

CAPÍTULO 2

INTRODUCCIÓN A ICA Y PCA

2.1 Análisis de componentes principales

2.1.1 Introducción

El Análisis de Componentes Principales (PCA), también llamado transformada de Karhunen-Loève, es una técnica estadística para el tratamiento de datos de forma compacta que elimina la redundancia de los datos originales. Se trata pues de obtener una representación de los datos lo mejor posible. Las técnicas PCA encuentran numerosas aplicaciones en diversas materias que requieren el tratamiento de datos (especialmente para la compresión). En nuestro caso, será una herramienta especialmente útil para el procesado previo de datos para la aplicación de técnicas de Análisis de Componentes Independientes (ICA).

Dado un vector aleatorio \mathbf{x} con n componentes, del cual conocemos una muestra $\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(T)$, el objetivo de PCA consiste en tratar de eliminar la redundancia existente entre cada una de sus componentes de forma que se obtenga un nuevo vector \mathbf{y} con m componentes, tal que $n \geq m$. Con esto conseguimos un vector de menor dimensión, sin redundancia entre sus componentes.

En PCA, no se tiene en cuenta el tipo de función de densidad que poseen los vectores. Sólo supondremos que se pueden estimar los estadísticos de primer y segundo orden de la muestra.

PCA trata de buscar un nuevo sistema de ejes coordenados formado por una base ortogonal para representar al vector \mathbf{x} de forma que se maximiza la varianza de las

proyecciones del vector en dicho sistema. Así conseguimos que las componentes de \mathbf{x} en el nuevo sistema no estén correladas.

Para ilustrar el funcionamiento de PCA, en la siguiente figura se representa la distribución de probabilidad conjunta de dos variables aleatorias x_1 y x_2 obtenidas cada una de ellas a partir de la combinación lineal de otras dos variables aleatorias s_1 y s_2 independientes entre sí.

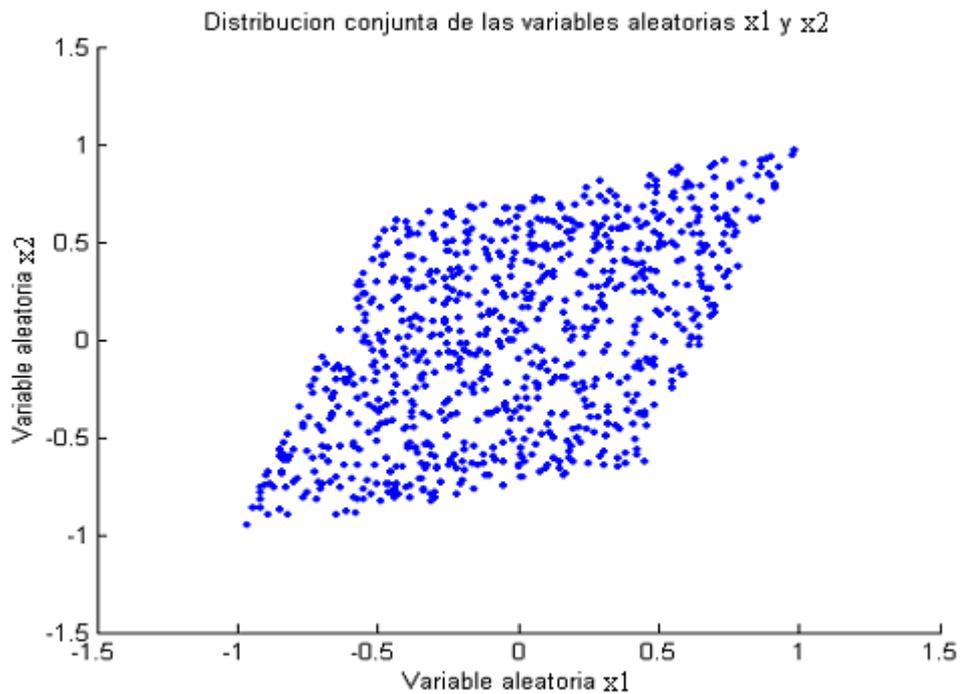


Figura 2.1: Distribución de probabilidad conjunta de la variables aleatorias x_1 y x_2 .

Si a las variables aleatorias x_1 y x_2 se les realiza un procesamiento PCA, las nuevas variables que se obtienen no estarán correladas entre sí y tendrán una distribución de probabilidad conjunta como se ilustra a continuación:

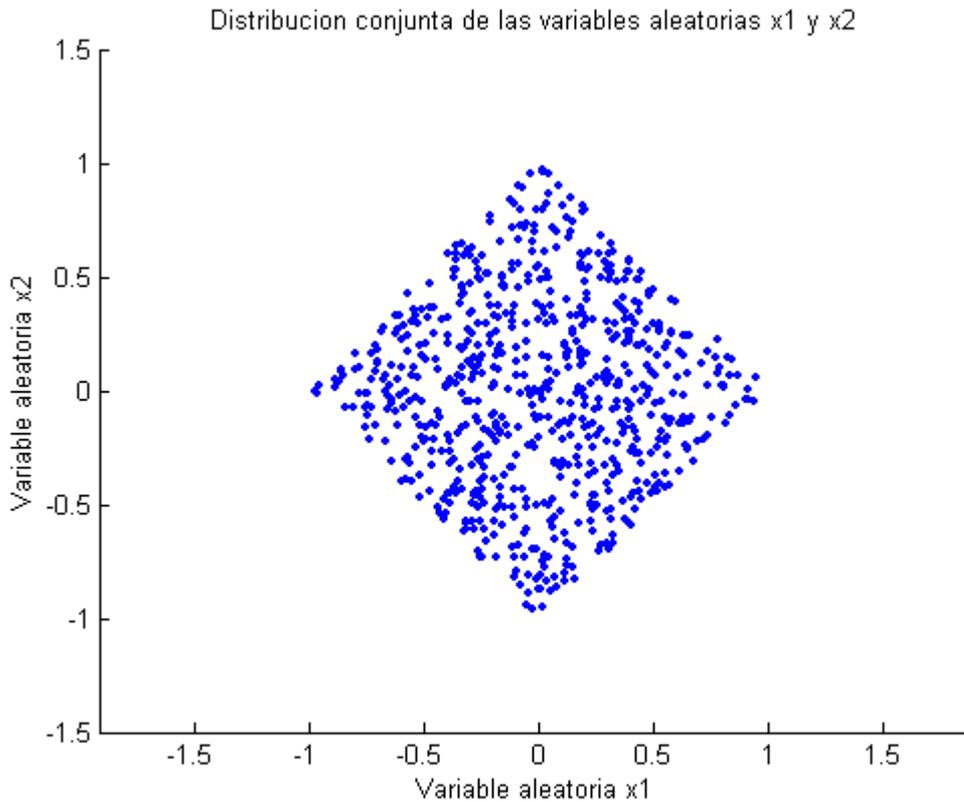


Figura 2.2: Nueva distribución de probabilidad conjunta de las variables tras un procesamiento PCA.

Tal como puede verse en la *figura 2.2* las nuevas variables obtenidas estarán no correladas entre sí y su función de distribución conjunta estará contenida en un cuadrado.

Las técnicas PCA persiguen objetivos distintos que las técnicas ICA. Sin embargo, se verá que PCA puede ser útil como un procesamiento previo al uso de técnicas ICA. También pueden usarse para blanquear o decorrelar los datos antes de aplicarles un determinado procesado.

2.1.2 Descripción del método

En PCA se trata de buscar las componentes principales de las observaciones que consisten en los autovalores de la matriz de covarianzas de éstas. Para poder aplicar el método, debemos centrar el vector \mathbf{x} . Para ello, simplemente, tenemos que sustraerle la media:

$$\mathbf{x} = \mathbf{x} - E\{\mathbf{x}\} \quad (2.1)$$

Denotaremos como \mathbf{y} la nueva representación de los datos. Se trata de un vector de dimensión m . En general, se cumplirá que $m \leq n$ siendo n la dimensión de los datos originales.

La nueva representación de los datos tras aplicar las técnicas PCA vendrá dada por:

$$\mathbf{y} = \mathbf{w}^T \mathbf{x} \quad (2.2)$$

Donde \mathbf{w} es una matriz cuyas filas son vectores de dimensión n , tales que maximizan la varianza entre las componentes del vector \mathbf{y} . Para dicha matriz sus columnas serán ortogonales entre sí, es decir:

$$\mathbf{w}_i^T \mathbf{w}_j = \delta_{ij} \quad (2.3)$$

Puede demostrarse, véase [8], que las columnas de la matriz \mathbf{w} coinciden con una base ortogonal expandida a partir de los autovectores $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_m$ de la matriz de covarianzas, \mathbf{C}_x , del vector \mathbf{x} . Dichos autovectores serán distintos entre sí y se ordenarán en función del valor de sus autovalores $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$. La mayor parte de la información estará contenida en el primer autovector \mathbf{e}_1 . La cantidad de información contenida en cada autovalor será mayor cuanto mayor sea su autovalor asociado, de forma que en las primeras componentes se concentrará la mayor parte de la información. Al valor $y_1 = \mathbf{e}_1^T \mathbf{x}$ se le denomina *componente principal* de \mathbf{x} , y es la que contiene más información de todas.

En el caso de que $m < n$, estaremos cometiendo un error en la representación de los datos. Como ventaja estaremos comprimiendo la información contenida en el vector inicial. El error cuadrático mínimo que podemos cometer se dará cuando tomemos una base ortogonal expandida a partir de los m autovectores $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_m$ de la matriz de covarianzas. Dicho error vendrá dado por:

$$J_{MSE}^{PCA} = \sum_{i=m+1}^n d_i \quad (2.4)$$

Si $m = n$ no se cometerá error alguno en la representación de los datos en la nueva base formada por los autovectores de la matriz de covarianzas de los datos. Este caso será de especial interés, ya que al aplicar un análisis PCA a los datos, conseguiremos decorrelarlos.

2.2 Análisis de componentes independientes

2.2.1 Introducción

El Análisis de Componentes Independientes (ICA) es una técnica pensada para obtener o extraer una representación de cada una de las distintas componentes o señales independientes entre sí que forman parte de una señal de mezcla. Veremos que esto será posible siempre que las distintas señales que componen la mezcla sean independientes entre sí; es decir, el valor de una de ellas no está relacionado con el de las otras, y que no más de una de ellas tenga una distribución de tipo gaussiana.

Entre las numerosas aplicaciones de ICA, nos centraremos en las aplicaciones para BSS (Separación ciega de fuentes) que consisten en la recuperación de un conjunto de señales a partir de mezclas lineales de ellas sin saber nada sobre las fuentes que queremos recuperar.

2.2.2 Planteamiento del modelo en presencia de ruido

Consideraremos que observamos m variables aleatorias en presencia de ruido a las que denotaremos como x_1, x_2, \dots, x_m . Estas variables aleatorias son el resultado de combinaciones de otras variables aleatorias, a las que denotaremos como s_1, s_2, \dots, s_n , que queremos obtener a partir de las variables que hemos observado.

En general, podremos expresar cada una de las variables aleatorias que hemos observado de la forma:

$$x_i = a_{i1}s_1 + a_{i2}s_2 + \dots + a_{in}s_n + n_i \quad \text{con } i = 1, \dots, m \quad (2.5)$$

Donde los coeficientes a_{ij} son coeficientes reales *a priori* desconocidos, con $i = 1, \dots, m$ y $j = 1, \dots, n$, y n_i es el ruido recibido en el i -ésimo receptor. En general, supondremos que las variables aleatorias s_i son estadísticamente independientes y que, como máximo, una de ellas tiene una distribución de tipo gaussiana. También supondremos que el número de variables aleatorias observadas cumple que $m > n$ en el caso de que las observaciones se hagan en presencia de ruido y que $m \geq n$ en ausencia de ruido.

En el presente documento se empleará notación vectorial para referirse a las fuentes y a las observaciones. Denotaremos como \mathbf{x} al vector que contiene cada una de las mezclas observadas x_1, x_2, \dots, x_m , como \mathbf{s} al vector que contiene cada una de las componentes independientes s_1, s_2, \dots, s_n que queremos separar y como \mathbf{n} al vector que contiene las fuentes de ruido recibidas en cada sensor. Asimismo denotaremos como \mathbf{A} la matriz de mezclas cuyos elementos son los coeficientes a_{ij} . De forma compacta, podemos expresar el proceso de mezcla de la siguiente forma:

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{n} \quad (2.6)$$

En la siguiente figura se ilustra gráficamente el proceso de mezcla y de estimación de las señales originales a partir de sus mezclas:

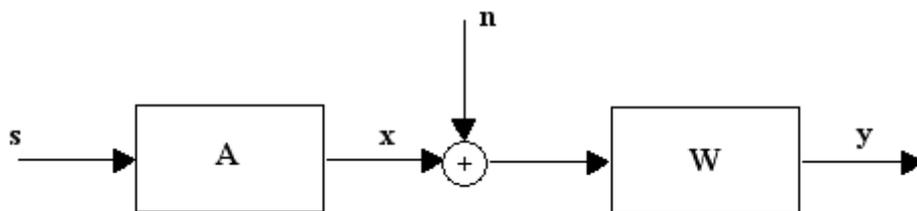


Figura 2.3: Esquema del proceso de mezclas y de separación

Nuestro objetivo será determinar las componentes independientes (Ics) a partir de las observaciones. Para ello, trataremos de buscar una estimación de inversa de la matriz de mezclas \mathbf{W} , que nos permita recuperar las señales originales a partir de las observaciones. Normalmente la matriz \mathbf{A} será invertible, aunque si el número de observaciones es mayor que el de fuentes (tendremos redundancia) la matriz de mezclas no será cuadrada. En estos casos, la identificación de las fuentes a partir de las mezclas también es posible y nos limitaremos a buscar una matriz que al multiplicarla por la matriz de mezclas nos dé la matriz identidad.

2.2.3 Ilustración del funcionamiento de los métodos ICA

Para mostrar de forma cualitativa el correcto funcionamiento de ICA, consideraremos dos fuentes estadísticamente independientes entre sí, s_1 y s_2 , distribuidas de forma uniforme en un cuadrado.

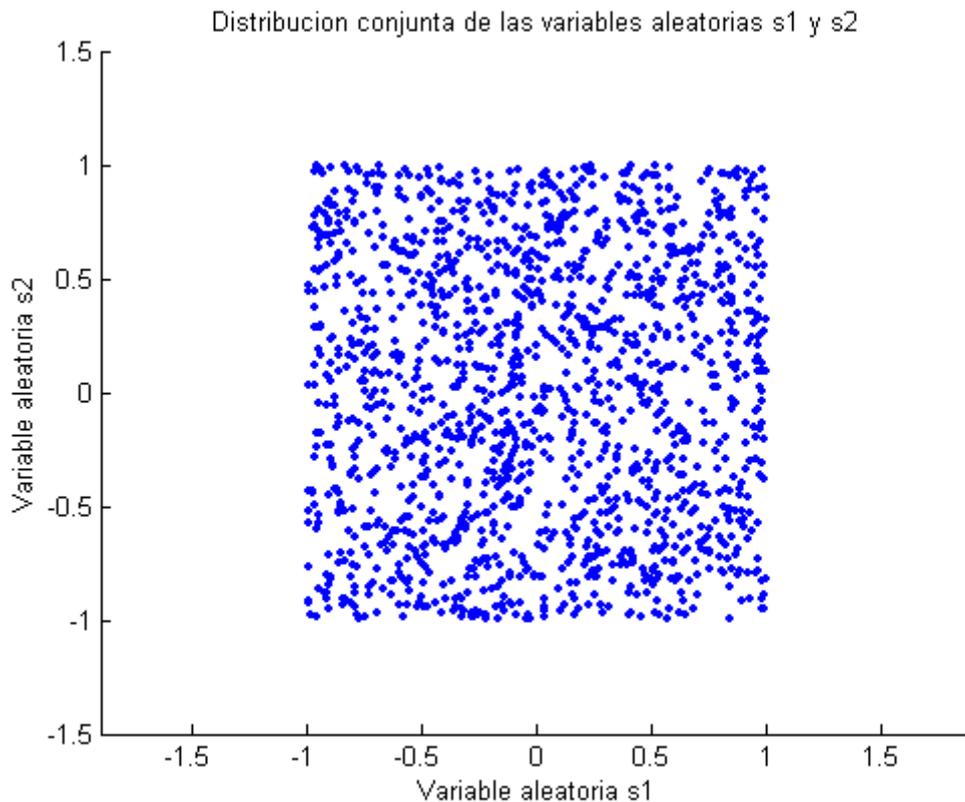


Figura 2.4: Distribución conjunta de las variables aleatorias s_1 y s_2 . Cada una de ellas con distribución uniforme. Eje horizontal s_1 . Eje vertical s_2

Podemos ver como las dos variables aleatorias son independientes entre sí. Si fijamos el valor de una de ellas, no podemos determinar el valor de la otra.

Ahora supongamos que mezclamos ambas variables aleatorias mediante una matriz de mezcla \mathbf{A} de dimensión 2×2 para obtener otras dos nuevas variables aleatorias x_1 y x_2 . Tal como puede verse en la distribución conjunta de ambas variables, las dos nuevas variables aleatorias no son independientes entre sí. Si fijamos el valor de una de ellas, el valor de la otra estará condicionado por el valor que tome la primera.

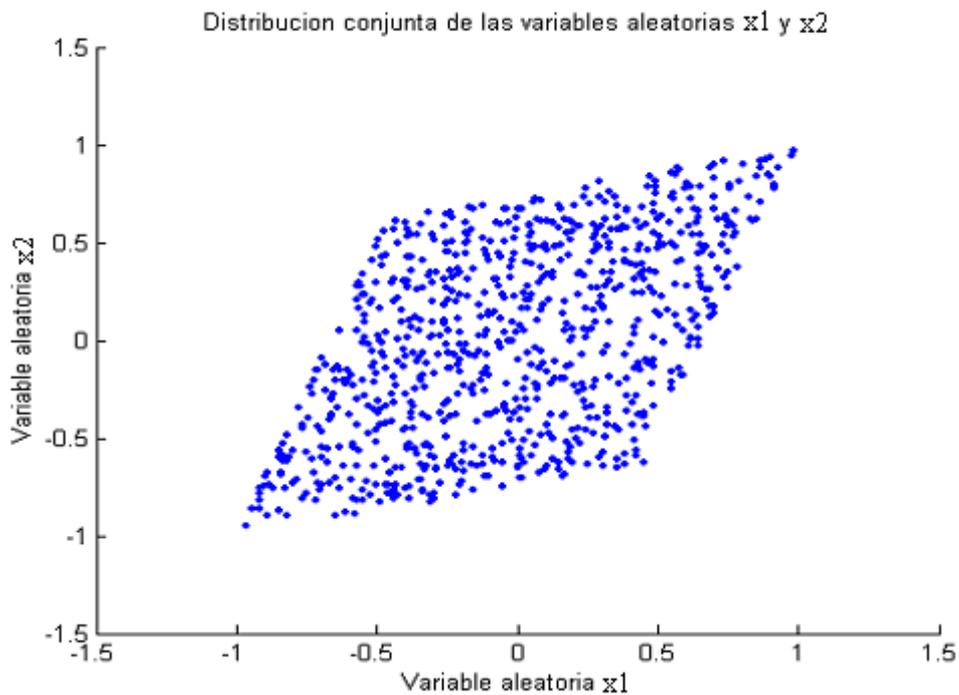


Figura 2.5: Distribución conjunta de las variables x_1 y x_2 . Eje horizontal: variable x_1 . Eje vertical: variable x_2

El problema de obtención de las componentes independientes originales debe hacerse a partir de las observaciones y consiste en estimar la matriz \mathbf{A} a partir de las observaciones. De forma intuitiva podemos ver que las direcciones del paralelepípedo de la *Figura 2.5* coinciden con las columnas de la matriz \mathbf{A} . Este método no es útil para separar las fuentes, sin embargo podemos ver de forma cualitativa que el problema de separación tiene solución si las variables de mezcla son estadísticamente independientes. En apartados posteriores, se comentarán en detalle algunos métodos analíticos para la obtención de las distintas ICs de una mezcla utilizados en el presente documento.

2.2.4 Restricciones de los métodos ICA

Existen varias limitaciones que poseen los métodos ICA y que es conveniente destacar. Estas limitaciones o condicionantes no son demasiado restrictivos y son asumibles en la mayoría de las aplicaciones de interés. A continuación se describen las principales restricciones de ICA:

- **Las componentes independientes deben ser estadísticamente independientes**

Como ya se vio en apartados anteriores, necesitamos que las variables aleatorias que se mezclan s_1, s_2, \dots, s_n sean independientes entre sí. Dos variables aleatorias son independientes entre sí cuando una de ellas no da información de la otra. Matemáticamente, dado un conjunto de variables aleatorias, se dice que son estadísticamente independientes si su función de densidad conjunta es igual al producto de las funciones de densidad marginales:

$$p(s_1, s_2, \dots, s_n) = p_1(s_1)p_2(s_2)\dots p_n(s_n) \quad (2.7)$$

Este condicionante, en la práctica, no resulta ser demasiado restrictivo: en la mayoría de las aplicaciones, las señales que sufren el proceso de mezcla son independientes. Así, en el caso que posteriormente se tratará de señales de voz de distintas personas, esta suposición se cumple.

- **Las componentes independientes deben tener distribuciones no gaussianas**

Para que la identificación de las distintas componentes sea posible, es necesario que tengan funciones de distribución no gaussianas. Con funciones de distribución gaussianas, la identificación de la matriz \mathbf{A} no es siempre posible. Para establecer las condiciones de separación cuando existen fuentes gaussianas, partiremos del siguiente teorema:

Teorema 2.1: *Teorema de Darmois-Skitovich:* Sean dos combinaciones lineales de fuentes estadísticamente independientes:

$$\begin{aligned} y_1[n] &= G_{11}s_1[n] + G_{12}s_2[n] \\ y_2[n] &= G_{21}s_1[n] + G_{22}s_2[n] \end{aligned}$$

Si $y_1[n]$ e $y_2[n]$ son independientes y $\{G_{1j} \neq 0, G_{2j} \neq 0\}$, entonces s_j es gaussiana.

Para variables gaussianas, el hecho de que no estén correladas implica independencia. Por tanto, la propiedad de independencia en la que se basa ICA no puede ser utilizada para la identificación de variables gaussianas. Así, en el caso de que la matriz \mathbf{A} sea ortogonal, bien por casualidad o porque hayamos preblanqueado los datos, la función de distribución conjunta de las fuentes originales y de las mezclas son idénticas, por lo que no puede obtenerse ninguna información adicional de las fuentes a partir de las mezclas y no podrá determinarse la matriz \mathbf{A} .

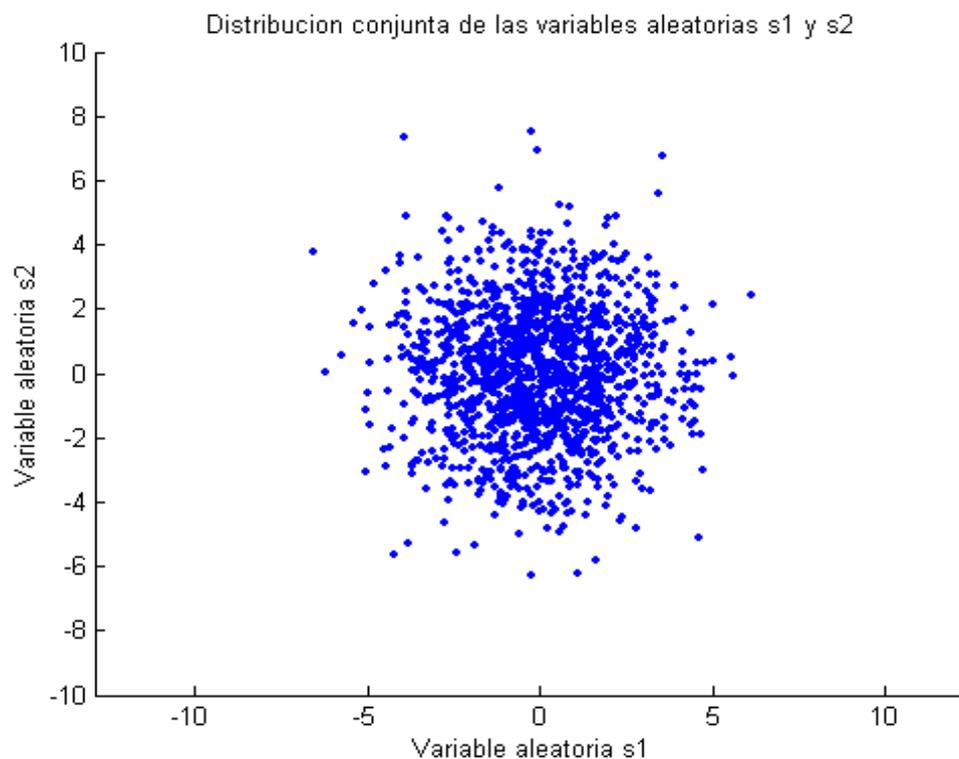


Figura 2.6: Representación de la función de distribución de dos variables aleatorias de tipo gaussiano. Dicha distribución coincide con la de las mezclas decorreladas de las fuentes gaussianas.

- **No es posible determinar las amplitudes de las componentes independientes ni el orden de las mismas.**

Las amplitudes no pueden determinarse porque las matrices \mathbf{A} y \mathbf{s} son desconocidas. Un factor que multiplique al vector de fuentes \mathbf{s} puede ser cancelado, dividiendo una de las columnas de \mathbf{A} por el mismo factor. Por tanto, ICA podrá determinar las Ics originales salvo una constante multiplicativa. Decir que dicha constante puede tener signo negativo e invertir las señales. Esto no supone una limitación para la mayoría de los casos de interés.

Tampoco puede determinarse el orden de las Ics. La razón es la misma de antes: desconocemos la matriz de mezclas y las fuentes originales. Podemos escribir las señales recibidas en los receptores de la forma $\mathbf{x} = \mathbf{A}\mathbf{P}^{-1}\mathbf{P}\mathbf{s}$, donde \mathbf{P} es una matriz de permutación. Entonces vemos que el valor de la nueva matriz de mezclas vendrá dado por $\mathbf{A}\mathbf{P}^{-1}$. Es decir, podremos determinar la matriz de mezclas salvo una permutación. Esta limitación no supondrá un problema, ya que normalmente nos centraremos en obtener todas las componentes independientes. Decir que mediante técnicas como la deflación (véase implementación del algoritmo de separación ciega con pequeños retardos) es posible extraer las señales originales ordenadas de mayor a menor energía.

2.2.5 Comparación entre ICA y PCA

Aunque ambos métodos están relacionados entre sí, persiguen objetivos distintos: el objetivo de PCA es decorrelar los datos y reducir el número de variables, mientras el de ICA es obtener por separado cada una de las componentes independientes que forman parte de una mezcla.

Por otra parte ICA usa el concepto de independencia que es más restrictivo que el de correlación, ya que si dos variables son independientes serán no correladas, pero no necesariamente al revés.

Dos variables aleatorias son no correladas si su covarianza es cero:

$$\text{cov}(y_1, y_2) = E\{y_1 y_2\} - E\{y_1\} E\{y_2\} = 0 \quad (2.8)$$

Y si dos variables aleatorias son independientes porque cumplen la condición (2.7), también serán no correladas porque en ese caso:

$$E(y_1 y_2) = E(y_1) E(y_2) \quad (2.9)$$

Aunque ambas técnicas persiguen objetivos distintos y se basan en propiedades distintas, no son excluyentes. Las técnicas PCA pueden usarse para preblanquear los datos antes de aplicarles un análisis ICA y simplificar así de forma considerable el problema. El preblanqueo de los datos consiste en multiplicar los datos recibidos por una matriz \mathbf{V} de forma que los nuevos datos recibidos no estén correlados entre sí. Los nuevos datos obtenidos tras preblanquear vendrán dados por:

$$\mathbf{z} = \mathbf{V}\mathbf{x} = \mathbf{V}\mathbf{A}\mathbf{s} \quad (2.10)$$

Y se cumplirá que $E\{\mathbf{z}\mathbf{z}^T\} = \mathbf{I}$, porque al preblanquear los nuevos datos que se obtienen ya no estarán correlados.

Con el preblanqueo de los datos se consigue que la matriz de mezclas, $\mathbf{V}\mathbf{A}$, sea ortogonal haciendo más sencilla su identificación, ya que sólo tenemos que buscar en el espacio de matrices ortogonales. Al ortogonalizar la matriz \mathbf{A} , se pasa de tener n^2 grados de libertad a tener $\frac{n(n-1)}{2}$ grados de libertad, por lo que se simplifica extraordinariamente el problema de identificación.

2.3 Resumen del capítulo

En el presente capítulo, se han analizado dos técnicas para el tratamiento de datos: PCA e ICA. Ambas técnicas persiguen objetivos distintos y están basadas en propiedades distintas. La primera de ellas está basada busca eliminar la redundancia de

un conjunto de datos intentando eliminar la correlación entre las distintas muestras que conforman los datos. La segunda de ellas está basada en la propiedad de independencia y busca la obtención de una serie de fuentes estadísticamente independientes entre sí a partir de una mezcla de ellas.

Pese a las diferencias que presentan ICA y PCA, no son técnicas excluyentes. En capítulos posteriores, se utilizarán las técnicas PCA para preblanquear los datos y facilitar la resolución del modelo ICA.