

Trabajo fin de máster
Sara Martínez Santalla
Máster en Técnicas Estadísticas, USC
Curso 2017-2018

English

Modelling of the distance decay of biological similarity

Biological communities tend to be more different in their species composition when the spatial distance between them is high. This pattern of gradual decrease in species composition similarity with spatial distance is called distance-decay of biological similarity and is considered to follow a negative exponential or power-law function. Several approaches have been suggested to model it, of which the most common are linear models and generalized linear models (GLMs). Given the nature of the data (i.e. pairwise comparison of biological communities in different localities), the response variable is a distance matrix of a pairwise index of biological similarity (Sørensen's or Simpson's) and thus the observations are not independent. This circumstance violates one of the main assumptions of GLMs and linear models, and must be taken into account when fitting them.

The aim of this work is to find a method for accurately fitting a GLM to the relationship between biological similarity and spatial distance data. GLMs are explored in this dissertation due to their simplicity and thus ease of interpretation from an ecological standpoint, while they allow to represent the functional forms proposed for the distance-decay. In particular, the main challenge of this study is to overcome the distortions in the GLM fitting process due to the lack of independence. With this aim, we have tested two kinds of bootstrap, the classical moving blocks bootstrap and a manual selection blocks bootstrap. In order to assess the performance of each bootstrap method, we have carried a simulation study covering the bias of the estimated parameters, the nominal level for the likelihood ratio test of the deviance and the comparison of AIC for candidate models. Finally, both bootstrap methods have been applied to the real data and the results have been compared. However, the main results have been inconsistent with the simulation. Thus, the actual focus of the research is going to be the improvement of the simulation and the reassessment of their performance for both bootstrap methods.

Modelado del descenso de la similitud biológica en función de la distancia espacial

Las comunidades biológicas tienden a diferenciarse más en su composición específica cuando la distancia espacial entre ellas es alta. Este patrón de descenso gradual de la similitud en la composición específica de las comunidades se denomina decaimiento de la similitud biológica con la distancia, y se trata como una función de tipo exponencial o power-law. Se han sugerido distintas

aproximaciones para modelarla, de las cuales las más comunes son los modelos lineales y los modelos lineales generalizados (Generalized Linear Models, GLMs). Dada la naturaleza de los datos (ej. comparación a pares de las comunidades biológicas de distintas localizaciones) la variable respuesta es una matriz de distancias de un índice pairwise de similitud biológica (el de Sørensen o el de Simpson) y por lo tanto las observaciones no son independientes. Esta circunstancia incumple una de las principales asunciones de los GLMs y los modelos lineales, y debe ser tenida en cuenta cuando se trabaja con ellos.

El objetivo de este trabajo es encontrar un método para ajustar de forma adecuada un GLM a la relación entre los datos de similitud biológica y distancia espacial. Se ha elegido trabajar con este modelo por su simplicidad y su fácil interpretación desde un punto de vista biológico, y porque permite representar las formas funcionales con las que se describe el decaimiento de la similitud biológica con la distancia. Particularmente el mayor reto del estudio es evitar las irregularidades en el proceso de ajuste de un GLM debidas a la falta de independencia. Con este objetivo hemos estudiado dos tipos de bootstrap, el bootstrap de bloques móviles (Moving Block Bootstrap) y un bootstrap con bloques seleccionados manualmente. Para evaluar el comportamiento de cada método se ha realizado un estudio de simulación que incluye el cálculo del sesgo de los parámetros, del nivel nominal para el test de razón de similitudes de la deviance, y la comparación basada en el AIC de los modelos candidatos. Finalmente, ambos métodos bootstrap han sido aplicados a los datos reales y se han comparado los resultados, encontrando inconsistencias con respect a las simulaciones. Por lo tanto, el foco actual del estudio se centra en mejorar la simulación de los datos y reevaluar la eficiencia de ambos métodos bootstrap.

Modelado do descenso da similitude biolóxica coa distancia espacial

As comunidades biolóxicas tenden a diferenciarse máis na súa composición específica cando a distancia espacial entre elas é alta. Este patrón de diminución gradual da similitude na composición específica das comunidades denomínase descenso da similitude biolóxica coa distancia, e trátase coma unha función de tipo exponencial o power-law. Na literatura súxírense distintas aproximacións para a súa modelización, das cales as máis comúns son os modelos lineais e os modelos lineais xeneralizados (Generalized Linear Models, GLMs). Dada a natureza dos datos (ex. comparación a pares das comunidades biolóxicas de distintas localizacións) a variable resposta é unha matriz de distancias dun índice pairwise de similitude biolóxica (o de Sørensen ou o de Simpson) e polo tanto as observacións non son independentes. Esta circunstancia incumple unha das principais asuncións dos GLMs e os modelos lineais, e debe ser tida en consideración cando se traballa con eles.

O obxectivo deste traballo é atopar un método para axustar de forma adecuada un GLM á relación entre os datos de similitude biolóxica e distancia espacial. Elexiuse traballar con este modelo pola súa simplicidade e fácil interpretación dende un punto de vista biolóxico, e porque permite representar as formas funcionais coas que se describe o descenso da similitude biolóxica coa distancia. Particularmente o maior reto do estudo é evitar as irregularidades no proceso de axuste dun GLM debidas á falta de independencia. Con este obxectivo estudamos dous tipos de bootstrap, o bootstrap de bloques móbiles (Moving Block Bootstrap) e un bootstrap con bloques seleccionados manualmente. Para avaliar o comportamento de cada método realizouse un estudo de simulación que inclúe o cálculo do sesgo dos parámetros, do nivel nominal para o test de razón de similitudes da deviance, e a comparación baseada no AIC dos modelos candidatos. Finalmente, ámbolos métodos bootstrap foron aplicados aos datos reais e comparáronse os resultados, atopándose inconsistencias con respecto á simulación. Polo tanto, os futuros obxectivos desta liña de estudo inclúen mellorar o funcionamento da simulación e reevaluar os resultados de ambos tipos de bootstrap.