

Cálculo recursivo de la probabilidad de presencia de secuencias en cadenas markovianas

Javier Marengo Tomás Tetzlaff

Instituto de Ciencias, Universidad Nacional de General Sarmiento, Argentina
{tetzlaff, jmarengo}@ungs.edu.ar

El problema de determinar la probabilidad de que una secuencia esté presente en un fragmento de un lenguaje aleatorio es común en diversas áreas. Particularmente en el caso de cadenas de ADN, la aparición por azar de una secuencia de bases dentro de cierta porción del genoma es una hipótesis nula que resulta de interés para cotejar alternativas en las cuales una presencia mayor de la secuencia indicaría la posibilidad de que cumpla un rol biológico relacionado con el rol de la porción del genoma. En este trabajo presentamos un procedimiento eficiente para calcular la probabilidad de que algún elemento de un conjunto de secuencias fijas esté presente en una cadena aleatoria de longitud dada. Dentro de la familia de cadenas generadas por un proceso de Markov con un número finito de estados, consideramos variantes de interés biológico para las reglas probabilísticas que definen las cadenas.