Distribución Gaussiana Multivariable

Cecilia Galarza

Procesos Estocásticos
Facultad de Ingeniería, Universidad de Buenos Aires

2do cuatrimestre 2024

Media y Covarianza de un VeA

Media de un vector aleatorio

Sea $\mathbf{X} \in \mathbb{R}^n$ un VeA. Luego,

$$\mathbb{E}[\mathbf{X}] = \mu_{\mathbf{X}} = \int_{-\infty}^{\infty} \boldsymbol{x}\, f_{\mathbf{X}}(\boldsymbol{x}) d\boldsymbol{x} = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{n} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} f_{X_1 \cdots X_n}(x_1, \cdots x_n) dx_1 \cdots dx_n$$

$$= \begin{bmatrix} \int_{-\infty}^{\infty} x_1 dx_1 \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1 \cdots X_n}(x_1, x_2 \cdots x_n) dx_2 \cdots dx_n \\ \vdots \\ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1 \cdots X_n}(x_1, \cdots x_{n-1}, x_n) dx_1 \cdots dx_{n-1} \int_{-\infty}^{\infty} x_n dx_n \end{bmatrix}$$

$$= \begin{bmatrix} \int_{-\infty}^{\infty} x_1 f_{X_1}(x_1) dx_1 \\ \vdots \\ \int_{-\infty}^{\infty} x_n f_{X_n}(x_n) dx_n \end{bmatrix} = \begin{bmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_n] \end{bmatrix}$$

Repaso varianza y correlación

Recordemos que si X es una V.A cuya media es μ_X, su *varianza* está definida como

$$\mathbb{V}[X] = \mathbb{E}[(X - \mu)^2].$$

Igualmente, si X e Y son dos V.A, definimos la *covarianza* entre X e Y como

$$Cov[X, Y] = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)].$$

Decimos que X e Y están *descorrelacionadas* si $Cov[X, Y] = 0$

Matriz de covarianza

Ahora, sea $\mathbf{X} \in \mathbb{R}^n$ un VeA. La **Matriz de Covarianza** de $\mathbf{X}$ está definida del siguiente modo:

$$\mathbf{Cov}(\mathbf{X}) = C_{\mathbf{X}} = \mathbb{E}[(\mathbf{X} - \mu_{\mathbf{X}})(\mathbf{X} - \mu_{\mathbf{X}})^t]$$

Matriz de covarianza

$$
\begin{aligned}
C_{\mathbf{X}} &= \mathbb{E}\left\{ \begin{bmatrix} X_1 - \mu_{X_1} \\ \vdots \\ X_n - \mu_{X_n} \end{bmatrix} \begin{bmatrix} X_1 - \mu_{X_1} & \cdots & X_n - \mu_{X_n} \end{bmatrix} \right\} \\
&= \mathbb{E}\left\{ \begin{bmatrix} (X_1 - \mu_{X_1})^2 & \cdots & (X_1 - \mu_{X_1})(X_n - \mu_{X_n}) \\ & \ddots & \\ (X_n - \mu_{X_n})(X_1 - \mu_{X_1}) & \cdots & (X_n - \mu_{X_n})^2 \end{bmatrix} \right\} \\
&= \begin{bmatrix} \underbrace{\mathbb{E}[(X_1 - \mu_{X_1})^2]}_{\mathbb{V}[X_1]} & \cdots & \underbrace{\mathbb{E}[(X_1 - \mu_{X_1})\mathbb{E}[(X_n - \mu_{X_n})]}_{Cov[X_1, X_n]} \\ & \ddots & \\ Cov[X_n, X_1] & \cdots & \mathbb{V}[X_n] \end{bmatrix}
\end{aligned}
$$

Propiedades de $C_{\mathbf{X}}$

- Es una matriz simétrica, $C_{\mathbf{X}} = C_{\mathbf{X}}^t$
- Es una matriz definida positiva, $C_{\mathbf{X}} \geq 0$.
  - Todos los autovalores son positivos, $\lambda_i \geq 0, i = 1, \cdots n$
  - Los autovectores forman una base ortonormal, $p_i^t p_j = \delta_{ij}, i, j = 1, \cdots n$
  - Sea P la matriz de autovectores y Λ la de autovalores. Luego,

$$C_{\mathbf{X}} = P \Lambda P^t$$

donde $PP^t = P^t P = \mathbf{I}_n$ y $\Lambda = diag(\lambda_i)$.

Propiedades de $C_{\mathbf{X}}$: VeA degenerado

Qué sucede si $C_{\mathbf{X}}$ tiene autovalores nulos? Es decir, $C_{\mathbf{X}}$ es una matriz singular?

- Recordemos que sucede con $X \in \mathbb{R}$, $\mathbb{E}[X] = \mu_X$ y $\mathbb{V}(X) = \sigma_X^2$

$$\sigma_X^2 = 0 \Longleftrightarrow \mathbb{P}[X = \mu_X] = 1$$

  La V.A se comporta como una constante igual a su media.

- Para el caso *n*-dimensional tenemos el siguiente resultado:

  Teorema. $C_{\mathbf{X}}$ es singular si y solo si existe un vector $\boldsymbol{v} \neq \mathbf{0}$ tal que $\mathbb{P}(\boldsymbol{v}^t(\mathbf{X} - \mu_{\mathbf{X}}) = 0) = 1$.

En este caso decimos que **X** es un VeA *degenerado*.

Propiedades de $C_{\mathbf{X}}$: VeA degenerado

- Un modo interpretar este resultado es ver que si $\mathbf{X}$ es un VeA degenerado, entonces existe una dirección fija ($\boldsymbol{v}$) que va a ser ortogonal a todas las realizaciones ($\boldsymbol{x} - \mu_{\mathbf{X}}$). Es decir que con probabilidad 1 NO vamos a observar realizaciones de $\mathbf{X}$ que se encuentren en el espacio generado por $\boldsymbol{v}$ desplazado en $\mu_{\mathbf{X}}$.
- Es decir, hay componentes del VeA redundantes que se pueden determinar a partir de las demás.

VeA degenerados: demostración del teorema

- $C_{\mathbf{X}}$ singular entonces $\exists \boldsymbol{v} \neq \mathbf{0}$: $\mathbb{P}(\boldsymbol{v}^t(\mathbf{X} - \mu_{\mathbf{X}}) = 0) = 1$.
  $C_{\mathbf{X}}$ singular implica que $C_{\mathbf{X}}$ existe $\boldsymbol{v} \neq \mathbf{0}$ tal que $\boldsymbol{C}_{\mathbf{X}} \boldsymbol{v} = \mathbf{0}$.
  Entonces $\boldsymbol{v}^t \boldsymbol{C}_{\mathbf{X}} \boldsymbol{v} = 0$. Ahora definimos $Y = \boldsymbol{v}^t(\mathbf{X} - \mu_{\mathbf{X}})$. Para cualquier $a > 0$, la cota de Markov establece que

  $$\mathbb{P}(Y^2 \geq a) \leq \frac{\mathbb{E}[Y^2]}{a} = \frac{\boldsymbol{v}^t \boldsymbol{C}_{\mathbf{X}} \boldsymbol{v}}{a} = 0,$$

  Luego, $\mathbb{P}(Y^2 = 0) = 1$, y por ende, $\mathbb{P}(\boldsymbol{v}^t(\mathbf{X} - \mu_{\mathbf{X}}) = 0) = 1$.
- $\exists \boldsymbol{v} \neq \mathbf{0}$: $\mathbb{P}(\boldsymbol{v}^t(\mathbf{X} - \mu_{\mathbf{X}}) = 0) = 1$ luego, $C_{\mathbf{X}}$ es singular.
  $Y = \boldsymbol{v}^t(\mathbf{X} - \mu_{\mathbf{X}}) \in \mathbb{R}$ es una V.A. de media nula. Vimos que si $\mathbb{P}[Y = 0] = 1$ implica que su varianza es nula, es decir

  $$\sigma_Y^2 = \mathbb{E}[Y^2] = \boldsymbol{v}^t \boldsymbol{C}_{\mathbf{X}} \boldsymbol{v} = 0.$$

  Esto implica que $\boldsymbol{C}_{\mathbf{X}}$ es singular.

$\square$

Transformación afín de $\mathbf{X}$

Sea $\mathbf{X}$ un VeA con covarianza $C_{\mathbf{X}}$. Consideramos una transformación afín formada a partir de la matriz $A \in \mathbb{R}^{m\times n}$ y el vector $b \in \mathbb{R}^m$,

$$\mathbf{Y} = A\mathbf{X} + b.$$

Entonces

$$\mathbb{E}[\mathbf{Y}] = \mathbb{E}[A\mathbf{X} + b] = A\mathbb{E}[\mathbf{X}] + b$$

$$\begin{aligned}
\mathbf{Cov}(\mathbf{Y}) &= \mathbb{E}\left\{[\mathbf{Y} - \mathbb{E}[\mathbf{Y}]][\mathbf{Y} - \mathbb{E}[\mathbf{Y}]]^t\right\} \\
&= \mathbb{E}\left\{[A\mathbf{X} + b - A\mathbb{E}[\mathbf{X}] - b][A\mathbf{X} + b - A\mathbb{E}[\mathbf{X}] - b]^t\right\} \\
&= \mathbb{E}\left\{[A(\mathbf{X} - \mathbb{E}[\mathbf{X}])]\,[A(\mathbf{X} - \mathbb{E}[\mathbf{X}])]^t\right\} \\
&= \mathbb{E}\left\{[A(\mathbf{X} - \mathbb{E}[\mathbf{X}])][(\mathbf{X} - \mathbb{E}[\mathbf{X}])]^t A^t\right\} \\
&= A\mathbb{E}\left\{[\mathbf{X} - \mathbb{E}[\mathbf{X}]][\mathbf{X} - \mathbb{E}[\mathbf{X}]]^t\right\} A^t = A C_{\mathbf{X}} A^t
\end{aligned}$$

Transformación afín (recap importante!)

$\mathbf{Y} = A\mathbf{X} + b$

$$\mathbb{E}[\mathbf{Y}] = A\mathbb{E}[\mathbf{X}] + b$$

$$\mathbf{Cov}(\mathbf{Y}) = C_{\mathbf{Y}} = AC_{\mathbf{X}}A^t$$

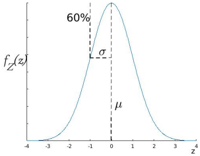
Observaciones:

- b desplaza la media
- b no afecta la covarianza

Distribución Gaussiana Multivariable

Repaso de distribución gaussiana o normal

La VA Z tiene distribución gaussiana, o $Z \sim \mathcal{N}(\mu, \sigma^2)$ si



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

Algunas propiedades:

- $f_Z(z)$ es máxima en μ y simétrica,

$$f_Z(\mu - z) = f_Z(\mu + z) \quad , \quad \forall z \in \mathbb{R}.$$

- $\mathbb{E}[Z] = \mu \quad$ y $\quad \mathbb{V}[Z] = \mathbb{E}\left[(Z-\mu)^2\right] = \sigma^2$.

Repaso de distribución gaussiana o normal

gaussiana estándar o normal estándar , $\mathcal{N}(0,1)$.

$$F_Z(z) = \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2} du.$$

Función $Q(z)$:

$$Q(z) = 1 - \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{z}^{+\infty} e^{-u^2/2} du.$$

Ambas funciones se obtienen de tablas.

Distribución gaussiana Multivariable (Definición 1)

Un VeA $\mathbf{X}$ tiene una distribución gaussiana multivariable si existe un vector $\boldsymbol{\mu}_{\mathbf{X}}$ y una matriz simétrica $C_{\mathbf{X}} \geq 0$ tal que la PDF conjunta es:

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

Decimos que $\mathbf{X} \sim \mathcal{N}(\mu_{\mathbf{X}}, C_{\mathbf{X}})$. Es posible demostrar que

- Esperanza: $\mathbb{E}[\mathbf{X}] = \boldsymbol{\mu}_{\mathbf{X}}$.
- Covarianza: $\mathbf{Cov}[\mathbf{X}] = C_{\mathbf{X}}$

Cuando $\boldsymbol{\mu}_{\mathbf{X}} = 0$ y $C_{\mathbf{X}} = \mathbf{I}_n$ tenemos una distribución normal estándar.

Distribución Gaussiana Multivariable (Definición 2)

$\mathbf{X}$ tiene una distribución gaussiana multivariable si y sólo si, $\forall \mathbf{A} \in \mathbb{R}^{m \times n}$, $m \leq n$, $\mathbf{Y} = \mathbf{AX}$ sigue una distribución gaussiana. Esta última queda especificada por

$$\mathbb{E}[\mathbf{Y}] = \mathbf{A}\mu_{\mathbf{X}} \qquad \mathbf{Cov}(\mathbf{Y}) = \mathbf{A}C_{\mathbf{X}}\mathbf{A}^T$$

Observaciones:

- Se puede partir de la Definición 1 y probar la Definición 2 como una propiedad o viceversa.
- La distribución gaussiana queda completamente definida por la esperanza y su matriz de covarianza.

Distribuciones marginales

Si $\mathbf{X}$ es un VeA gaussiano, luego sus componentes y todo subvector son gaussianos.

Este resultado es una consecuencia de que toda transformación lineal de un VeA gaussiano produce otro VeA gaussiano.

- Como $X_i = \mathbf{1}_i^T \mathbf{X}$,, donde $\mathbf{1}_i = [0, \ldots, 0, 1, 0, \ldots, 0]^t$, entonces X_i es el resultado de una transformación lineal de $\mathbf{X}$. Por lo tanto, X_i es gaussiana con

$$\mathbb{E}[X_i] = \mathbf{1}_i^t \mu_{\mathbf{X}} = (\mu_{\mathbf{X}})_i, \qquad \mathbb{V}[X_i] = \mathbf{1}_i^t C_{\mathbf{X}} \mathbf{1}_i = (C_{\mathbf{X}})_{i,i}.$$

- En forma más general, un subvector $\mathbf{X}_I = [X_{i_1}, \ldots, X_{i_k}]^t$ de $\mathbf{X}$ es gaussiano.

$$\mathbf{X}_I = \begin{bmatrix} \mathbf{1}_{i_1}^t \\ \vdots \\ \mathbf{1}_{i_k}^t \end{bmatrix} \mathbf{X} = \mathbf{1}_I^t \mathbf{X} \quad \longrightarrow \quad \begin{cases} \mathbb{E}[\mathbf{X}_I] = \mathbf{1}_I^t \mu_{\mathbf{X}} = (\mu_{\mathbf{X}})_I \\ \mathbf{Cov}(\mathbf{X}_I) = \mathbf{1}_I^t C_{\mathbf{X}} \mathbf{1}_I = (C_{\mathbf{X}})_{I,I} \end{cases}$$

Gaussiana condicionada por Gaussiana

PDF condicional

Sea $\mathbf{X} \in \mathbb{R}^n$, $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{X}}, C_{\mathbf{X}})$,

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \quad , \quad \mathbf{X}_1 \in \mathbb{R}^{n_1} \ , \ \mathbf{X}_2 \in \mathbb{R}^{n_2} \qquad n_1 + n_2 = n.$$

Buscamos la distribución condicional de $\mathbf{X}_1$ cuando $\mathbf{X}_2 = \boldsymbol{x}_2$. Vamos a demostrar que

$f_{\mathbf{X}_1|\mathbf{X}_2}(\boldsymbol{x}_1|\mathbf{X}_2 = \boldsymbol{x}_2)$ es gaussiana con

- $\mathbb{E}[\mathbf{X}_1|\mathbf{X}_2 = \boldsymbol{x}_2] = \mu_{\mathbf{X}_1} + C_{\mathbf{X}_1,\mathbf{X}_2} C^{-1}_{\mathbf{X}_2,\mathbf{X}_2}(\boldsymbol{x}_2 - \mu_{\mathbf{X}_2})$
- $\mathbf{Cov}(\mathbf{X}_1|\mathbf{X}_2 = \boldsymbol{x}_2) = C_{\mathbf{X}_1} - C_{\mathbf{X}_1,\mathbf{X}_2} C^{-1}_{\mathbf{X}_2} C_{\mathbf{X}_2,\mathbf{X}_1}.$

PDF condicional cuando $n = 2$

- Consideremos primero el caso bidimensional, es decir, $\mathbf{X} = [X_1, X_2]^t$.

$$
\begin{aligned}
f_{X_1|X_2}(x_1|X_2 = x_2) &= \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_2}(x_2)} \\
&= \frac{\frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{1/2}} e^{-\frac{1}{2}\frac{1}{1-\rho^2}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{x_1-\mu_1}{\sigma_1}\frac{x_2-\mu_2}{\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]}}{\frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2}\frac{(x_2-\mu_2)^2}{\sigma_2^2}}} \\
&= \frac{1}{\sqrt{2\pi\sigma_1^2(1-\rho^2)}} e^{-\frac{1}{2\sigma_1^2(1-\rho^2)}\left[x_1 - \rho\frac{\sigma_1}{\sigma_2}(x_2-\mu_2) - \mu_1\right]^2}.
\end{aligned}
$$

- Luego, $f_{X_1|X_2}(x_1|X_2 = x_2)$ corresponde a una pdf gaussiana con

$$
\mathbb{E}[X_1|X_2 = x_2] = \mu_1 + \frac{\rho\sigma_{X_1}}{\sigma_2}(x_2 - \mu_2).
$$

$$
\mathbb{V}[X_1|X_2 = x_2] = (1 - \rho^2)\sigma_1^2.
$$

PDF condicional cuando $n = 2$

- La media condicional $\mu_{X_1|X_2=x_2}$ es una función afín de x_2.
- La varianza condicional $\sigma^2_{X_1|X_2=x_2}$ es independiente del valor de x_2.
- Si $\rho = 0$, X_1 y X_2 son independientes. Luego, $f_{X_1|X_2}(x_1|X_2 = x_2) = f_{X_1}(x_1)$. Esto queda claro en las ecuaciones anteriores considerando $\rho = 0$, $\mu_{X_1|X_2=x_2} = \mu_1$ y $\sigma^2_{X_1|X_2=x_2} = \sigma^2_1$.
- Si $|\rho| = 1$, las variables están perfectamente correlacionadas. Luego al observar $\mathbf{X}_2 = \boldsymbol{x}_2$ podemos determinar perfectamente $\mathbf{X}_1$, es decir $\sigma^2_{X_1|X_2=x_2} = 0$. Nuevamente, esto se verifica en las ecuaciones anteriores.

PDF condicional: caso general

Partimos de la definición

$$f_{\mathbf{X}_1|\mathbf{X}_2}(\boldsymbol{x}_1|\boldsymbol{x}_2) = \frac{f_{\mathbf{X}_1,\mathbf{X}_2}(\boldsymbol{x}_1,\boldsymbol{x}_2)}{f_{\mathbf{X}_2}(\boldsymbol{x}_2)} = \frac{f_{\mathbf{X}}(\boldsymbol{x})}{f_{\mathbf{X}_2}(\boldsymbol{x}_2)}$$

$$= \frac{\frac{1}{(2\pi)^{n/2}\det(C_{\mathbf{X}})^{1/2}}}{\frac{1}{(2\pi)^{n_2/2}\det(C_{\mathbf{X}_2})^{1/2}}} \frac{e^{-\frac{1}{2}(\boldsymbol{x}-\mu_{\mathbf{X}})^t C_{\mathbf{X}}^{-1}(\boldsymbol{x}-\mu_{\mathbf{X}})}}{e^{-\frac{1}{2}(\boldsymbol{x}_2-\mu_{\mathbf{X}_2})^t C_{\mathbf{X}_2}^{-1}(\boldsymbol{x}_2-\mu_{\mathbf{X}_2})}}$$

Para calcular $f_{\mathbf{X}}$ necesitamos invertir $C_{\mathbf{X}}$ y calcular su determinante. Para ello, particionamos la matriz de covarianza

$$C_{\mathbf{X}} = \begin{bmatrix} C_{\mathbf{X}_1} & C_{\mathbf{X}_1,\mathbf{X}_2} \\ C_{\mathbf{X}_2,\mathbf{X}_1} & C_{\mathbf{X}_2} \end{bmatrix}.$$

Presentamos dos resultados auxiliares para matrices en bloques.

Inversión de matrices por bloques

Sea $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ una matriz no singular particionada del siguiente modo:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{1,1} & \boldsymbol{A}_{1,2} \\ \boldsymbol{A}_{2,1} & \boldsymbol{A}_{2,2} \end{bmatrix}$$

Definimos

$$\mathbf{A}_* = \mathbf{A}_{1,1} - \mathbf{A}_{1,2}\mathbf{A}_{2,2}^{-1}\mathbf{A}_{2,1}.$$

Esta matriz se la conoce como el complemento de Schur del bloque $\mathbf{A}_{2,2}$ de la matriz $\mathbf{A}$. Asumimos que todas las inversas involucradas existen. Luego:

- $$\boldsymbol{A}^{-1} = \begin{bmatrix} \mathbf{A}_*^{-1} & -\mathbf{A}_*^{-1}\boldsymbol{A}_{1,2}\boldsymbol{A}_{2,2}^{-1} \\ -\boldsymbol{A}_{2,2}^{-1}\boldsymbol{A}_{2,1}\mathbf{A}_*^{-1} & \boldsymbol{A}_{2,2}^{-1} + \boldsymbol{A}_{2,2}^{-1}\boldsymbol{A}_{2,1}\mathbf{A}_*^{-1}\boldsymbol{A}_{1,2}\boldsymbol{A}_{2,2}^{-1} \end{bmatrix}$$
- $$\det(\boldsymbol{A}) = \det(\boldsymbol{A}_{2,2})\det(\mathbf{A}_*).$$

PDF condicional: caso general

Retomamos la derivación de $f_{\mathbf{X}_1|\mathbf{X}_2}$

- Aplicando el resultado de inversión por bloques a la matriz $C_{\mathbf{X}}$ calculamos el complemento de Schur

$$C_{\mathbf{X}*} = C_{\mathbf{X}_1} - C_{\mathbf{X}_1,\mathbf{X}_2} C_{\mathbf{X}_2}^{-1} C_{\mathbf{X}_2,\mathbf{X}_1}$$

Luego,

$$\begin{aligned} C_{\mathbf{X}}^{-1} &= \begin{bmatrix} C_{\mathbf{X}_1} & C_{\mathbf{X}_1,\mathbf{X}_2} \\ C_{\mathbf{X}_2,\mathbf{X}_1} & C_{\mathbf{X}_2} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} C_{\mathbf{X}*}^{-1} & -C_{\mathbf{X}*}^{-1} C_{\mathbf{X}_1,\mathbf{X}_2} C_{\mathbf{X}_2}^{-1} \\ -C_{\mathbf{X}_2}^{-1} C_{\mathbf{X}_2,\mathbf{X}_1} C_{\mathbf{X}*}^{-1} & C_{\mathbf{X}_2}^{-1} + C_{\mathbf{X}_2}^{-1} C_{\mathbf{X}_2,\mathbf{X}_1} C_{\mathbf{X}*}^{-1} C_{\mathbf{X}_1,\mathbf{X}_2} C_{\mathbf{X}_2}^{-1} \end{bmatrix} \end{aligned}$$

$$\det(C_{\mathbf{X}}) = \det(C_{\mathbf{X}_2}) \det(C_{\mathbf{X}*}).$$

PDF condicional: caso general

- Utilizando la partición de $C_{\mathbf{X}}^{-1}$, calculamos:

$$
\begin{aligned}
(\boldsymbol{x}-\mu_{\mathbf{X}})^t C_{\mathbf{X}}^{-1}(\boldsymbol{x}-\mu_{\mathbf{X}}) &= \begin{bmatrix} \boldsymbol{x}_1-\mu_{\mathbf{X}_1} \\ \boldsymbol{x}_2-\mu_{\mathbf{X}_2} \end{bmatrix}^t \begin{bmatrix} C_{\mathbf{X}_1} & C_{\mathbf{X}_1,\mathbf{X}_2} \\ C_{\mathbf{X}_2,\mathbf{X}_1} & C_{\mathbf{X}_2} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{x}_1-\mu_{\mathbf{X}_1} \\ \boldsymbol{x}_2-\mu_{\mathbf{X}_2} \end{bmatrix} \\
&= (\boldsymbol{x}_1-\mu_{\mathbf{X}_1})^t C_{\mathbf{X}*}^{-1}(\boldsymbol{x}_1-\mu_{\mathbf{X}_1}) \\
&\quad - (\boldsymbol{x}_1-\mu_{\mathbf{X}_1})^t C_{\mathbf{X}*}^{-1} C_{\mathbf{X}_1,\mathbf{X}_2} C_{\mathbf{X}_2}^{-1}(\boldsymbol{x}_2-\mu_{\mathbf{X}_2}) - (\boldsymbol{x}_2-\mu_{\mathbf{X}_2})^t C_{\mathbf{X}_2}^{-1} C_{\mathbf{X}_2,\mathbf{X}_1} C_{\mathbf{X}*}^{-1}(\boldsymbol{x}_1-\mu_{\mathbf{X}_1}) \\
&\quad + (\boldsymbol{x}_2-\mu_{\mathbf{X}_2})^t \left\{ C_{\mathbf{X}_2}^{-1} + C_{\mathbf{X}_2}^{-1} C_{\mathbf{X}_2,\mathbf{X}_1} C_{\mathbf{X}*}^{-1} C_{\mathbf{X}_1,\mathbf{X}_2} C_{\mathbf{X}_2}^{-1} \right\} (\boldsymbol{x}_2-\mu_{\mathbf{X}_2})
\end{aligned}
$$

$$
= \Bigg[\boldsymbol{x}_1 - \underbrace{\left(\mu_{\mathbf{X}_1} + C_{\mathbf{X}_1,\mathbf{X}_2} C_{\mathbf{X}_2}^{-1} C_{\mathbf{X}_2,\mathbf{X}_1}(\boldsymbol{x}_2-\mu_{\mathbf{X}_2})\right)}_{\mu_*}\Bigg]^t C_{\mathbf{X}*}^{-1} \Bigg[\boldsymbol{x}_1 - \underbrace{\left(\mu_{\mathbf{X}_1} + C_{\mathbf{X}_1,\mathbf{X}_2} C_{\mathbf{X}_2}^{-1} C_{\mathbf{X}_2,\mathbf{X}_1}(\boldsymbol{x}_2-\mu_{\mathbf{X}_2})\right)}_{\mu_*}\Bigg]
$$

$$
+ (\boldsymbol{x}_2-\mu_{\mathbf{X}_2})^t C_{\mathbf{X}_2}^{-1}(\boldsymbol{x}_2-\mu_{\mathbf{X}_2}).
$$

Definimos $\mu_* = \mu_{\mathbf{X}_1} + C_{\mathbf{X}_1,\mathbf{X}_2} C_{\mathbf{X}_2}^{-1} C_{\mathbf{X}_2,\mathbf{X}_1}(\boldsymbol{x}_2-\mu_{\mathbf{X}_2})$.

PDF condicional: caso general

- Incorporando esta expresión y la de $\det(C_{\mathbf{X}})$ en $f_{\mathbf{X}_1|\mathbf{X}_2}(\boldsymbol{x}_1|\boldsymbol{x}_2)$ obtenemos

$$e^{-\frac{1}{2}\left[(\boldsymbol{x}-\mu_{\mathbf{X}})^t C_{\mathbf{X}}^{-1}(\boldsymbol{x}-\mu_{\mathbf{X}})-(\boldsymbol{x}_2-\mu_{\mathbf{X}_2})^t C_{\mathbf{X}_2}^{-1}(\boldsymbol{x}_2-\mu_{\mathbf{X}_2})\right]} = e^{-\frac{1}{2}(\boldsymbol{x}_1-\mu_*)^t C_{\mathbf{X}*}^{-1}(\boldsymbol{x}_1-\mu_*)}$$

$$\frac{\frac{1}{(2\pi)^{n/2}\det(C_{\mathbf{X}})^{1/2}}}{\frac{1}{(2\pi)^{n_2/2}\det(C_{\mathbf{X}_2})^{1/2}}} = \frac{1}{(2\pi)^{n_1/2}\det(C_{\mathbf{X}}*)^{1/2}}$$

- Luego,

$$f_{\mathbf{X}_1|\mathbf{X}_2}(\boldsymbol{x}_1|\boldsymbol{x}_2) = \frac{1}{(2\pi)^{n_1/2}\det(C_{\mathbf{X}}*)^{1/2}} e^{-\frac{1}{2}(\boldsymbol{x}_1-\mu_*)^t C_{\mathbf{X}*}^{-1}(\boldsymbol{x}_1-\mu_*)}.$$

Ésta es una pdf gaussiana de media μ_* y covarianza $C_{\mathbf{X}*}$, es decir,

$$f_{\mathbf{X}_1|\mathbf{X}_2}(\boldsymbol{x}_1|\boldsymbol{x}_2) = \mathcal{N}(\mu_{\mathbf{X}_1} + C_{\mathbf{X}_1,\mathbf{X}_2} C_{\mathbf{X}_2}^{-1}(\boldsymbol{x}_2 - \mu_{\mathbf{X}_2}), C_{\mathbf{X}_1} - C_{\mathbf{X}_1,\mathbf{X}_2} C_{\mathbf{X}_2}^{-1} C_{\mathbf{X}_2,\mathbf{X}_1}).$$

Transformación afín de un VeA gaussiano

Transformaciones afines de un VeA gaussiano

Vimos que si $\mathbf{X} \in \mathbb{R}^n$ es un VeA Gaussiano, luego toda transformación lineal del mismo es también gaussiano. Analicemos qué ocurre cuando tomamos una transformación afín. Sean $A \in \mathbb{R}^{m \times n}$ y $b \in \mathbb{R}^m$. Noten que m puede ser distinto a n, es decir A puede ser rectangular. Definimos

$$\mathbf{Y} = A\mathbf{X} + b.$$

- Vimos que en una transformación afín

$$\mu_{\mathbf{Y}} = A\mathbb{E}[\mathbf{X}] + b, \qquad C_{\mathbf{Y}} = AC_{\mathbf{X}}A^t$$

- Sabemos que $A\mathbf{X}$ sigue una distribución gaussiana. Luego, $\mathbf{Y} \sim \mathcal{N}(\mu_{\mathbf{Y}}, C_{\mathbf{Y}})$.
- Usando transformaciones afines, podemos transformar cualquier distribución gaussiana en una normal estándar y viceversa. Vamos a trabajar sólo con gaussianas no-degeneradas.

Transformaciones afines de un VeA gaussiano

Sea $\mathbf{X} \sim \mathcal{N}(\mu_{\mathbf{X}}, C_{\mathbf{X}})$, con $C_{\mathbf{X}} > 0$.

- Descomponemos $C_{\mathbf{X}} = P_{\mathbf{X}} \Lambda_{\mathbf{X}} P_{\mathbf{X}}^t$
- Como $\Lambda_{\mathbf{X}} > 0$, entonces $\Lambda_{\mathbf{X}} = \Lambda_{\mathbf{X}}^{1/2} \Lambda_{\mathbf{X}}^{1/2}$
- Definimos una transformación $\mathbf{Z} = A\mathbf{X}$ con $A = \Lambda_{\mathbf{X}}^{-1/2} P_{\mathbf{X}}^t$. Entonces

$$C_{\mathbf{Z}} = \underbrace{\Lambda_{\mathbf{X}}^{-1/2} P^t}_{A} C_{\mathbf{X}} \underbrace{P_{\mathbf{X}} \Lambda_{\mathbf{X}}^{-1/2}}_{A^t} = \Lambda_{\mathbf{X}}^{-1/2} P_{\mathbf{X}}^t P_{\mathbf{X}} \Lambda_{\mathbf{X}} P_{\mathbf{X}}^t P_{\mathbf{X}} \Lambda_{\mathbf{X}}^{-1/2} = \mathbf{I}_n$$

- Completamos con el desplazamiento $b = -A\mu_{\mathbf{X}}$, $\mathbf{Z} = A\mathbf{X} + b$. Luego este desplazamiento no afecta la covarianza pero desplaza la media.

$$\mathbb{E}[\mathbf{Z}] = A\mu_{\mathbf{X}} + (-A\mu_{\mathbf{X}}) = 0$$

Transformaciones afines de $\mathbf{X}$

Normalización de VeA Gaussiano

Sea $\mathbf{X} \sim \mathcal{N}(\mu_{\mathbf{X}}, C_{\mathbf{X}})$. La transformación

$$\mathbf{Z} = \Lambda_{\mathbf{X}}^{-1/2} P_{\mathbf{X}}^{t} (\mathbf{X} - \mu_{\mathbf{X}})$$

resulta en un VeA normal estandar, es decir $\mathbf{Z} \sim \mathcal{N}(0, \mathbf{I}_n)$

Transformaciones afines de $\mathbf{X}$

Del mismo modo, partiendo de un VeA normal estándar, podemos obtener un VeA gaussiano con media y covarianza dadas.

Normalización de VeA Gaussiano

Sean $\mathbf{Z} \sim \mathcal{N}(0, \mathbf{I}_n)$, $\mathbf{Y} \sim \mathcal{N}(\mu_{\mathbf{Y}}, C_{\mathbf{Y}})$ donde $C_{\mathbf{Y}} = P_{\mathbf{Y}} \Lambda_{\mathbf{Y}} P_{\mathbf{Y}}^t$. Luego,

$$\mathbf{Y} = P_{\mathbf{Y}} \Lambda_{\mathbf{Y}}^{1/2} \mathbf{Z} + \mu_{\mathbf{Y}}$$

Transformaciones afines de $\mathbf{X}$ (recap)

$$\mathbf{X} \sim \mathcal{N}(\mu_{\mathbf{X}}, C_{\mathbf{X}}) \qquad C_{\mathbf{X}} = P_{\mathbf{X}} \Lambda_{\mathbf{X}} P_{\mathbf{X}}^t$$

$$\mathbf{Y} \sim \mathcal{N}(\mu_{\mathbf{Y}}, C_{\mathbf{Y}}) \qquad C_{\mathbf{Y}} = P_{\mathbf{Y}} \Lambda_{\mathbf{Y}} P_{\mathbf{Y}}^t$$

$$\begin{tikzcd}
\mathbf{X} \sim \mathcal{N}(\mu_{\mathbf{X}}, C_{\mathbf{X}}) \arrow[rr, bend left, "\Lambda_{\mathbf{X}}^{-1/2} P_{\mathbf{X}}^t (\mathbf{X} - \mu_{\mathbf{X}})"] \arrow[dr, bend right] & & \mathbf{Z} \sim \mathcal{N}(0, \mathbf{I}) \arrow[dl, bend left, "P_{\mathbf{Y}} \Lambda_{\mathbf{Y}}^{1/2} \mathbf{Z} + \mu_{\mathbf{Y}}"] \\
& \mathbf{Y} \sim \mathcal{N}(\mu_{\mathbf{Y}}, C_{\mathbf{Y}}) &
\end{tikzcd}$$

Relación entre VeA Gaussiano y la distribución χ^2

Distribución χ^2

Sea $\mathbf{X} \in \mathbb{R}^n$ un VeA Gaussiano a partir del cual formamos la V.A.

$$Y = (\mathbf{X} - \mu_{\mathbf{X}})^T C_{\mathbf{X}}^{-1} (\mathbf{X} - \mu_{\mathbf{X}}).$$

$$Y \sim \chi_n^2$$

Y tiene una distribución chi-cuadrado con n grados de libertad.

Relación entre VeA Gaussiano y la distribución χ^2

1. Sea $\boldsymbol{Z} = C_{\mathbf{X}}^{-1/2}(\mathbf{X} - \mu_{\mathbf{X}})$. Por ser transformación lineal de VeA gaussiano es gaussiano. Más aún, $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$.
2. Entonces, cada $Z_i \sim \mathcal{N}(0, 1)$ y son independientes.
3. $$\begin{aligned} Y &= (\mathbf{X} - \mu_{\mathbf{X}})^T C_{\mathbf{X}}^{-1}(\mathbf{X} - \mu_{\mathbf{X}}) \\ &= (\mathbf{X} - \mu_{\mathbf{X}})^T C_{\mathbf{X}}^{-1/2} C_{\mathbf{X}}^{-1/2}(\mathbf{X} - \mu_{\mathbf{X}}) = \boldsymbol{Z}^t \boldsymbol{Z} = \sum_{i=1}^{n} Z_i^2. \end{aligned}$$
4. Cada término Z_i^2 tiene una distribución chi-cuadrado con 1 grado de libertad.
5. Luego, $Y \sim \chi_n^2$.

Función Característica de una gaussiana

Función característica de V.A. normal estándar

Vamos a trabajar el problema para una sola variable aleatoria. Para ello, partimos de la definición de Función Característica. Para el caso de la normal estándar $\mathcal{N}(0,1)$ tenemos,

$$\Phi_Z(\omega) = \mathbb{E}[e^{\jmath\omega Z}] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{\jmath\omega z} e^{-\frac{z^2}{2}} dz$$
$$= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{\omega^2}{2}} e^{-\frac{(z-\jmath\omega)^2}{2}} dz = e^{-\frac{\omega^2}{2}} \underbrace{\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{u^2}{2}} du}_{1} = e^{-\frac{\omega^2}{2}}.$$

La FC resulta entonces

$$\Phi_Z(\omega) = e^{-\frac{\omega^2}{2}}, \qquad \omega \in \mathbb{R}.$$

Función característica de V.A. gaussiana

A partir de la normal estándar, podemos obtener cualquier distribución gaussiana. En particular, si $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$,

$$X = \sigma_X Z + \mu_X.$$

Luego, aplicando las propiedades de las funciones características,

$$\Phi_X(\omega) = e^{\jmath\omega\mu_X}\Phi_Z(\sigma_X\omega) = e^{\jmath\omega\mu_X}e^{-\frac{\sigma_X^2\omega^2}{2}}$$

Observaciones

- $\Phi_Z(\omega) = e^{-\frac{\omega^2}{2}}$ y $f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$ tienen la misma forma funcional. Difieren solamente en un factor de escala.
- La media μ_X determina un factor de fase de $\Phi_X(\omega)$. El módulo sólo depende de σ_X.
- A mayor valor de σ_X, $f_X(x)$ es una "campana" más dispersa alrededor de μ_X y $\Phi_X(\omega)$ es una "campana" más concentrada alrededor del origen.

Momentos de VA gaussiana

- Comenzamos con la normal estándar. A partir de $\Phi_Z(\omega)$ hallamos sus momentos utilizando el teorema de momentos. Desarrollamos en serie de Taylor:

$$\begin{aligned}\Phi_Z(\omega) = e^{-\frac{\omega^2}{2}} &= \sum_{k=0}^{\infty} \frac{\left(-\omega^2/2\right)^k}{k!} \\ &= \sum_{k=0}^{\infty} \left(-\frac{1}{2}\right)^k \frac{(2k)!}{k!} \frac{\omega^{2k}}{(2k)!} = \sum_{k=0}^{\infty} \Phi_Z^{(k)}(0) \frac{\omega^k}{k!}.\end{aligned}$$

- Igualando los términos de igual potencia, se tiene que

$$\left.\frac{d^{(k)}\Phi_Z(\omega)}{d\omega^{(k)}}\right|_{\omega=0} = \Phi_Z^{(k)}(0) = \begin{cases} 0 & \text{si } k \text{ es impar,} \\ \left(-\frac{1}{2}\right)^{(k/2)} \frac{k!}{(k/2)!} & \text{si } k \text{ es par.} \end{cases}$$

Momentos de VA gaussiana

- El teorema de momentos establece que

$$\mathbb{E}[Z^k] = \frac{1}{\jmath^k}\Phi_Z^{(k)}(0).$$

Por ende,

$$\mathbb{E}[Z^{(2m+1)}] = 0 \qquad \mathbb{E}[Z^{2m}] = \frac{1}{2^m}\frac{(2m)!}{m!}$$

Verificamos el resultado con los primeros dos momentos:

$$\mathbb{E}[Z] = \frac{1}{j}\Phi_Z^{(1)}(0) = 0 \quad \mathbb{V}(Z) = \mathbb{E}[Z^2] = \frac{1}{j^2}\Phi_Z^{(2)}(0) = -\left(-\frac{1}{2}\right)2 = 1.$$

- En forma general, para $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$

$$\mathbb{E}[(X-\mu_X)^{(2m+1)}] = 0 \qquad \mathbb{E}[(X-\mu_X)^{2m}] = \sigma_X^{2m}\frac{1}{2^m}\frac{(2m)!}{m!}$$

Independencia y Correlación

Observación

- En general

$$X_1, \cdots X_n \text{ V.A independientes} \Longrightarrow X_1, \cdots X_n \text{ V.A descorrelacionadas}$$

Claramente, si X_i, X_j son independientes, su covarianza es

$$Cov(X_i, X_j) = \mathbb{E}\left[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])\right] = \mathbb{E}\left[X_i - \mathbb{E}[X_i]\right]\mathbb{E}\left[X_j - \mathbb{E}[X_i]\right] = 0 \quad \square$$

Pero

$$X_1, \cdots X_n \text{ V.A descorrelacionadas} \not\Longrightarrow X_1, \cdots X_n \text{ V.A independientes}$$

- Se plantea una excepción cuando $X_1, \cdots X_n$ son V.A conjuntamente gaussianas

Gaussianas Independientes

El vector $\mathbf{X} \sim \mathcal{N}(\mu_{\mathbf{X}}, C_{\mathbf{X}})$ tiene componentes descorrelacionadas (es decir,$C_{\mathbf{X}}$ es diagonal) si y sólo si sus componentes son independientes, es decir,

$$C_{\mathbf{X}} \text{ es diagonal } \iff f_{\mathbf{X}}(\boldsymbol{x}) = \prod_{i=1}^{n} f_{X_i}(x_i).$$

En este caso,

$$X_1, \cdots X_n \text{ V.A descorrelacionadas} \iff X_1, \cdots X_n \text{ V.A independientes}$$

Demostracion

- $(\Longrightarrow)$: $C_{\mathbf{X}} = \operatorname{diag}(\sigma_1^2, \cdots \sigma_n^2)$. Luego,

$$(\boldsymbol{x} - \mu_{\mathbf{X}})^T \, C_{\mathbf{X}}^{-1} \, (\boldsymbol{x} - \mu_{\mathbf{X}}) = \sum_{i=1}^{n} \frac{(x_i - \mu_i)}{\sigma_i^2} \quad , \quad \det(C_{\mathbf{X}}) = \prod_{i=1}^{n} \sigma_i^2$$

Usando las propiedades de la función exponencial,

$$f_{\mathbf{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2} \det(C_{\mathbf{X}})^{1/2}} e^{-\frac{1}{2}(\boldsymbol{x}-\mu_{\mathbf{X}})^T C_{\mathbf{X}}^{-1}(\boldsymbol{x}-\mu_{\mathbf{X}})} = \prod_{i=1}^{n} \underbrace{\frac{1}{(2\pi\sigma_i^2)^{1/2}} e^{-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}}}_{f_{X_i}(x_i)} .$$

- $(\Longleftarrow)$: $f_{\mathbf{X}}(\boldsymbol{x}) = \prod f_{X_i}(x_i)$. Por ser $\mathbf{X}$ una gaussiana multivariable, sus componentes tienen que ser VA gaussianas. Luego,

$$f_{X_i}(x_i) = \frac{e^{-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}}}{\sqrt{2\pi\sigma_i^2}} \Longrightarrow f_{\mathbf{X}}(\boldsymbol{x}) = \prod_{i=1}^{n} \underbrace{\frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}}}_{f_{X_i}(x_i)} \Longrightarrow C_{\mathbf{X}} = \operatorname{diag}(\sigma_i^2)$$

Independencia y PDF condicional

Vimos que si $\mathbf{X}_1$ y $\mathbf{X}_2$ son dos VeA conjuntamente gaussianos, la PDF condicional es una gaussiana de media y matriz de covarianza

$$\mathbb{E}[\mathbf{X}_1|\mathbf{X}_2 = \boldsymbol{x}_2] = \mu_{\mathbf{X}_1} + C_{\mathbf{X}_1,\mathbf{X}_2} C_{\mathbf{X}_2}^{-1}(\boldsymbol{x}_2 - \mu_{\mathbf{X}_2}) \quad \mathbf{Cov}(X_1|X_2) = C_{\mathbf{X}_1} - C_{\mathbf{X}_1,\mathbf{X}_2} C_{\mathbf{X}_2}^{-1} C_{\mathbf{X}_2,\mathbf{X}_1}$$

- Si $\mathbf{X}_1$ y $\mathbf{X}_2$ están descorrelacionados, $C_{\mathbf{X}_1,\mathbf{X}_2} = 0$. Por ende,

  $$\mathbb{E}[\mathbf{X}_1|\mathbf{X}_2 = \boldsymbol{x}_2] = \mu_{\mathbf{X}_1} \quad \mathbf{Cov}(X_1|X_2) = C_{\mathbf{X}_1}$$

  y $f_{\mathbf{X}_1|\mathbf{X}_2} = f_{\mathbf{X}_1}$, es decir $\mathbf{X}_1$ y $\mathbf{X}_2$ son independientes.
- Por otro lado, si $\mathbf{X}_1$ y $\mathbf{X}_2$ son independientes $f_{\mathbf{X}_1|\mathbf{X}_2} = f_{\mathbf{X}_1}$. Para ello,

  $$\mu_{\mathbf{X}_1} + C_{\mathbf{X}_1,\mathbf{X}_2} C_{\mathbf{X}_2}^{-1}(\boldsymbol{x}_2 - \mu_{\mathbf{X}_2}) = \mu_{\mathbf{X}_1} \quad \text{y} \quad C_{\mathbf{X}_1} - C_{\mathbf{X}_1,\mathbf{X}_2} C_{\mathbf{X}_2}^{-1} C_{\mathbf{X}_2,\mathbf{X}_1} = C_{\mathbf{X}_1}$$

  La única posibilidad que esto se cumpla $\forall \boldsymbol{x}_2$ es que $C_{\mathbf{X}_1\mathbf{X}_2} = 0$.

Componentes principales

Conjuntos de nivel

- Una forma de visualizar y comprender mejor la forma de la PDF de un vector gaussiano es analizar los conjuntos de nivel. Dado un α constante, se define el conjunto de nivel

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

$$f_{\mathbf{X}}(\boldsymbol{x}) = \frac{1}{\sqrt{\det(2\pi C_{\mathbf{X}})}} e^{-(\boldsymbol{x}-\mu_{\mathbf{X}})^T C_{\mathbf{X}}^{-1}(\boldsymbol{x}-\mu_{\mathbf{X}})/2} = \alpha,$$

$$\Longrightarrow (\boldsymbol{x} - \mu_{\mathbf{X}})^T C_{\mathbf{X}}^{-1}(\boldsymbol{x} - \mu_{\mathbf{X}}) = \beta, \qquad \text{donde } \beta \text{ es otra constante.}$$

Esta ecuación describe a un elipsoide en $\mathbb{R}^n$.

Direcciones principales

Vimos que $C_{\mathbf{X}} = P\Lambda P^t$, con $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$ y $PP^t = \mathbf{I}_n$.
Reemplazando esta expresión en la ecuación de la elipsoide, tenemos

$$\beta = \underbrace{(\boldsymbol{x} - \mu_{\mathbf{X}})^t P}_{\boldsymbol{y}^t} \Lambda^{-1} \underbrace{P^t (\boldsymbol{x} - \mu_{\mathbf{X}})}_{\boldsymbol{y}}$$

Definimos un vector auxiliar $\boldsymbol{y} = P^t(\boldsymbol{x} - \mu_{\mathbf{X}})$. Luego recordando que $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$,

$$\beta = \boldsymbol{y}^t \Lambda^{-1} \boldsymbol{y} = \sum_{i=1}^{n} \frac{y_i^2}{\lambda_i}.$$

Los auto-vec de $C_{\mathbf{X}}$ son las *Direcciones principales* del elipsoide cuyos ejes tienen una longitud proporcional a los auto-val de $C_{\mathbf{X}}$.

Ejemplo: curvas de nivel en $\mathbb{R}^2$

- $\mathbf{X} = [X, Y] \sim \mathcal{N}(0, C_{\mathbf{X}})$.
- La matriz de covarianza es

$$C_{\mathbf{X}} = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix} \longrightarrow C_{\mathbf{X}}^{-1} = \frac{1}{\sigma_X^2\sigma_Y^2(1-\rho^2)} \begin{bmatrix} \sigma_Y^2 & -\rho\sigma_X\sigma_Y \\ -\rho\sigma_X\sigma_Y & \sigma_X^2 \end{bmatrix}.$$

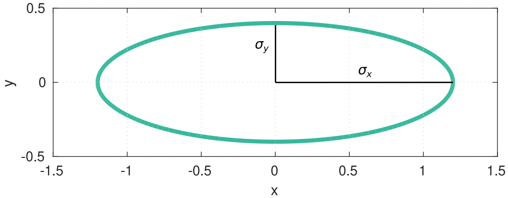
- Luego,

$$\boldsymbol{x}^T \boldsymbol{C}_{\mathbf{X}}{}^{-1} \boldsymbol{x} = \frac{1}{1-\rho^2}\left[\frac{x^2}{\sigma_X^2} - 2\rho\frac{x}{\sigma_X}\frac{y}{\sigma_Y} + \frac{y^2}{\sigma_Y^2}\right].$$

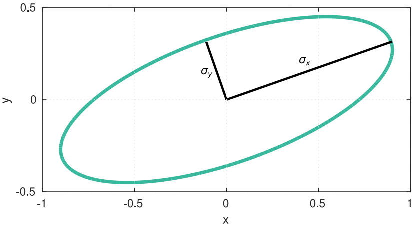
  Las curvas de nivel son $\{(x, y) \in \mathbb{R}^2 : \frac{x^2}{\sigma_X^2} - 2\rho\frac{x}{\sigma_X}\frac{y}{\sigma_Y} + \frac{y^2}{\sigma_Y^2} = c\}$

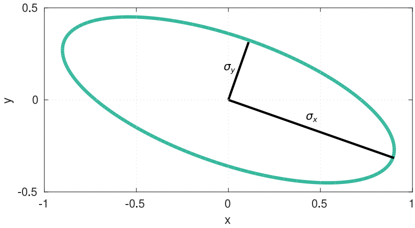
- Las longitudes de los semiejes principales son proporcionales a σ_X y σ_Y.

Ejemplo: curvas de nivel en $\mathbb{R}^2$



$\rho = 0 \longrightarrow X, Y$ descorrelacionadas





$\rho > 0$ $\rho < 0$

Análisis de componentes principales

- Los conjuntos de nivel de la pdf de $\mathbf{X} \sim \mathcal{N}(\mu_{\mathbf{X}}, C_{\mathbf{X}})$ son elipsoides cuyas direcciones son los auto-vec $\{p_1, \cdots p_n\}$. El largo de cada eje es proporcional al auto-val asociado $\{\lambda_1, \cdots, \lambda_n\}$. Supongamos que los auto-vec están ordenados de modo que $\lambda_1 \geq \cdots \geq \lambda_n$.
- Si a partir de un auto-val, los auto-val restantes son mucho más pequeños, es decir, $\lambda_1 \geq \cdots \geq \lambda_r \gg \lambda_{r+1} \geq \cdots \geq \lambda_n$, las realizaciones de $\mathbf{X}$ van a tener mayor variación en las direcciones $p_1, \cdots p_r$ que en las restantes direcciones de $\mathbb{R}^n$
- Una idea para *comprimir* la información contenida en el vector $\mathbf{X}$ es observar solamente las primeras r componentes.
- Es decir, vamos a aproximar $\mathbf{X}$ por un VeA en el subespacio generado por $\{p_1, \cdots p_r\}$.

Análisis de componentes principales

Sin falta de generalidad, vamos a considerar que $\mu_{\mathbf{X}} = 0$.
Las *componentes principales* de **X** son las proyecciones de **X** en cada una de sus direcciones principales. Es decir, la *i*-ésima componente principal tiene la forma

$$Y_i = p_i^t \mathbf{X}.$$

- Y_i es una V.A (pregunta: es gaussiana?)
- Para representar **X** como combinación lineal de $\{p_1, \cdots p_r\}$ utilizamos las primeras r componentes principales:

  $$\hat{\mathbf{X}}_r = \sum_{i=1}^{r} Y_i p_i = \sum_{i=1}^{r} p_i^t \mathbf{X} p_i.$$

  Obviamente, cuando $r = n$, $\hat{\mathbf{X}}_n = \mathbf{X}$.
- Este análisis se llama *Análisis de componentes principales* o PCA (principal component analysis).

Ejemplo: PCA en $\mathbb{R}^2$

Sea $\mathbf{X} = [X_1, X_2]^t$ de media nula y matriz de covarianza

$$\boldsymbol{C}_{\mathbf{X}} = \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix}.$$

Cuál es la media cuadrática del error de representación $\mathbb{E}[\|\mathbf{X} - \hat{\mathbf{X}}_r\|^2]$?

- En este caso, la diagonalización de $C_{\mathbf{X}}$ resulta en

$$\lambda_1 = \sigma^2(1+\rho),\ \lambda_2 = \sigma^2(1-\rho), \qquad p_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 2 \end{bmatrix},\ p_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Ejemplo: PCA en $\mathbb{R}^2$

- La primera componente principal de $\mathbf{X}$

$$Y_1 = p_1^t \mathbf{X} = \frac{X_1 + X_2}{\sqrt{2}}$$

y la mejor representación en un subespacio de dimensión 1 es

$$\hat{\mathbf{X}}_1 = Y_1 p_1 = \frac{X_1 + X_2}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

- La segunda componente principal de $\mathbf{X}$ es

$$Y_2 = p_2^t \boldsymbol{X} = \frac{X_1 - X_2}{\sqrt{2}}$$

Ejemplo: PCA en $\mathbb{R}^2$

- Claramente, si tomamos ambas componentes obtenemos

$$\hat{\mathbf{X}}_2 = Y_1 p_1 + Y_2 p_2 = \frac{X_1 + X_2}{2}\begin{bmatrix}1\\1\end{bmatrix} + \frac{X_1 - X_2}{2}\begin{bmatrix}1\\-1\end{bmatrix} = \begin{bmatrix}X_1\\X_2\end{bmatrix} = \mathbf{X}.$$

- Como $\mathbf{X} - \hat{\mathbf{X}}_1 = Y_2 p_2 = (p_2^t \boldsymbol{X})p_2$, el error de la representación es

$$\begin{aligned}\mathbb{E}\left[\|\mathbf{X} - \hat{\mathbf{X}}_1\|^2\right] &= \mathbb{E}\left[|p_2^t \mathbf{X}|^2 \|p_2\|^2\right] = \mathbb{E}\left[|p_2^t \mathbf{X}|^2\right] \qquad \textit{por ser } \|p_2\| = 1\\ &= \mathbb{E}\left[p_2^t \mathbf{X}\mathbf{X}^t p_2\right] = p_2^t C_{\mathbf{X}} p_2 = \lambda_2 = \sigma^2(1-\rho).\end{aligned}$$

- *Interpretación*: cuánto más fuerte sea la correlación entre las variables, menor será el error. En el caso degenerado las dos componentes se desplazan en un subespacio de dimensión 1. Como $\rho = 1$, $\mathbb{E}[\|\mathbf{X} - \hat{\mathbf{X}}_1\|^2] = 0$.

Reducción de la dimensionalidad en imágenes

Reconstrucciones de una imagen del dígito 4 de la base de datos MNIST usando (a) $r = 1$, (b) $r = 2$, (c) $r = 4$, (d) $r = 8$, (e) $r = 16$, (f) $r = 32$, (g) $r = 64$, (h) $r = 128$, (i) $r = 400$ componentes principales.

Ejercicios

Ejercicio: Vector gaussiano degenerado

Sea $Z \sim \mathcal{N}(0, 1)$. Definimos el VeA

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} Z \\ 2Z + 1 \end{bmatrix}.$$

Calcular la pdf conjunta $f_{\mathbf{X}}(\boldsymbol{x})$.

Solución

- De la definición de **X**, deducimos que $X_1 \sim \mathcal{N}(0,1)$.
- Igualmente, $X_2 = 2X_1 + 1$, y por ende X_2 es gaussiana por ser transformación afín de una gaussiana. Más aún, $X_2 \sim \mathcal{N}(1,4)$.
- Como $X_2 = 2X_1 + 1$, observamos que la realización x_1 determina sin ambigüedades la realización de X_2.
- Utilizando la ley de cadena obtenemos

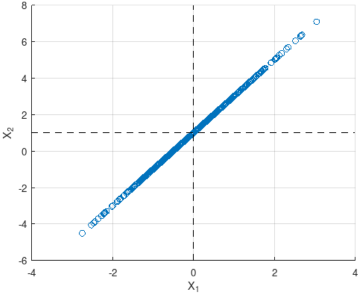
$$\begin{aligned} f_{\mathbf{X}}(\boldsymbol{x}) &= f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1) f_{X_2|X_1=x_1}(x_2|x_1) \\ &= \frac{1}{\sqrt{2\pi}} e^{-x_1^2/2} \delta\left(x_2 - (2x_1+1)\right). \end{aligned}$$

- No hay una única forma de escribir la PDF conjunta.

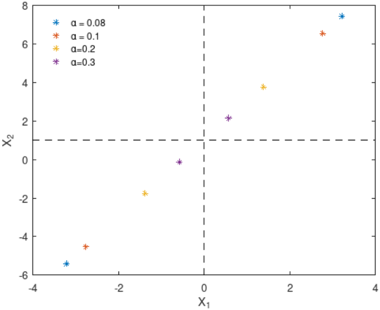
$$\begin{aligned} f_{\mathbf{X}}(\boldsymbol{x}) = f_{X_1,X_2}(x_1,x_2) &= f_{X_2}(x_2) f_{X_1|X_2=x_2}(x_1) \\ &= \frac{1}{\sqrt{8\pi}} e^{-(x_2-1)^2/2}\, \delta\left(x_1 - \frac{x_2-1}{2}\right). \end{aligned}$$

Solución

Realizaciones de **X**



Curvas de Nivel de $f_{\mathbf{X}}$



Ejercicio: V.As marginalmente pero no conjuntamente gaussianas

Sean $X \sim \mathcal{N}(0,1)$ y $B \sim \text{Ber}(1/2)$, independientes entre sí. Definimos

$$Y = (2B - 1)X.$$

Determinar si X e Y son conjuntamente gaussianas o no.

Solución

- Para que las V.A sean conjuntamente gaussianas, las dos deben ser gaussianas. X lo es. Tenemos que ver qué pasa con Y. Para ello, calculamos la cdf:

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) = \mathbb{P}(B=1)\mathbb{P}(X \leq y) + \mathbb{P}(B=0)\mathbb{P}(-X \leq y) \\ &= \frac{1}{2}\mathbb{P}(X \leq y) + \frac{1}{2}\mathbb{P}(X \geq -y) = \mathbb{P}(X \leq y). \end{aligned}$$

  Es decir, Y es gaussiana y tiene la misma distribución que X.
- La suma resulta $W = X + Y = 2BX$. Luego,

$$\begin{aligned} F_W(w) &= \mathbb{P}(W \leq w) = \mathbb{P}(B=1)\mathbb{P}(2X \leq w) + \mathbb{P}(B=0)\mathbb{P}(0 \leq w) \\ &= \frac{1}{2}\mathbb{P}(X \leq \frac{w}{2}) + \frac{1}{2}\mathbb{1}\{w \geq 0\} \longrightarrow W \text{ no es normal.} \end{aligned}$$

- Como hay una combinación lineal de X e Y que no es gaussiana, entonces X e Y no son conjuntamente gaussianas, pese a ser cada una marginalmente gaussianas.

Solución

Cómo es la distribución conjunta $f_{X,Y}(x, y)$?

- Como $Y = (2B - 1)X$, cuando $X = x$, $Y = x$ o $Y = -x$ dependiendo de si $B = 1$ o $B = 0$. Es decir que

$$f_{Y|X}(y|X = x) = \mathbb{P}[B = 1]\delta(y - x) + \mathbb{P}[B = 0]\delta(y + x)$$

- Utilizando la regla de la cadena

$$f_{X,Y}(x, y) = f_{Y|X}(y|X = x)f_X(x) = \frac{1}{2}\left[\delta(y - x) + \delta(y + x)\right] f_X(x)$$

Claramente, no es gaussiana.

Ejercicio:Gaussianas independendientes

Un transmisor de radio envía una señal $s > 0$ que es recibida en un receptor utilizando tres caminos. Las señales que llegan al receptor por cada camino son:

$$X_1 = s + N_1 \qquad X_2 = s + N_2 \qquad X_3 = s + N_3$$

donde N_1, N_2, N_3 son variables aleatorias gaussianas independientes con media cero y varianza unitaria.

1. Hallar la pdf conjunta de $\mathbf{X} = [X_1, X_2, X_3]^t$
2. Determine si las componentes de $\mathbf{X}$ son variables aleatorias independientes.
3. Hallar la probabilidad de que el mínimo de las tres señales sea positivo.
4. Hallar la probabilidad de que la mayoría de las señales sean positivas.

Ejercicio:Canal de comunicaciones

La señal transmitida por un canal de comunicaciones se suele modelar con un VeA $\mathbf{X} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$ de módulo A y módulo al cuadrado B

$$\mathbf{X} = \begin{bmatrix} U \\ V \end{bmatrix} \qquad A = \sqrt{U^2 + V^2} \qquad B = A^2.$$

Luego, $A \sim \text{Rayleigh}(\sigma)$ y $B \sim \text{Exp}(\frac{\sigma^2}{2})$.

$$f_A(a) = \frac{a}{\sigma^2} e^{-a^2/(2\sigma^2)} \quad , \quad f_B(b) = \frac{\sigma^2}{2} e^{-b\sigma^2/2} \qquad a, b \geq 0$$

Para combatir las perturbaciones del canal se agrega redundancia en el mensaje. En un modelo sencillo, se envían N versiones independientes de una misma señal en forma consecutiva, $\mathbf{X}_1, \cdots, \mathbf{X}_N$. Una señal es descartada si su amplitud al cuadrado B_k está por debajo de un umbral γ.

Hallar la probabilidad de que las N señales estén por debajo de ese umbral.

V.A conjuntamente gaussianas

Sean X e Y dos V.A conjuntamente gaussianas de media nula. Demuestre que

$$\mathbb{E}[X^2 Y^2] = \mathbb{E}[X^2]\mathbb{E}[Y^2] + 2\mathbb{E}^2[XY]$$

Ayuda: Utilice el resultado de gaussiana condicionada por gaussiana y la expresión de los momentos de una gaussiana.