

# Vector aleatorio normal o distribución gaussiana multivariable

Dr. Andrés Altieri

Procesos Estocásticos  
Facultad de Ingeniería, Universidad de Buenos Aires

Primer cuatrimestre de 2021

Esta presentación introduce el vector aleatorio normal. Cuando termine la presentación deben poder responder las siguientes preguntas:

- ¿Qué es un vector aleatorio normal?
- ¿Cómo es la densidad de un vector aleatorio normal no degenerado?
- ¿Cómo son las marginales de un vector aleatorio normal no degenerado?
- ¿Qué significa que un vector aleatorio normal tenga componentes descorrelacionadas?
- ¿Cómo generar muestras de un vector aleatorio normal?

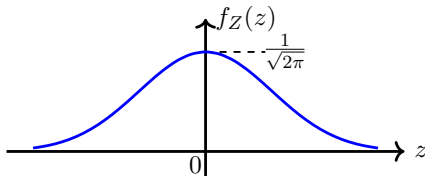
- 1 Definición general, clasificación y expresión de la densidad
- 2 Transformaciones lineales y marginales
- 3 Gráfica y curvas de nivel de un vector normal no degenerado
- 4 Propiedad importante: independencia y descorrelación
- 5 Simulación de vector aleatorio normal
- 6 Ejemplo de aplicación
- 7 Apéndices

## Repaso: variable aleatoria normal

- Recordemos que una variable **aleatoria normal estándar**  $Z$  es una variable continua con densidad:

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad z \in \mathbb{R}.$$

- La densidad tiene forma de campana, con un máximo en  $z = 0$ :



- La media de la normal estándar es 0 y su varianza es 1.

# Repaso: variable aleatoria normal

- Introduciendo el cambio de variable:

$$X = \sigma Z + \mu, \quad (\sigma > 0) \quad (1)$$

obtenemos una **variable aleatoria normal** (no estándar). Haciendo un cambio de variable con la técnica del Jacobiano es sencillo probar que:

$$f_X(x) = \frac{1}{\sigma} f_Z\left(\frac{x - \mu}{\sigma}\right),$$

es decir:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

- Tomando media y varianza en (1) vemos que la media de  $X$  es  $\mu$  y varianza  $\sigma^2$ .
- Denotamos a la variable normal como  $X \sim \mathcal{N}(\mu, \sigma^2)$ .

¿Cómo extender esta definición a múltiples dimensiones?

# Vector aleatorio normal estándar: definición

## Definición (Vector aleatorio normal estándar)

Llamaremos **vector aleatorio normal estándar** a un vector aleatorio  $\mathbf{Z} = [Z_1, \dots, Z_n]^T$  con componentes I.I.D. normales estándar, es decir:

$$f_{\mathbf{Z}}(\mathbf{z}) = \prod_{i=1}^n f_{Z_i}(z_i),$$

con  $Z_i \sim \mathcal{N}(0, 1)$ ,  $i = 1, \dots, n$ .

# Densidad del vector normal estándar

Si  $\mathbf{Z} = [Z_1, \dots, Z_n]^T$  es normal estándar podemos escribir su densidad como:

$$\begin{aligned} f_{\mathbf{Z}}(\mathbf{z}) &= \prod_{i=1}^n f_{Z_i}(z_i) \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{z_i^2}{2}} \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2} \sum_{i=1}^n z_i^2} \end{aligned}$$

$$f_{\mathbf{Z}}(\mathbf{z}) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{\|\mathbf{z}\|^2}{2}}.$$



# Vector aleatorio normal: definición (I)

## Definición (vector aleatorio normal)

- Diremos que un vector aleatorio  $\mathbf{X} = [X_1, \dots, X_m]^T$  es un vector aleatorio normal de media  $\mu_{\mathbf{X}}$  si existe una matriz  $\mathbf{A}$  en  $\mathbb{R}^{m \times n}$  tal que:

$$\mathbf{X} = \mathbf{AZ} + \mu_{\mathbf{X}},$$

donde  $\mathbf{Z} = [Z_1, \dots, Z_n]^T$  es un vector normal estándar y la igualdad quiere decir que tiene la misma distribución.

Notar que en ese caso  $\mathbf{C}_{\mathbf{X}} = \mathbb{E}[\mathbf{AA}^T]$  (ver Presentacion 1b).

Denotaremos a  $\mathbf{X}$  como:  $\mathbf{X} \sim \mathcal{N}(\mu_{\mathbf{X}}, \mathbf{C}_{\mathbf{X}})$ , donde  $\mathbf{C}_{\mathbf{X}}$  es su matriz de covarianza, que, según vimos en la Presentacion 1b, vale:  $\mathbf{C}_{\mathbf{X}} = \mathbf{AA}^T$ .

# Vector aleatorio normal: clasificación

- Partimos de un vector:  $\mathbf{X} \sim \mathcal{N}(\mu_{\mathbf{X}}, \mathbf{C}_{\mathbf{X}})$ , del cual conocemos  $\mathbf{A}$ ,  $\mathbf{b}$  tal que:

$$\mathbf{X} = \mathbf{AZ} + \mu_{\mathbf{X}},$$

con  $\mathbf{Z}$  normal estándar.

- Para hallar la densidad de  $\mathbf{X}$  debemos considerar dos casos:
  - ▶ Si la matriz  $\mathbf{C}_{\mathbf{X}}$  es inversible, entonces diremos que  $\mathbf{X}$  es un **vector aleatorio normal no degenerado**.
  - ▶ Si la matriz  $\mathbf{C}_{\mathbf{X}}$  no es inversible, entonces diremos que  $\mathbf{X}$  es un **vector aleatorio normal degenerado**.
- El caso que interesa en la práctica es el **no degenerado**.
- A continuación analizamos ambos casos.

## Observación

Notar que como  $\mathbf{C}_{\mathbf{X}} = \mathbf{AA}^T$  entonces podemos determinar si  $\mathbf{C}_{\mathbf{X}}$  es inversible o no, en función de cómo sea  $\mathbf{A}$ .

# Vector aleatorio normal no degenerado

## Lema (Densidad del VAN no degenerado)

*Si un vector aleatorio normal  $n$ -dimensional  $\mathbf{X}$  de media  $\mu_{\mathbf{X}}$  tiene una matriz de covarianza  $\mathbf{C}_{\mathbf{X}}$  inversible, su densidad es :*

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{C}_{\mathbf{X}})}} e^{-\frac{1}{2}(\mathbf{x}-\mu_{\mathbf{X}})^T \mathbf{C}_{\mathbf{X}}^{-1}(\mathbf{x}-\mu_{\mathbf{X}})}, \quad \mathbf{x} \in \mathbb{R}^n.$$

## Vector aleatorio normal no degenerado (II)

Para probar esto, sólo tenemos el dato de que:

$$\mathbf{X} = \mathbf{A}\mathbf{Z} + \mathbf{b},$$

para cierta  $\mathbf{A}$  y cierto  $\mathbf{b}$ , donde:

$$\mathbf{C}_X = \mathbf{A}\mathbf{A}^T.$$

con  $\mathbf{C}_X$  inversible.

Debemos determinar que implica que  $\mathbf{C}_X$  sobre  $\mathbf{A}$ :

- $\mathbf{A}$  es cuadrada: tenemos que<sup>1</sup>:

$$\det(\mathbf{C}_X) = \det(\mathbf{A}\mathbf{A}^T) = \det(\mathbf{A}) \det(\mathbf{A}^T) = \det(\mathbf{A})^2.$$

Por lo tanto, si  $\mathbf{A}$  es cuadrada,  $\mathbf{C}_X$  es inversible si y sólo si  $\mathbf{A}$  lo es.

- $\mathbf{A}$  es rectangular: con algo más de trabajo podemos probar que  $\mathbf{C}_X$  es inversible, si y sólo si  $\mathbf{A}$  tiene filas linealmente independientes. Este caso contiene al anterior.

---

<sup>1</sup>Recordar que sólo las matrices cuadradas tienen determinante y que son inversibles si es no nulo.

## Vector aleatorio normal no degenerado: caso $\mathbf{A}$ inversible (I)

A continuación probamos el resultado para el caso  $\mathbf{A}$  inversible:

- Tenemos  $\mathbf{X} = [X_1, \dots, X_n]^T$  un vector que se obtiene como:

$$\mathbf{X} = \mathbf{AZ} + \mathbf{b},$$

donde  $\mathbf{A}$  es inversible, y  $\mathbf{Z}$  es un vector normal.

- En este caso para hallar la densidad estamos en el caso de ejemplo de la Clase 1a: una transformación lineal inversible, de un vector aleatorio en todo  $\mathbb{R}^n$ .
- Por lo tanto la densidad de  $\mathbf{X}$  es:

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{f_{\mathbf{Z}}(\mathbf{A}^{-1}(\mathbf{x} - \mathbf{b}))}{|\det \mathbf{A}|} \quad (\mathbf{x} \in \mathbb{R}^n).$$

- Reemplazando en la densidad de  $\mathbf{Z}$

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} |\det(\mathbf{A})|} e^{-\frac{1}{2} \|\mathbf{A}^{-1}(\mathbf{x} - \mathbf{b})\|^2}$$

Debemos seguir simplificando...

## VAN no degenerado: caso $\mathbf{A}$ inversible (II)

- Por un lado vemos que:

$$|\det(\mathbf{A})| = \sqrt{\det(\mathbf{A})^2} = \sqrt{\det(\mathbf{A}) \det(\mathbf{A}^T)} = \sqrt{\det(\mathbf{C}_\mathbf{X})}.$$

- Además:

$$\begin{aligned} \|\mathbf{A}^{-1}(\mathbf{x} - \mathbf{b})\|^2 &= (\mathbf{A}^{-1}(\mathbf{x} - \mathbf{b}))^T \mathbf{A}^{-1}(\mathbf{x} - \mathbf{b}) \longrightarrow \|\mathbf{u}\|^2 = \mathbf{u}^T \mathbf{u} \\ &= (\mathbf{x} - \mathbf{b})^T (\mathbf{A}^{-1})^T \mathbf{A}^{-1}(\mathbf{x} - \mathbf{b}) \longrightarrow \mathbf{A}^T \mathbf{B} = \mathbf{B}^T \mathbf{A} \\ &= (\mathbf{x} - \mathbf{b})^T \mathbf{C}_\mathbf{X}^{-1}(\mathbf{x} - \mathbf{b}) \longrightarrow (\mathbf{A} \mathbf{A}^T)^{-1} = (\mathbf{A}^T)^{-1} \mathbf{A}^{-1} \end{aligned}$$

- Finalmente obtenemos:

$$f_\mathbf{X}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{C}_\mathbf{X})}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{b})^T \mathbf{C}_\mathbf{X}^{-1}(\mathbf{x}-\mathbf{b})}.$$

- Caso  $\mathbf{A}$  con filas LI: ver en el Apéndice.

# Vector aleatorio normal degenerado (I)

- Un vector aleatorio gaussiano degenerado es aquel que tiene una matriz de covarianza no inversible.
- De acuerdo a lo que vimos para el caso no degenerado, tenemos:

$\mathbf{C}_X$  no es inversible  $\iff \mathbf{A}$  tiene filas linealmente dependientes.

- Tenemos entonces el siguiente resultado:

## Lema (Dependencia lineal (ver ejercicio 1.8))

*Un VAN  $\mathbf{X} = \mathbf{AZ} + \mathbf{b}$  es degenerado si y sólo si con probabilidad 1 alguna(s) de las componentes del vector centrado*

$$\tilde{\mathbf{X}} = \mathbf{X} - \mathbf{b} = \mathbf{AZ} \sim \mathcal{N}(\mathbf{0}, \mathbf{AA}^T)$$

*son combinación lineal de las demás.*

- Demostración: ver en el Apéndice.

# Vector aleatorio normal degenerado (II)

## Conclusión

- Los vectores aleatorios gaussianos degenerados tienen información redundante, es decir, algunas de sus componentes son combinación lineal de las demás, lo que implica que su matriz de covarianza es no inversible.
- Eliminando las componentes linealmente dependientes obtenemos un vector no degenerado, que tiene la densidad hallada anteriormente.
- En la práctica no es habitual esta situación, por lo que siempre consideraremos que trabajamos con vectores no degenerados.



# Transformación lineal de un vector normal

## Lema (Transformación lineal)

*Cualquier transformación lineal de un vector aleatorio normal, tiene también distribución normal.*

Este Lema se desprende naturalmente de la definición de VAN. Es decir, si  $\mathbf{X} = [X_1, \dots, X_n]$  es normal, entonces, podemos expresarlo como:

$$\mathbf{X} = \mathbf{AZ} + \mu_{\mathbf{X}},$$

para cierta matriz  $\mathbf{A}$ , y  $\mathbf{Z}$  un vector normal estándar. Si definimos otro vector como  $\mathbf{Y} = [Y_1, \dots, Y_m]$  como una transformación lineal de  $\mathbf{X}$ :

$$\mathbf{Y} = \mathbf{BX} + \mathbf{b},$$

con  $\mathbf{B}$  de  $m \times n$  y  $\mathbf{b} \in \mathbb{R}^m$ , entonces:

$$\mathbf{Y} = \mathbf{BAZ} + \mathbf{B}\mu_{\mathbf{X}} + \mathbf{b}$$

será un vector aleatorio normal  $m$ -dimensional (porque cumple la definición). Si la matriz  $\mathbf{AB}$  tiene filas linealmente dependientes, será un vector degenerado.

# Marginales de un vector normal

- El Lema anterior implica que todos los subvectores de un vector aleatorio normal son a su vez, vectores normales. Si el vector original es no degenerado, también lo serán sus marginales.
- En particular, cada componente de un vector aleatorio normal es una variable aleatoria normal unidimensional de las que ya conocíamos.

# Ejemplo

## Ejemplo

Supongamos que  $\mathbf{X} = [X_1, \dots, X_n] \sim \mathcal{N}(\mu_{\mathbf{X}}, \mathbf{C}_{\mathbf{X}})$ . Entonces su componente  $i$ -ésima puede escribirse mediante una transformación lineal:

$$X_i = \mathbf{u}^T \mathbf{X},$$

donde:  $\mathbf{u} = [\underbrace{0, \dots, 0}_{i-1}, 1, \underbrace{0, \dots, 0}_{n-i}]^T$ . Entonces de acuerdo al teorema anterior,

$X_i \sim \mathcal{N}(\mu_i, \sigma_{X_i}^2)$ . La media y la varianza se obtienen con el resultado del ejemplo al final de la Clase 1b:

- $\mu_{X_i} = \mathbb{E}[\mathbf{u}^T \mathbf{X}] = \mathbf{u}^T \mathbb{E}[\mathbf{X}] = [\mu_{\mathbf{X}}]_i.$
- $\sigma_{X_i}^2 = \mathbf{u}^T \mathbf{C}_{\mathbf{X}} \mathbf{u} = [\mathbf{C}_{\mathbf{X}}]_{i,i}.$

# Ejemplo 1: cálculo de marginales y transformaciones (I)

## Enunciado

Sea  $\mathbf{X} = [X_1, X_2, X_3]^T$  un vector aleatorio normal con:

$$\mu_{\mathbf{X}} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad \mathbf{C}_{\mathbf{X}} = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 1 & -\frac{1}{2} \\ 0 & -\frac{1}{2} & 1 \end{bmatrix}$$

Halle:

- 1 La distribución de cada componente.
- 2 La distribución del vector  $[X_1, X_3]^T$ .
- 3 La distribución del vector  $\mathbf{Y} = \mathbf{B}\mathbf{X}$  con

$$\mathbf{B} = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 2 & 1 \end{bmatrix}.$$

En este caso escriba además la densidad.

## Ejemplo 1: cálculo de marginales y transformaciones (II)

- ❶ Como vimos en el ejemplo, cada marginal es normal:

$$X_i \sim \mathcal{N}([\mu_{\mathbf{X}}]_i, [\mathbf{C}_{\mathbf{X}}]_{i,i}),$$

es decir:

$$X_1 \sim \mathcal{N}(1, 2) \quad X_2 \sim \mathcal{N}(2, 1) \quad X_3 \sim \mathcal{N}(3, 1).$$

- ❷ El vector  $\tilde{\mathbf{X}} = [X_1, X_3]^T$  también será normal, pues es un subvector, que puede expresarse mediante una transformación lineal:

$$\tilde{\mathbf{X}} = \mathbf{A}\mathbf{X},$$

con

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Entonces  $\tilde{\mathbf{X}} \sim \mathcal{N}(\mathbf{A}\mu_{\mathbf{X}}, \mathbf{A}\mathbf{C}_{\mathbf{X}}\mathbf{A}^T)$ .

## Ejemplo 1: cálculo de marginales y transformaciones (III)

Haciendo las operaciones vemos que la matriz elimina las componentes de  $\mu_{\mathbf{X}}$  y  $\mathbf{C}_{\mathbf{X}}$  que no corresponden a  $\tilde{\mathbf{X}}$ , es decir:

$$\mu_{\mathbf{X}} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \longrightarrow \mu_{\tilde{\mathbf{X}}} = \mathbf{A}\mu_{\mathbf{X}} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$
$$\mathbf{C}_{\mathbf{X}} = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 1 & -\frac{1}{2} \\ 0 & -\frac{1}{2} & 1 \end{bmatrix} \longrightarrow \mathbf{C}_{\tilde{\mathbf{X}}} = \mathbf{A}\mathbf{C}_{\mathbf{X}}\mathbf{A}^T = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

- Por último, este punto es igual al anterior, pero con otra matriz de transformación, por lo que:  $\mathbf{Y} \sim \mathcal{N}(\mathbf{B}\mu_{\mathbf{X}}, \mathbf{B}\mathbf{C}_{\mathbf{X}}\mathbf{B}^T)$ . No podemos interpretar las operaciones con  $\mathbf{B}$  como una eliminación las componentes de  $\mu_{\mathbf{X}}$  y  $\mathbf{C}_{\mathbf{X}}$  que no se necesitan, debemos hacer la cuenta.

$$\mu_{\mathbf{Y}} = \mathbf{B}\mu_{\mathbf{X}} = \begin{bmatrix} 5 \\ 7 \end{bmatrix} \quad \mathbf{C}_{\mathbf{Y}} = \mathbf{B}\mathbf{C}_{\mathbf{X}}\mathbf{B}^T = \begin{bmatrix} 10 & 5 \\ 5 & 3 \end{bmatrix}$$

## Ejemplo 1: cálculo de marginales y transformaciones (IV)

Por último para escribir la densidad, busquemos la inversa de  $\mathbf{C}_Y$ :

$$\mathbf{C}_Y^{-1} = \begin{bmatrix} \frac{3}{5} & -1 \\ -1 & 2 \end{bmatrix},$$

y el determinante:

$$\det(\mathbf{C}_Y) = 5.$$

La densidad es entonces:

$$\begin{aligned} f_Y(\mathbf{y}) &= \frac{1}{2\pi\sqrt{5}} e^{-\frac{1}{2}(\mathbf{y}-\mu_Y)^T \mathbf{C}_Y^{-1}(\mathbf{y}-\mu_Y)} \\ &= \frac{1}{2\sqrt{5}\pi} e^{-\frac{1}{2}\left(\frac{3}{5}y_1^2 - 2y_1y_2 + 2y_2^2\right)} \end{aligned}$$

# Marginales normales no implican un vector normal

- Es importante mencionar que el hecho de que  $X_1, \dots, X_n$  sean variables normales (cada una por separado), es decir:

$$X_i \sim \mathcal{N}(\mu_i, \sigma_i^2), \quad 1 \leq i \leq n,$$

no implica que  $[X_1, \dots, X_n]^T$  sea un vector aleatorio normal, es decir que juntas formen un vector aleatorio normal.

- Recordemos que **cada conjunto de marginales puede corresponder a infinitas conjuntas** (ver el ejemplo de la Clase 1a).
- Entonces decir que  $X_1, \dots, X_n$  son normales cada una considerada sola, no implica que juntas formen un vector normal.



# Aclaraciones importantes de nomenclatura

- Al **vector aleatorio normal** se lo llama también **vector aleatorio gaussiano**.
- A veces también se dice que ciertas variables aleatorias  $X_1, \dots, X_n$  son **conjuntamente gaussianas**. Esto es equivalente a decir que  $\mathbf{X} = [X_1, \dots, X_n]$  es un vector aleatorio normal, es decir,  $\mathbf{X} \sim \mathcal{N}(\mu_{\mathbf{X}}, \mathbf{C}_{\mathbf{X}})$  para cierto vector  $\mu_{\mathbf{X}}$  y cierta matriz  $\mathbf{C}_{\mathbf{X}}$ .

## Mucha atención

Debemos tener cuidado con la afirmación: “ $X_1, \dots, X_n$  son variables aleatorias gaussianas” porque es una frase ambigua. ¿Qué significa?

- ▶ ¿Que son conjuntamente gaussianas? (o sea, juntas son un vector aleatorio normal)
- ▶ ¿Que cada una por separada es gaussiana pero juntas no? (en ese caso no podríamos saber cuál es su conjunta sin otras hipótesis).

Evitaremos esta frase en lo posible. Si la usamos, nos estaremos refiriendo siempre a que son conjuntamente gaussianas. Los estudiantes deben evitarla, porque el docente no puede saber si están entendiendo o no los conceptos.

# Forma de la distribución no degenerada

Al igual que en el caso unidimensional, el vector aleatorio normal tendrá una forma de campana, pero en un espacio de mayor dimensión. Para ello observamos que:

- El máximo de la densidad está también en el vector de media: la función del exponente:

$$(\mathbf{x} - \mathbf{b})^T \mathbf{C}_X^{-1} (\mathbf{x} - \mathbf{b})$$

tiene un mínimo absoluto, nulo, cuando  $\mathbf{x} = \mathbf{b}$ , por lo que

$$e^{-\frac{1}{2}(\mathbf{x}-\mathbf{b})^T \mathbf{C}_X^{-1} (\mathbf{x}-\mathbf{b})}$$

tendra un único máximo absoluto en ese mismo punto.

- Las curvas de nivel del vector son hiperelipsoides, es decir, también será monótonamente decreciente con la distancia a la media, al igual que en el caso unidimensional.

A continuación estudiamos esto.

# Curvas de nivel del VAN no degenerado (I)

Podemos definir las curvas de nivel  $\alpha$  de la densidad como:

$$C(\alpha) = \{\mathbf{x} \in \mathbb{R}^n : f_{\mathbf{X}}(\mathbf{x}) = \alpha\}.$$

A continuación veremos que estas curvas de nivel son elipses cuyos ejes no son los ejes canónicos en general:

- Dado un valor de  $\alpha$ , despejamos:

$$\begin{aligned}f_{\mathbf{X}}(\mathbf{x}) &= \alpha \\e^{-\frac{1}{2}(\mathbf{x}-\mu_{\mathbf{X}})^T \mathbf{C}_{\mathbf{X}}^{-1}(\mathbf{x}-\mu_{\mathbf{X}})} &= \sqrt{(2\pi)^n \det(\mathbf{C}_{\mathbf{X}})} \alpha \\-\frac{1}{2}(\mathbf{x}-\mu_{\mathbf{X}})^T \mathbf{C}_{\mathbf{X}}^{-1}(\mathbf{x}-\mu_{\mathbf{X}}) &= \log \left( \sqrt{(2\pi)^n \det(\mathbf{C}_{\mathbf{X}})} \alpha \right) \\(\mathbf{x}-\mu_{\mathbf{X}})^T \mathbf{C}_{\mathbf{X}}^{-1}(\mathbf{x}-\mu_{\mathbf{X}}) &= -2 \log \left( \sqrt{(2\pi)^n \det(\mathbf{C}_{\mathbf{X}})} \alpha \right) \\(\mathbf{x}-\mu_{\mathbf{X}})^T \mathbf{C}_{\mathbf{X}}^{-1}(\mathbf{x}-\mu_{\mathbf{X}}) &= \tilde{\alpha}.\end{aligned}$$

$$\text{donde } \tilde{\alpha} = -2 \log \left( \sqrt{(2\pi)^n \det(\mathbf{C}_{\mathbf{X}})} \alpha \right).$$

## Curvas de nivel del VAN no degenerado (II)

- Esta es la expresión implícita de una hiperelipsoide, centrada en el vector media  $\mu_{\mathbf{X}}$ .
- Para ver esto primero centramos la ecuación, con un cambio de la forma:

$$\mathbf{y} = \mathbf{x} - \mu_{\mathbf{X}}$$

obteniendo:

$$\mathbf{y}^T \mathbf{C}_{\mathbf{X}}^{-1} \mathbf{y} = \tilde{\alpha}.$$

- Luego diagonalizamos la matriz de covarianza, según vimos en la presentación anterior:

$$\mathbf{C}_{\mathbf{X}} = \mathbf{P} \mathbf{D} \mathbf{P}^T \implies \mathbf{C}_{\mathbf{X}}^{-1} = \mathbf{P} \mathbf{D}^{-1} \mathbf{P}^T,$$

y obtenemos:

$$\mathbf{y}^T \mathbf{P} \mathbf{D}^{-1} \mathbf{P}^T \mathbf{y} = \tilde{\alpha}.$$

- Por último introducimos el cambio:

$$\mathbf{P} \mathbf{u} = \mathbf{y},$$

y expresamos la curva de nivel como:

$$\mathbf{u}^T \mathbf{D}^{-1} \mathbf{u} = \tilde{\alpha}.$$

## Curvas de nivel del VAN no degenerado (III)

- La transformación  $\mathbf{P}$  representa un cambio de coordenadas entre la base ortonormal formada por las columnas de  $\mathbf{P}$ ,  $\{\mathbf{p}_1, \dots, \mathbf{p}_n\}$  y la base canónica de  $\mathbb{R}^n$ .
- Para transformar un vector  $\mathbf{u}$  de la nueva base a la base canónica original, hacemos:

$$\mathbf{y} = \mathbf{P}\mathbf{u}.$$

- Los ejes principales del nuevo sistema de coordenadas:

$$\{\mathbf{e}_1, \dots, \mathbf{e}_n\} \text{ donde } \mathbf{e}_i = \underbrace{[0, \dots, 0]_{i-1}}_{i-1}, \underbrace{[1, 0, \dots, 0]_{n-i}}_{n-i}^T.$$

corresponden a los autovectores de  $\mathbf{C}_X$  en el viejo sistema de coordenadas:

$$\mathbf{P}\mathbf{e}_i = \mathbf{p}_i.$$

- Además como la matriz  $\mathbf{P}$  es ortonormal, representa rotaciones o reflexiones por lo que curva de nivel en las nuevas coordenadas solamente tendrá sus ejes rotados o reflejados respecto de los vectores canónicos de  $\mathbb{R}^n$ .

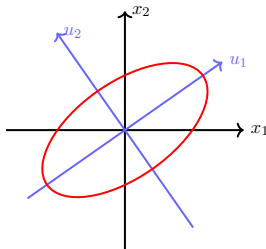
## Curvas de nivel del VAN no degenerado (IV)

- Al diagonalizar tenemos:

$$\mathbf{u}^T \mathbf{D}^{-1} \mathbf{u} = \sum_{i=1}^n \left( \frac{u_i}{\sqrt{\lambda_i}} \right)^2 = \tilde{\alpha},$$

donde  $\{\lambda_1, \dots, \lambda_n\}$  son los autovalores de  $\mathbf{C}_{\mathbf{X}}$ . Esta es la ecuación implícita de una hiperelipsoide, cuyos radios son las raíces de los autovalores de  $\mathbf{C}_{\mathbf{X}}$ .

- El siguiente gráfico muestra el caso particular para un VAN bidimensional:



$x_1$  y  $x_2$  son los ejes canónicos de  $\mathbb{R}^2$   
 $u_1$  y  $u_2$  son las direcciones de los  
autovectores de  $\mathbf{C}_{\mathbf{X}}$ ,  
que son los ejes de la elipse

## Ejemplo 2: gráfico de la densidad y las curvas de nivel

### Enunciado

Grafique las curvas de nivel para el vector aleatorio  $\mathbf{X} \sim \mathcal{N}(\mu_{\mathbf{X}}, \mathbf{C}_{\mathbf{X}})$ , con:

$$\mu_{\mathbf{X}} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad \mathbf{C}_{\mathbf{X}} = \frac{1}{5} \begin{bmatrix} 9 & 2 \\ 2 & 6 \end{bmatrix}.$$

Diagonalizamos  $\mathbf{C}_{\mathbf{X}} = \mathbf{P}\mathbf{D}\mathbf{P}^T$ , y obtenemos:

$$\mathbf{D} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \quad \mathbf{P} = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix} = [\mathbf{p}_1, \mathbf{p}_2].$$

## Ejemplo 2: gráfico de la densidad y las curvas de nivel (II)

Luego la elipse correspondiente a la curva de nivel  $\alpha$  en el sistema de coordenadas  $(u_1, u_2)$  dado por los autovectores de  $\mathbf{P}$  tiene la ecuación:

$$\mathbf{u}^T \mathbf{D}^{-1} \mathbf{u} = \tilde{\alpha},$$

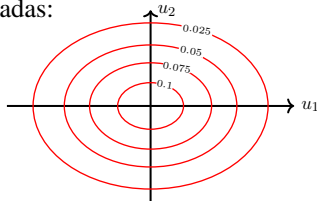
es decir:

$$\left( \frac{u_1}{\sqrt{2}} \right)^2 + u_2^2 = \tilde{\alpha},$$

donde  $\tilde{\alpha} = -2 \log \left( 2\pi \sqrt{\det(\mathbf{C}_X)} \alpha \right)$ . Para volver a las coordenadas originales debemos aplicar la transformación:

$$\mathbf{y} = \mathbf{P}\mathbf{u} \longrightarrow \mathbf{x} = \mathbf{y} + \mu_X$$

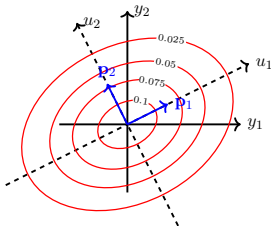
En coordenadas de  $(u_1, u_2)$ , los ejes principales de la elipse están alineadas con los ejes del sistema de coordenadas:



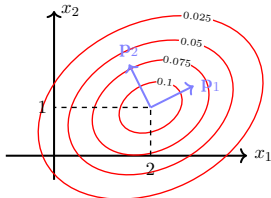


## Ejemplo 2: gráfico de la densidad y las curvas de nivel (III)

La transformación  $\mathbf{Pu} = \mathbf{y}$  rota las elipses para que estén alineadas con las coordenadas del sistema original:



Por último, la transformación  $\mathbf{x} = \mathbf{y} + \boldsymbol{\mu}_X$  nos devuelve al sistema de coordenadas original, donde las elipses están centradas en la media del vector aleatorio:



# Independencia y descorrelación

- Sabemos que el siguiente resultado vale siempre:

Si  $X_1, X_2$  son variables independientes  $\implies X_1$  y  $X_2$  son descorrelacionadas.

- Esto implica lo siguiente:

Si  $\mathbf{X} = [X_1, \dots, X_n]^T$  tiene componentes independientes entonces su matriz de covarianza es diagonal.

- Para ver esto, basta con observar que:

$$\begin{aligned} X_i, X_j, i \neq j \text{ son independientes} &\implies C(X_i, X_j) = 0, \quad i \neq j \\ &\implies [\mathbf{C}_X]_{i,j} = 0, \quad i \neq j \\ &\implies \mathbf{C}_X \text{ es diagonal.} \end{aligned}$$

- En general, descorrelación no implica independiencia, es decir, es una condición más débil:

Si  $X_1, X_2$  son variables descorrelacionadas  $\nRightarrow$  son independientes.

Las variables normales son uno de los únicos casos donde estas dos condiciones son equivalentes.

# Independencia y descorrelación para el VAN (I)

En el caso del vector aleatorio normal tenemos el siguiente resultado:

## Lema (Independencia si y sólo si descorrelación)

*Sea  $\mathbf{X} = [X_1, \dots, X_n]^T$  un vector aleatorio normal no degenerado. Entonces las componentes de  $\mathbf{X}$  son independientes si y sólo si son descorrelacionadas.*

Para hacer la demostración debemos probar las dos implicancias, aunque una es válida siempre:

- $\Rightarrow$ : esta implicación es válida siempre, porque independencia siempre implica descorrelación.
- $\Leftarrow$ : esta debemos probarla. Para ello, utilizamos que si  $\mathbf{X}$  tiene componentes independientes, entonces  $\mathbf{C}_\mathbf{X}$  es diagonal. Partimos de la densidad de  $\mathbf{X}$ :

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{C}_\mathbf{X})}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{b})^T \mathbf{C}_\mathbf{X}^{-1}(\mathbf{x}-\mathbf{b})},$$

y debemos mostrar que se factoriza como el producto de las densidades de sus componentes.

# Independencia y descorrelación para el VAN (II)

- Debemos trabajar el exponente. Como  $\mathbf{C}_X$  es diagonal, tenemos:

$$[\mathbf{C}_X]_{i,j} = \begin{cases} 0 & \text{si } i \neq j \\ \sigma_{X_i}^2 & \text{si } i = j. \end{cases}$$

- Usando esto podemos simplificar el exponente como:

$$(\mathbf{x} - \mathbf{b})^T \mathbf{C}_X^{-1} (\mathbf{x} - \mathbf{b}) = \sum_{i=1}^n \frac{(x_i - \mu_{X_i})^2}{\sigma_{X_i}^2},$$

y el determinante:

$$\det(\mathbf{C}_X) = \prod_{i=1}^n \sigma_{X_i}^2.$$

# Independencia y descorrelación para el VAN (III)

- Por lo tanto reemplazando en la densidad:

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= \frac{1}{\sqrt{(2\pi)^n \prod_{i=1}^n \sigma_{X_i}^2}} e^{-\frac{1}{2} \sum_{i=1}^n \frac{(x_i - \mu_{X_i})^2}{\sigma_{X_i}^2}} \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi \sigma_{X_i}^2}} e^{-\frac{(x_i - \mu_{X_i})^2}{2\sigma_{X_i}^2}} \\ &= \prod_{i=1}^n f_{X_i}(x_i). \end{aligned}$$

observamos que logramos factorizarla como el producto de  $n$  densidades normales unidimensionales.

# Simulación de vector aleatorio normal (I)

- Se desean generar muestras de un vector aleatorio normal estándar en  $\mathbb{R}^n$ :

$$\mathbf{X} \sim \mathcal{N}(\mu_{\mathbf{X}}, \mathbf{C}_{\mathbf{X}}).$$

- Es sencillo generar vectores aleatorios normales en  $\mathbb{R}^n$ ,  $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ , pues solamente hay que generar  $n$  variables  $\mathcal{N}(0, 1)$  independientes, y colocarlas en un vector.
- Para generar muestras de  $\mathbf{X}$  debemos hallar una transformación lineal tal que:

$$\mathbf{X} = \mathbf{AZ} + \mu_{\mathbf{X}},$$

pues entonces, generamos muestras independientes de  $\mathbf{Z}$  y al transformarlas, obtendremos muestras independientes de  $\mathbf{X}$ .

## Simulación de vector aleatorio normal (II)

- Para hallar la transformación debemos encontrar una  $\mathbf{A}$  tal que:

$$\mathbf{C}_X = \mathbf{A}\mathbf{A}^T.$$

- Esta matriz no es única, de hecho hay infinitas. Veamos algunos ejemplos:
- Diagonalizamos la matriz  $\mathbf{C}_X$  como

$$\mathbf{C}_X = \mathbf{P}\mathbf{D}\mathbf{P}^T$$

y vemos que como  $\mathbf{C}_X$  tiene autovalores no negativos, podemos escribir:

$$\mathbf{D} = \mathbf{D}^{\frac{1}{2}}\mathbf{D}^{\frac{1}{2}},$$

donde

$$[\mathbf{D}^{\frac{1}{2}}]_{i,j} = \begin{cases} \sqrt{\lambda_i} & \text{si } i = j \\ 0 & \text{en otro caso.} \end{cases},$$

y  $\{\lambda_1, \dots, \lambda_n\}$  son los autovalores de  $\mathbf{C}_X$ .

## Simulación de vector aleatorio normal (III)

- Con la factorización:

$$\mathbf{C}_X = \mathbf{P} \mathbf{D}^{\frac{1}{2}} \mathbf{D}^{\frac{1}{2}} \mathbf{P}^T$$

podemos proponer varios ejemplos de  $\mathbf{A}$ :

- La primera consiste en escribir la factorización como:

$$\mathbf{C}_X = \left( \mathbf{P} \mathbf{D}^{\frac{1}{2}} \right) \left( \mathbf{P} \mathbf{D}^{\frac{1}{2}} \right)^T,$$

donde vemos que tomando

$$\mathbf{A}_1 = \mathbf{P} \mathbf{D}^{\frac{1}{2}},$$

se cumple lo pedido.

- Otra opción es incorporar en el medio el término  $\mathbf{P}^T \mathbf{P} = \mathbf{I}$ :

$$\mathbf{C}_X = \mathbf{P} \mathbf{D}^{\frac{1}{2}} \mathbf{P}^T \mathbf{P} \mathbf{D}^{\frac{1}{2}} \mathbf{P}^T.$$

Se define la raíz cuadrada de la matriz  $\mathbf{C}_X$  como:

$$\mathbf{C}_X^{\frac{1}{2}} = \mathbf{P} \mathbf{D}^{\frac{1}{2}} \mathbf{P}^T.$$

Se la llama la raíz cuadrada porque es la única matriz simétrica SDP tal que:

$$\mathbf{C}_X = \mathbf{C}_X^{\frac{1}{2}} \mathbf{C}_X^{\frac{1}{2}}.$$

También cumple lo pedido.



## Simulación de vector aleatorio normal (IV)

- 4 Por último, se podría usar la factorización de Cholesky, que realiza la descomposición:

$$\mathbf{C}_\mathbf{X} = \mathbf{L}\mathbf{L}^T,$$

donde  $\mathbf{L}$  es una matriz triangular inferior.

Una vez obtenida la matriz  $\mathbf{A}$  el algoritmo sería el siguiente:

### Algoritmo para generar vectores normales

Se desea generar  $m$  muestras independientes del vector normal:

$$\mathbf{X} = [X_1, \dots, X_n]^T \sim \mathcal{N}(\mu_\mathbf{X}, \mathbf{C}_\mathbf{X}).$$

- 1 Generar  $m$  muestras independientes del vector  $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ .
- 2 Obtener la matriz  $\mathbf{A}$  por alguno de los modos anteriores.
- 3 Para cada realización  $\mathbf{z}_i$ ,  $1 \leq i \leq n$ , hacer:

$$\mathbf{x}_i = \mathbf{A}\mathbf{z}_i + \mu_\mathbf{X}.$$

## Ejemplo 3: generar y graficar muestras de un VAN (I)

### Enunciado

Genere realizaciones del vector aleatorio normal del Ejemplo 2:  $\mathbf{X} \sim \mathcal{N}(\mu_{\mathbf{X}}, \mathbf{C}_{\mathbf{X}})$ , con:

$$\mu_{\mathbf{X}} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad \mathbf{C}_{\mathbf{X}} = \frac{1}{5} \begin{bmatrix} 9 & 2 \\ 2 & 6 \end{bmatrix}.$$

Realice un histograma 2D con un mapa de colores y compare las curvas de nivel que obtuvo en el Ejemplo 2 con el histograma.

**Nota:** este ejemplo lo resolveremos utilizando el algoritmo que acabamos de explicar. Numpy y los demás paquetes de cálculo tienen rutinas específicas para generar dichos vectores, que utilizan este algoritmo. Ver los ejemplos resueltos.

## Ejemplo 3: generar y graficar muestras de un VAN (II)

El siguiente es un ejemplo de código que implementa lo pedido:

```
import numpy as np
from matplotlib import pyplot as plt
m = 200000 # cantidad de vectores a generar

# media del vector a generar: el reshape permite luego sumarlo
# con la matriz mas abajo
muX = np.array ([2,1]).reshape(2,1)
Cx = 1/5 * np.array ([[9, 2], [2, 6]]) # matriz de covarianza a generar

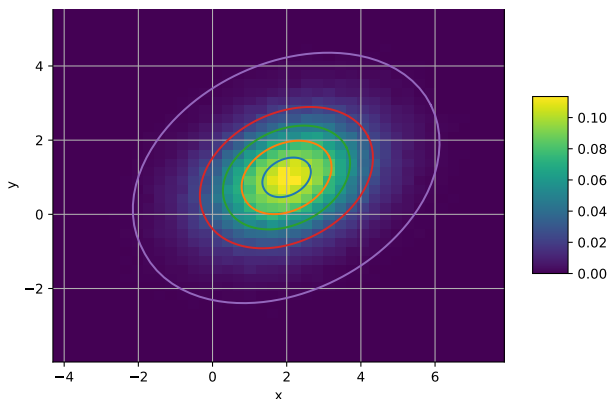
A = np.linalg.cholesky(Cx)
# matriz que tiene en cada columna un vector normal estandar (2 x m):
Z = np.random.randn(muX.size ,m)

# matriz con las muestras del vector con la distribucion deseada:
X = np.matmul(A, Z) + muX

# Grafico
(H, xedges, yedges) = np.histogram2d(X[0], X[1], density=True, bins=40)
fig = plt.figure()
ax = fig.add_subplot(111, title='', aspect='equal')
Xvals, Yvals = np.meshgrid(xedges, yedges)
surf=ax.pcolormesh(Xvals, Yvals, H)
fig.colorbar(surf, shrink=0.5, aspect=5)
```

### Ejemplo 3: generar y graficar muestras de un VAN (III)

Generado una corrida del código anterior y luego superponiendo las curvas de nivel halladas en el Ejemplo 2, obtenemos el siguiente resultado:



Las curvas de nivel de mayor a menor toman valores: 0.1, 0.075, 0.05, 0.025.

Comparando las curvas visualmente se observa que la densidad estimada mediante el histograma es compatible con las curvas de nivel calculadas teóricamente.

## Ejemplo 4: procesamiento de una señal ruidosa (I)

### Enunciado

Se observa un vector  $\mathbf{Y} \in \mathbb{R}^n$  de la forma:

$$\mathbf{Y} = \mathbf{N} + \mathbf{u}$$

donde  $\mathbf{u}$  es un vector determinístico de interés (señal) y  $\mathbf{N} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_\mathbf{N})$ , es un vector de ruido (no degenerado) que corrompe la medición de  $\mathbf{u}$ .

- 1 Halle la densidad de  $\mathbf{Y}$ .
- 2 Halle una transformación lineal de  $\mathbf{Y}$  que descorrelacione las componentes del ruido  $\mathbf{N}$ , es decir, una matriz  $\mathbf{B} \in \mathbb{R}^{n \times n}$  tal que al hacer  $\mathbf{B}\mathbf{Y}$ , el ruido procesado tenga componentes independientes.
- 3 Halle un vector  $\mathbf{v} \in \mathbb{R}^n$  tal que  $\|\mathbf{v}\| = 1$  y además al procesar la señal  $\mathbf{Y}$  como:

$$X = \mathbf{v}^T \mathbf{Y},$$

la varianza del ruido que queda en  $X$  sea mínima.

## Ejemplo 4: procesamiento de una señal ruidosa (II)

- ❶ Para hallar la densidad de  $\mathbf{Y}$ , notamos que se trata simplemente de una transformación lineal del ruido, que consiste en sumar el vector fijo  $\mathbf{u}$ . Esto solamente cambia su media:

$$\mathbf{Y} \sim \mathcal{N}(\mathbf{u}, \mathbf{C}_N).$$

- ❷ Al aplicar una transformación lineal a  $\mathbf{Y}$ , obtenemos lo siguiente:

$$\mathbf{B}\mathbf{Y} = \mathbf{B}(\mathbf{u} + \mathbf{N}) = \mathbf{B}\mathbf{u} + \mathbf{B}\mathbf{N} = \mathbf{B}\mathbf{u} + \tilde{\mathbf{N}}.$$

Se modifica la señal  $\mathbf{u}$ , pero también el ruido. Según vimos, el nuevo ruido  $\tilde{\mathbf{N}}$  tendrá distribución:

$$\tilde{\mathbf{N}} \sim \mathcal{N}(\mathbf{0}, \mathbf{B}\mathbf{C}_N\mathbf{B}^T).$$

El objetivo es que al transformar la señal  $\mathbf{Y}$ , el ruido tenga componentes descorrelacionadas. En este caso al ser ruido normal, esto significa que también serán independientes. Es decir, se busca que  $\mathbf{B}\mathbf{C}_N\mathbf{B}^T$  sea una matriz diagonal.

## Ejemplo 4: procesamiento de una señal ruidosa (III)

- 2 Cuando vimos como generar muestras de un vector aleatorio normal, hallamos varias matrices  $\mathbf{A}$  inversibles, tales que:

$$\mathbf{N} = \mathbf{A}\mathbf{Z} + \mu_{\mathbf{N}},$$

con  $\mathbf{Z}$  un vector normal estándar.

Entonces, invirtiendo esas transformaciones tenemos:

$$\mathbf{A}^{-1}(\mathbf{N} - \mu_{\mathbf{N}}) = \mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}),$$

es decir, obtenemos un ruido con componentes descorrelacionadas y varianza unitaria. Hay infinitas de estas matrices. Por ejemplo, con la raíz cuadrada de  $\mathbf{C}_{\mathbf{N}}$  tendríamos:

$$\mathbf{N} = \mathbf{C}_{\mathbf{N}}^{\frac{1}{2}}\mathbf{Z} + \mu_{\mathbf{N}} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{\mathbf{N}}),$$

La media es nula en este caso, entonces:

$$\mathbf{Z} = \left(\mathbf{C}_{\mathbf{N}}^{\frac{1}{2}}\right)^{-1} \mathbf{N} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}).$$

Podríamos tomar entonces:

$$\mathbf{B} = \left(\mathbf{C}_{\mathbf{N}}^{\frac{1}{2}}\right)^{-1}.$$

## Ejemplo 4: procesamiento de una señal ruidosa (IV)

- Queremos procesar la señal  $\mathbf{Y}$  como:

$$X = \mathbf{v}^T \mathbf{Y}.$$

Notar que esto equivale a hacer una combinación lineal de las componentes de  $\mathbf{Y}$ . La variable  $X$  será una variable escalar. Recordemos que al hacer  $\mathbf{v}^T \mathbf{Y}$  estamos haciendo el producto interno canónico entre dos vectores de  $\mathbb{R}^n$ .

Observemos que al procesar la señal:

$$X = \mathbf{v}^T \mathbf{Y} = \mathbf{v}^T \mathbf{u} + \mathbf{v}^T \mathbf{N}$$

el término  $\mathbf{v}^T \mathbf{u}$  es una constante que corresponde a procesar la señal sola. Entonces la varianza de  $X$  es la varianza del ruido procesado:

$$\mathbb{V}(X) = \mathbb{V}(\mathbf{v}^T \mathbf{N}).$$

El ejercicio pide minimizar la varianza del ruido que queda luego del filtrado, esto equivale a minimizar la varianza de  $X$ .



## Ejemplo 4: procesamiento de una señal ruidosa (V)

- Notar que si minimizamos sin más  $\mathbb{V}(X)$  estaríamos resolviendo un problema mal definido, pues tomando  $\mathbf{v} = 0$  tenemos que  $\mathbb{V}(X) = 0$ , es decir, se elimina el ruido, pero en ese caso  $X = 0$  y  $\mathbf{v}^T \mathbf{u} = 0$ , por lo que se perdió la señal también. Por eso se incorpora la restricción  $\|\mathbf{v}\| = 1$ .

El problema a resolver es entonces:

Hallar  $\mathbf{v}_{\text{opt}}$  tal que:  $\mathbb{V}(X) = \mathbb{V}(\mathbf{v}_{\text{opt}}^T \mathbf{N})$  sea mínima y  $\|\mathbf{v}_{\text{opt}}\| = 1$ .

Matemáticamente:

$$\mathbf{v}_{\text{opt}} = \arg \min_{\mathbf{v}: \|\mathbf{v}\|=1} \mathbb{V}(\mathbf{v}^T \mathbf{N}).$$

Rescribimos la varianza como:

$$\mathbb{V}(\mathbf{v}^T \mathbf{N}) = \mathbb{E}[(\mathbf{v}^T \mathbf{N})^2] = \mathbb{E}[\mathbf{v}^T \mathbf{N} \mathbf{N}^T \mathbf{v}] = \mathbf{v}^T \mathbf{C}_{\mathbf{N}} \mathbf{v}.$$

Entonces el objetivo es minimizar la forma cuadrática que define la matriz de covarianza de  $\mathbf{N}$ .

## Ejemplo 4: procesamiento de una señal ruidosa (VI)

Para minimizar o maximizar la varianza podemos recurrir a las siguientes cotas, que valen para cualquier  $\mathbf{v}$ :

$$\lambda_{\min} \leq \frac{\mathbf{v}^T \mathbf{C}_N \mathbf{v}}{\|\mathbf{v}\|^2} \leq \lambda_{\max},$$

donde  $\lambda_{\min}$  y  $\lambda_{\max}$  son los autovalores máximos y mínimos de  $\mathbf{C}_N$ . El cociente se llama cociente de Rayleigh, y en el Apéndice puede verse la demostración de la cota, que es bastante sencilla.

En este caso,  $\|\mathbf{v}\|^2 = 1$ , entonces tenemos:

$$\lambda_{\min} \leq \mathbb{V}(\mathbf{v}^T \mathbf{N}) \leq \lambda_{\max}.$$

Vemos que la varianza nunca va a ser menor que  $\lambda_{\min}$  de  $\mathbf{C}_N$ . Para lograr la igualdad, elegimos a  $\mathbf{v} = \mathbf{p}_{\min}$ , es decir, el autovector (normalizado) asociado a  $\lambda_{\min}$ . En ese caso tenemos:

$$\mathbf{p}_{\min}^T \mathbf{C}_N \mathbf{p}_{\min} = \mathbf{p}_{\min} \lambda_{\min} \mathbf{p}_{\min} = \lambda_{\min} \|\mathbf{p}_{\min}\|^2 = \lambda_{\min}.$$

Notar que si elegimos  $\mathbf{p}_{\max}$  el autovector asociado a  $\lambda_{\max}$  estaremos maximizando la varianza del ruido, es decir, aumentando el ruido.

# Vector aleatorio normal degenerado

- Para probar el resultado asumimos que la media es nula, dado que no cambia el resultado.
- Luego, supongamos que  $\{\mathbf{f}_1, \dots, \mathbf{f}_n\}$  son las filas de  $\mathbf{A}$  pensadas como columnas. Si son LD entonces, existe alguna, por ejemplo la  $i$ -ésima tal que:

$$\mathbf{f}_i = \sum_{j=1, j \neq i}^n \alpha_j \mathbf{f}_j.$$

- Además para cada  $1 \leq j \leq n$  se tiene  $X_j = \mathbf{f}_j^T \mathbf{Z}$ .
- Combinando estas ecuaciones para  $X_i$  que es la componente de la fila LD:

$$X_i = \mathbf{f}_i^T \mathbf{Z} = \sum_{j=1, j \neq i}^n \alpha_j \mathbf{f}_j^T \mathbf{Z} = \sum_{j=1, j \neq i}^n \alpha_j X_j.$$

donde se obtiene el resultado buscado, es decir, la componente  $X_i$  resulta ser una combinación lineal de las demás, y entonces

$$\mathbb{P} \left( X_i = \sum_{j=1, j \neq i}^n \alpha_j X_j \right) = 1.$$

## VAN no degenerado: $\mathbf{A}$ rectangular con filas LI

- La demostración en este caso es un poquito más complicada, porque no podemos usar directamente la técnica del Jacobiano, dado que  $\mathbf{A}$  no es cuadrada (ni inversible).
- Para que las filas de  $\mathbf{A}$  sean LI la matriz debe tener mas columnas que filas, es decir, debe ser una matriz alargada.
- Lo que se hace es extender el vector  $\mathbf{X}$  y por consecuencia la matriz  $\mathbf{A}$ , de modo que  $\mathbf{A}$  resulte cuadrada e inversible. Para ello se agregan filas a  $\mathbf{A}$ , es decir:

$$\tilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} = \begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix} \mathbf{Z} + \tilde{\mathbf{b}}.$$

- Luego estamos en el caso anterior con matriz inversible y conocemos la densidad.
- La matriz  $\mathbf{B}$  se elige de modo que  $\mathbf{X}$  e  $\mathbf{Y}$  resulten independientes, y por lo que se factoriza:

$$f_{\tilde{\mathbf{X}}}(\tilde{\mathbf{x}}) = f_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{Y}}(\mathbf{y}).$$

- Por ultimo se marginaliza el vector y se obtiene la marginal de  $\mathbf{X}$ , que resulta como la del Lema.

# Demostración de las cotas sobre el cociente de Rayleigh

- Partimos de una forma cuadrática  $\mathbf{v}^T \mathbf{C} \mathbf{v}$ , la cual vamos a acotar.
- Diagonalizamos la matriz:

$$\mathbf{C} = \mathbf{P} \mathbf{D} \mathbf{P}^T$$

y reemplazamos:

$$\mathbf{v}^T \mathbf{C} \mathbf{v} = \mathbf{v}^T \mathbf{P} \mathbf{D} \mathbf{P}^T \mathbf{v}.$$

- Luego aplicamos el cambio  $\mathbf{v} = \mathbf{P} \mathbf{u}$  y entonces podemos acotar:

$$\mathbf{u}^T \mathbf{D} \mathbf{u} = \sum_{i=1}^n \lambda_i u_i^2 \geq \lambda_{\min} \sum_{i=1}^n u_i^2 = \lambda_{\min} \|\mathbf{u}\|^2.$$

- Esto implica que:

$$\frac{\mathbf{u}^T \mathbf{D} \mathbf{u}}{\|\mathbf{u}\|^2} \geq \lambda_{\min}.$$

- Pero como la matriz  $\mathbf{P}$  es ortogonal conserva las normas, entonces:

$$\|\mathbf{v}\|^2 = \|\mathbf{P} \mathbf{u}\|^2.$$

- Entonces se tiene lo pedido:

$$\frac{\mathbf{v}^T \mathbf{C} \mathbf{v}}{\|\mathbf{v}\|^2} \geq \lambda_{\min}.$$