**INDEPENDENT COMPONENT ANALYSIS**

Tabla de contenido

[Usos 1](#_Toc82636457)

[Metodología 3](#_Toc82636458)

[Características 4](#_Toc82636459)

[Aplicación del algoritmo 5](#_Toc82636460)

[REFERENCIAS 5](#_Toc82636461)

Es una técnica estadística y computacional para revelar factores ocultos que subyacen a conjuntos de variables, medidas o señales aleatorias.

El método se originó a partir de la investigación de procesamiento de señales, donde se mezclan fuentes de señales desconocidas con un nuevo conjunto de señales.

# Usos

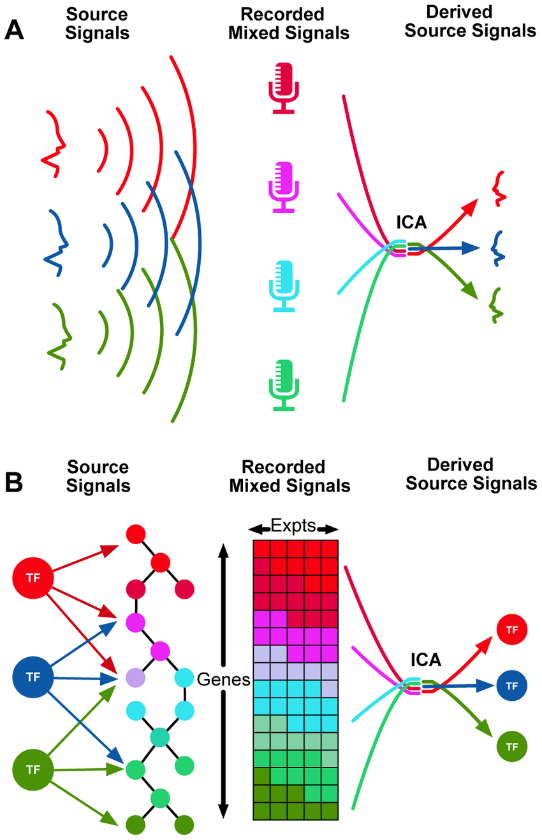
La forma más fácil en que lo entendí fue con el ejemplo del “Problema del cóctel”. En su forma más simple, imagine a dos personas conversando en un cóctel. Por alguna razón, tienes dos micrófonos colocados cerca de los asistentes a la fiesta.

Ambas voces son escuchadas por ambos micrófonos a diferentes volúmenes según la distancia entre la persona y el micrófono. En otras palabras, grabamos dos archivos que incluyen el audio de los dos asistentes a la fiesta mezclados.

El problema entonces es, ¿cómo podemos separar las dos voces en cada archivo para obtener grabaciones aisladas de cada hablante?

Este problema se resuelve fácilmente con el análisis de componentes independientes (ICA), que transforma un conjunto de vectores en un conjunto máximamente independiente.

Volviendo a nuestro "Problema de fiesta de cóctel", ICA convertirá las dos grabaciones de audio mezcladas en dos grabaciones sin mezclar de cada altavoz individual. Se puede observar que el número de entradas y salidas es el mismo, y dado que las salidas son mutuamente independientes.



# Metodología

Hay dos suposiciones clave hechas en ICA. Los componentes independientes ocultos que estamos tratando de descubrir deben ser uno, estadísticamente **independientes,** y dos, no gaussianos. Semánticamente, por independiente me refiero a que la información sobre x no te da información sobre y viceversa. Matemáticamente, esto se traduce en,



La suposición no gaussiana simplemente significa que los componentes independientes tienen distribuciones que no son gaussianas, lo que significa que no parece una curva de campana.

El primer supuesto es el punto de partida de ICA. Queremos desenredar la información para derivar un conjunto de factores independientes. Si no hay varios generadores independientes de información para descubrir, realmente no hay necesidad de ICA.

La necesidad del segundo supuesto radica en las matemáticas. ICA utiliza la idea de la no gaussianidad para descubrir componentes independientes. La no gaussianidad cuantifica qué tan lejos está la distribución de una variable aleatoria de ser gaussiana. Ejemplos de medidas de no gaussianidad son la curtosis y la neguentropía. La razón por la que esta medida es útil se deriva del Teorema del límite central. Específicamente, un resultado que establece que la suma de dos variables aleatorias independientes tiene una distribución más cercana a la de Gauss que cualquiera de las variables originales. ICA combina esta idea, medidas no gaussianas y la suposición no gaussiana para descubrir componentes independientes ocultos en los datos.

En otras palabras, la no gaussianidad de y se maximiza cuando es directamente proporcional a uno de los componentes independientes. Esto nos permite enmarcar ICA como un problema de optimización.

# Características

El método ICA tiene múltiples características, pero las más destacadas son:

* El número de entradas es igual al número de salidas
* Asume que los componentes independientes son estadísticamente independientes
* Asume que los componentes independientes no son gaussianos
* Las entradas deben ser valores auto escalados (restar cada columna por su media y dividir por su desviación estándar)

Se diferencia del método PCA por su diferencia de enfoque. EL PCA se usa para comprimir información (reducción de dimensionalidad) mientras que el ICA se encarga de separar la información.

# Aplicación del algoritmo

1. **Análisis financiero** (predecir los precios del mercado de valores)

**In econometrics,** we often have parallel time series, and ICA could decompose them into independent components that would give an insight to the structure of the data set

1. **Neurociencia** (clasificación de picos neuronales, análisis de los cambios en la expresión génica a lo largo del tiempo en experimentos de secuenciación de ARN de una sola célula, reconocimiento facial.)

**Bran Imaging** (we often have different sources in the brain emit signals that are mixed up in the sensors outside of the head, just like in the basic blind)

1. **OTROS** (image feature extraction, where we want to find features that are as independent as possible)

# REFERENCIAS

<https://towardsdatascience.com/independent-component-analysis-ica-a3eba0ccec35>

https://www.cs.helsinki.fi/u/ahyvarin/whatisica.shtml

http://scielo.sld.cu/scielo.php?script=sci\_arttext&pid=S1815-59282014000200007#e1