológico y epidemiológico.

1. Veamos primero un ejemplo de ecología, que puede tener importancia epidemiológica futura. Figueroa et al. (2020) modelaron con MaxEnt y estudiaron las posibles variaciones del nicho potencial del zancudo común, del complejo *Culex pipiens* que en Chile incluye a las especies *C. pipiens* y *C. quinquefasciatus* (Figura 1). En este trabajo mostraron que ante diferentes escenarios de cambio climático, la distribución de estas especies se puede extender hacia el sur ya que aumenta su idoneidad climática. Esto cobra validez actual cuando se está reportando aumento de las poblaciones de mosquitos en Concepción (Abril, 2026). Por otra parte, esto puede tener importancia epidemiológica en el caso del ingreso a Chile de la encefalitis por WN virus, cuyo vector es precisamente esta especie.
2. Tapia-Garay et al. (2018) estudiaron las zonas de riesgo de la enfermedad de Chagas en Chile bajo la influencia de diferentes escenarios de cambio climático (Figura 2). En este estudio se determinó con el método MaxEnt, por una parte el nicho

potencial de la vinchuca *Triatoma infestans*, vector principal de la enfermedad de Chagas, basados en datos de presencia de esta especie, y, por otra parte, con el mismo método, el «nicho» de la enfermedad de Chagas basados en el registro de los casos reportados.

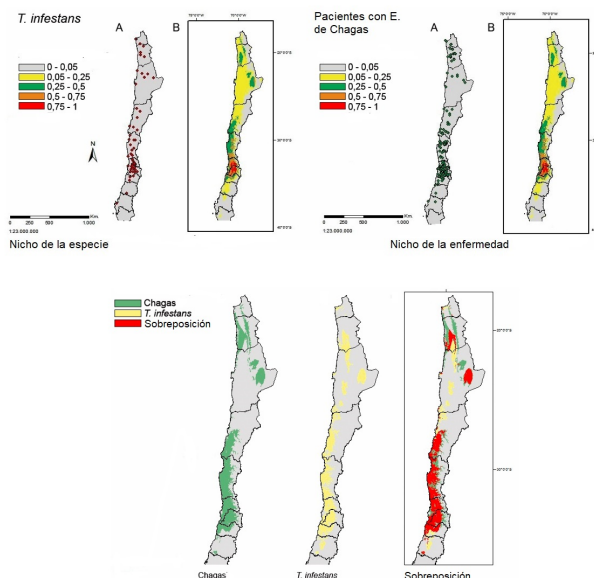
En este estudio se demostró que si bien ambos nichos son muy parecidos, al superponerlos se encontraron zonas en que la especie *T. infestans* se encuentra presente pero no se reportan casos de la enfermedad de Chagas, y otras zonas en que se encuentran enfermos pero no el vector *T. infestans*. Las primeras zonas se pueden interpretar como áreas de sub-reporte y las segundas pueden corresponder a áreas donde los otros vectores (*Mepraia gajardoi* y *M. spinolai*) son los responsables de la enfermedad, o bien a migraciones internas.



**Figura 1.** Modelación con MaxEnt, del nicho potencial actual del complejo *Culex pipiens* (*Culex pipiens* *cpx*: *C. pipiens* y *C. quinquefasciatus*) y ante un escenario de cambio climático RCP 4.5 en Chile, con MaxEnt. Modificado de Figueroa et al. (2020).

En resumen, en este artículo defino el concepto de nicho de una enfermedad y establezco la equivalencia con el nicho Grinnelliano de las especies. Muestro el método de Máxima Entropía (MaxEnt) que permite modelar la distribución espacial de especies y el riesgo de enfermedades bajo determinadas condiciones ambientales. Muestro la solidez del método y sus raíces

en la física, estadística y teoría de la información. Específicamente presento la relación entre divergencia de Kullback-Leibler, máxima verosimilitud y máxima entropía. Finalmente presento dos ejemplos de nuestro equipo de trabajo, donde se utilizan estos conceptos y la metodología.



**Figura 2.** Modelación con MaxEnt del nicho potencial de *Triatoma infestans*, principal vector de la enfermedad de Chagas en Chile, y del nicho de la enfermedad. Abajo la superposición de ambos nichos. Modificado de Tapia-Garay et al. (2018).

## Referencias

- Baldwin, R. A. (2009). Use of Maximum Entropy Modeling in Wildlife Research. *Entropy*, 11(4), 854-866. <https://doi.org/10.3390/e11040854>
- Bhatt, S., Gething, P. W., Brady, O. J., Messina, J. P., Farlow, A. W., Moyes, C. L., Drake, J. M., Brownstein, J. S., Hoen, A. G., Sankoh, O., Myers, M. F., George, D. B., Jaenisch, T., Wint, G. R., Simmons, C. P., Scott, T. W., Farrar, J. J., & Hay, S. I. (2013). The global distribution and burden of dengue. *Nature*, 496(7446), 504-507. <https://doi.org/10.1038/nature12060>
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- Boltzmann, L. (1872). Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen. *Sitzungsberichte Akademie der Wissenschaften*, 66, 275-370. [https://doi.org/10.1142/9781848161337\\_0015](https://doi.org/10.1142/9781848161337_0015)

- Choi, S. (2020). Kullback matching and maximum likelihood estimation [technical report]. *Pohang University of Science and Technology*.
- Elton, C. (1927). *Animal ecology*. The MacMillan Company.
- Figueroa, D., Scott, S., González, C. R., Bizama, G., Flores-Mara, R., Bustamante, R., & Canals, M. (2020). Estimating the climate change consequences on the potential distribution of *Culex pipiens* L. 1758, to assess the risk of West Nile virus establishment in Chile. *GAYANA*, 84(1), 46-53. <https://doi.org/10.4067/S0717-65382020000100046>
- Grinnell, J. (1917). The Niche-Relationships of the California Thrasher. *The Auk*, 34(4), 427-433. <https://doi.org/10.2307/4072271>
- Hirzel, A. H., & Le Lay, G. (2008). Habitat suitability modelling and niche theory. *Journal of Applied Ecology*, 45, 1372-1381. <https://doi.org/10.1111/j.1365-2664.2008.01524.x>
- Hirzel, A. H., Le Lay, G., Helfer, V., Randin, C., & Guisan, A. (2006). Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, 199, 142-152. <https://doi.org/10.1016/j.ecolmodel.2006.05.017>
- Hutchinson, G. E. (1957). Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, 22, 415-427. <https://doi.org/10.1101/SQB.1957.022.01.039>
- Jaynes, E. T. (1957a). Information Theory and Statistical Mechanics. *Physical Review*, 106(4), 620-630. <https://doi.org/10.1103/PhysRev.106.620>
- Jaynes, E. T. (1957b). Information Theory and Statistical Mechanics. II. *Physical Review*, 108(2), 171-190. <https://doi.org/10.1103/PhysRev.108.171>
- Mateo, R. G., Felicísimo, A. M., & Muñoz, J. (2011). Modelos de distribución de especies: Una revisión sintética. *Revista Chilena de Historia Natural*, 84, 217-240. <https://doi.org/10.4067/S0716-078X2011000200008>
- Ostfeld, R. S., Glass, G. E., & Keesing, F. (2005). Spatial epidemiology: an emerging (or re-emerging) discipline. *Trends in Ecology & Evolution*, 20, 328-336. <https://doi.org/10.1016/j.tree.2005.03.009>
- Parham, P. E., & Michael, E. (2010). Modeling the effects of weather and climate change on malaria transmission. *Environmental Health Perspectives*, 118(5), 620-626. <https://doi.org/10.1289/ehp.0901256>
- Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M., & Araújo, M. B. (2011). *Ecological Niches and Geographic Distributions*. Princeton University Press.
- Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3-4), 231-259. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>
- Phillips, S. J., Dudík, M., & Schapire, R. E. (2004). A Maximum Entropy Approach to Species Distribution Modeling. En *Proceedings of the 21st International Conference on Machine Learning* (pp. 655-662).
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379-423, 623-656.
- Shannon, C. E. (1971). *Scientific American*, 225, 180.
- Tapia-Garay, V., Figueroa, D. P., Maldonado, A., Frías-Laserre, D., Gonzalez, C. R., Parra, A., Canals, L., Apt, W., Alvarado, S., Cáceres, D., & Canals, M. (2018). Assessing the risk zones of Chagas' disease in Chile, in a world marked by global climatic change. *Memórias do Instituto Oswaldo Cruz*, 113(1), 24-29. <https://doi.org/10.1590/0074-02760170172>