

Computación segura entre 2 partes de profundidades en estadística (2PC4SD)

Ignacio Cascos
ignacio.cascos@uc3m.es

12 de febrero de 2026

1. Computación segura entre 2 partes para el análisis de datos

En muchos entornos, los datos necesarios para estimar un modelo estadístico se encuentran distribuidos entre entidades que no pueden o quieren compartirlos por razones de confidencialidad. Los protocolos seguros de dos partes (*2-Party Computation*, 2PC) permiten que dos participantes, con datos privados D_1 y D_2 calculen conjuntamente el resultado de una función estadística $f(D_1, D_2)$ sin revelar información adicional. Para reducir el número de interacciones entre las partes, puede recurrirse a esquemas de cifrado homomórfico (*Homomorphic Encryption*, HE) que permiten realizar operaciones directamente sobre los datos cifrados, con el inconveniente de asumir un elevado coste computacional.

Aplicadas a la estimación conjunta de parámetros, estas técnicas posibilitan un análisis distribuido que preserva tanto la privacidad de los datos como la validez estadística de los resultados.

2. Profundidad simplicial

2.1. Introducción

En Estadística multivariante, el término *profundidad* hace referencia al grado de centralidad de un punto respecto de una nube en su misma dimensión. Si tenemos una nube de n puntos en el plano $\{(x_i, y_i)\}_{i=1}^n$ en posición general (no hay más de dos puntos alineados sobre la misma recta), podemos construir los $\binom{n}{3}$ triángulos cuyos vértices son tres puntos distintos de la nube, y la *profundidad simplicial* de cualquier punto (x, y) es la proporción de esos triángulos que contienen a (x, y) .

Alternativamente, podemos considerar las envolventes convexas de subconjuntos de puntos de la nube de un tamaño k fijado de antemano, digamos $k = 4, 5, 6, \dots$ y calcular la proporción de esos $\binom{n}{k}$ subconjuntos que contienen a (x, y) . La envolvente convexa de un conjunto es el menor conjunto convexo que lo contiene, es decir aquel que contiene los segmentos entre dos puntos cualquiera del mismo. Para un conjunto finito de puntos, podemos pensar que la ubicación de cada uno en el plano está marcada por una chincheta y la envolvente convexa es la región delimitada por una goma elástica que está sujeta por las chinchetas más externas y las contiene a todas.

El problema puede generalizarse a cualquier dimensión, si tenemos nubes de puntos en dimensión $d \geq 3$, el menor número de puntos para el que tiene sentido considerar su envolvente convexa es $k = d + 1$. Así, podemos considerar los $\binom{n}{d+1}$ subconjuntos de $d+1$ puntos y calcular la fracción de símplexes con esos puntos como vértices que contienen al punto de referencia.

El cómputo eficiente de profundidades es todo un reto y puede involucrar ideas sofisticadas de Geometría Computacional. Rousseeuw y Ruts propusieron en 1996 un algoritmo de complejidad en el tiempo $O(n \log n)$ para calcular la profundidad simplicial bivalente respecto de una nube de tamaño n , que luego se demostró que era óptimo en el sentido de la eficiencia y puede generalizarse para las envolventes convexas k puntos en el plano para cualquier $k \geq 3$. Puedes decargarte el artículo pinchando aquí Rousseeuw y Ruts (1996).

2.2. El Reto

Dos entidades disponen de sendos conjuntos de datos bidimensionales (puntos en el plano) de tamaños n_1 y n_2 . Para un punto fijado de antemano (por ejemplo el origen de coordenadas) desean calcular el número de triángulos cuyos vértices son tres puntos de la muestra conjunta que contienen dicho punto fijo, sin desvelar información alguna sobre los puntos (datos) de sus respectivas muestras. Desean además, generalizar el procedimiento para considerar las envolventes convexas de subconjuntos de $k \geq 3$ puntos.

Describe el procedimiento que han de seguir, usando las técnicas comentadas anteriormente, para conseguir su objetivo.