

# **Estimación de máxima verosimilitud en modelos gráficos gaussianos**

---

Autor: Daniel Brito Sotelo

Tutor: Bojan Mihaljevic

Grado en Matemáticas, Facultad de Ciencias

Julio 2022

- **Introduction to our problem**

---
- **Recurrent neural network approach**

---
- **Binary classification of stock yearly return**

---
- **Conclusions**

---

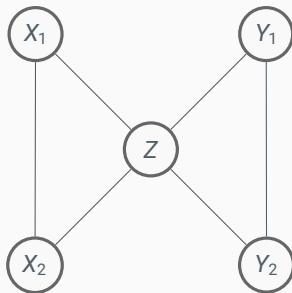
ip

## Definición: Independencia condicional

$$X \perp\!\!\!\perp Y|Z \iff f_{XYZ}(x, y, z) = f_{X|Z=z}(x)f_{Y|Z=z}(y)$$

## Definición: Grafo de independencia

Un grafo de independencia para una variable aleatoria  $X = (X_1, \dots, X_n)$  es un grafo no dirigido  $G = (V, E)$  en el que si  $(X_i, X_j) \notin E$  entonces  $X_i \perp\!\!\!\perp X_j | X \setminus \{X_i, X_j\}$ .



### Ejemplo

En este grafo tenemos:

- $X_1 \perp\!\!\!\perp Y_1 | (X_2, Y_2, Z)$
- $X_2 \perp\!\!\!\perp Y_2 | (X_1, Y_1, Z)$
- $X_1 \perp\!\!\!\perp Y_2 | (X_2, Y_1, Z)$

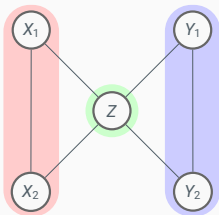
## Propiedades de Markov

Dado un grafo de independencia  $G$  tenemos las siguientes propiedades

- Por pares:  $X_i$  y  $X_j$  disjuntos entonces  $X_i \perp\!\!\!\perp X_j | X \setminus \{X_i, X_j\}$ .
- Local: Para  $X_i$ , siendo  $X_N$  sus vecinos, entonces  $X_i \perp\!\!\!\perp X \setminus X_N | X_N$ .
- Global: Si el grafo se puede separar en 3 conjuntos disjuntos de vértices  $X_A, X_B, X_C$  donde  $X_A$  y  $X_B$  están separados por  $X_C$ , entonces  $X_A \perp\!\!\!\perp X_B | X_C$ .

## Teorema

Dado un grafo de independencia  $G$  las propiedades de Markov son equivalentes.



## Ejemplo

Aplicando las propiedades de Markov

- Local:  $X_1 \perp\!\!\!\perp (Y_1, Y_2) | (X_2, Z)$
- Global:  $(X_1, X_2) \perp\!\!\!\perp (Y_1, Y_2) | Z$

## Definición: Distribución normal multivariante

$X \sim N_n(\mu, \Sigma)$  sí y sólo si la densidad de  $X$  es

$$f_X(x) = |\Sigma|^{-1/2} (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2}(x-\mu)^t \Sigma^{-1}(x-\mu)}$$

Distribución condicionada  $X_a|X_b \sim N_{|X_b|}(\mu_{a|b}, \Sigma_{a|b})$  con

- $\mu_{a|b} = \mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(x_b - \mu_b)$ .
- $\Sigma_{a|b} = \text{Var}(X_a|X_b) = \Sigma_{aa} + \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}$ .

Distribución marginal:  $X_a \sim N_{|X_a|}(\mu_a, \Sigma_{aa})$ .

## Definición

La matriz de covarianzas de un vector aleatorio  $X = (X_a, X_b)$  es

$$\Sigma = \text{Var}(X_a, X_b) = \begin{bmatrix} \text{Var}(X_a) & \text{Cov}(X_a, X_b) \\ \text{Cov}(X_b, X_a) & \text{Var}(X_b) \end{bmatrix} = \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix}.$$

## Definición

La matriz de precisión de  $X \sim N_n(\mu, \Sigma)$  es  $D = \Sigma^{-1}$ .

## Teorema

Dado un vector aleatorio  $X = (X_a, X_b)$  su matriz de precisión es

$$D = \begin{bmatrix} \text{Var}(X_a|X_b)^{-1} & -\text{Var}(X_a|X_b)^{-1}\text{Cov}(X_a, X_b)\text{Var}(X_b)^{-1} \\ -\text{Var}(X_b|X_a)^{-1}\text{Cov}(X_b, X_a)\text{Var}(X_a)^{-1} & \text{Var}(X_b|X_a)^{-1} \end{bmatrix}.$$

## Corolario

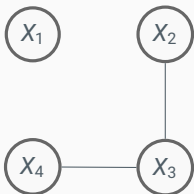
Los elementos de la matriz de precisión cumplen

$$d_{ij} = \text{Cov}(X_i, X_j | X \setminus \{X_i, X_j\}) \text{ para } i \neq j.$$

Por tanto si  $d_{ij} = 0$  tenemos que  $X_i \perp\!\!\!\perp X_j | X \setminus \{X_i, X_j\}$ .

## Definición

Dado un vector  $X \sim N_n(\mu, \Sigma)$  diremos que satisface un modelo gráfico gaussiano con grafo de independencia  $G = (V, E)$  si  $d_{ij} = 0$  cuando  $(i, j) \notin E$ .



## Ejemplo

Una variable  $X = (X_1, X_2, X_3, X_4)$  que satisface este modelo gráfico gaussiano tendrá una matriz de precisión de la forma

$$D = \begin{bmatrix} d_{11} & 0 & 0 & 0 \\ 0 & d_{22} & d_{23} & 0 \\ 0 & d_{32} & d_{33} & d_{34} \\ 0 & 0 & d_{43} & d_{44} \end{bmatrix}.$$

## Definición: logverosimilitud normal multivariante

Para  $X \sim N_n(\mu, \Sigma)$  su logverosimilitud  $l(\mu, D)$  cumple

$$l(\mu, D) \propto \frac{N}{2} \log |D| - \frac{N}{2} \text{tr}(DS) - \frac{N}{2} (\bar{x} - \mu)^t D (\bar{x} - \mu),$$

donde  $S$  es la matriz de covarianzas muestrales. Como los modelos gráficos no nos imponen restricciones sobre la media  $\mu$  podemos considerarla igual a la media muestral  $\bar{x}$  y reescribir

$$l(D) \propto \frac{N}{2} \log |D| - \frac{N}{2} \text{tr}(DS).$$

o equivalentemente para la matriz  $\Sigma$

$$l(\Sigma) \propto -\frac{N}{2} \log |\Sigma| - \frac{N}{2} \text{tr}(\Sigma^{-1}S).$$

Nos interesa, dado un modelo gráfico gaussiano con grafo de independencia  $G$  hallar la matriz de covarianzas que, cumpliendo las restricciones de independencias condicionadas del grafo, maximice la logverosimilitud.



Podemos enfocar la maximización de la verosimilitud como un problema de optimización convexa,

$$\begin{aligned} & \underset{D}{\text{maximizar}} \quad l(D) = \log |D| - \text{tr}(SD) \\ & \text{sujeto a } D \in \mathcal{D} = \{D : d_{ij} = 0 \text{ para } (i,j) \notin E\}. \end{aligned}$$

## Proposición

$l(D) \propto \log |D| - \text{tr}(DS)$  es cóncava para todo  $D \in \mathbb{S}_{\succ 0}^n$ .

## Dual

El problema dual asociado es

$$\begin{aligned} & \text{maximizar} \quad \log |\Sigma| + n \\ & \text{sujeto a } \sigma_{ij} = s_{ij} \quad \forall (i,j) \in E \text{ ó } i = j. \end{aligned}$$

Podemos aplicar el descenso de coordenadas sobre el problema dual

- Tomamos coordenada  $\sigma_{ij}$ .
- Maximizamos la verosimilitud en esa dirección.

---

**Algorithm 1** Descenso de coordenadas para la estimación de  $\Sigma$  por máxima verosimilitud

---

**Entrada:**  $S, G = (V, E)$  grafo de independencia, tolerancia

**Salida:** Estimación de máxima verosimilitud de  $\Sigma$

```
1:  $\tilde{\Sigma} \leftarrow S$ 
2: repeat
3:   for  $(i, j) \notin E$  do
       maximizar  $\log |\Sigma| + n$ 
        $\Sigma \in \mathbb{S}_{>0}^n$ 
       sujeto a  $\Sigma_{uv} = \tilde{\Sigma}_{uv} \forall (u, v) \neq (i, j)$ 
4:   end for
5:    $\tilde{\Sigma} \leftarrow \Sigma$ 
6: until  $\|\Sigma - \tilde{\Sigma}\|_1 < \text{tolerancia}$ 
```

---

# Solución cerrada para el paso de optimización iterativa

Queremos dar una solución cerrada para el paso de optimización

$$\begin{aligned} & \underset{\Sigma \in \mathbb{S}_{>0}^n}{\text{maximizar}} \quad \log |\Sigma| + n \\ & \text{sujeto a} \quad \Sigma_{uv} = \tilde{\Sigma}_{uv} \quad \forall (u, v) \neq (i, j) \end{aligned}$$

## Solución cerrada

Usando el complemento de Schur  $\Sigma \setminus R$  con  $R = V \setminus \{i, j\}$  obtenemos que

$$\sigma_{ij} = \Sigma_{iR} \Sigma_{RR}^{-1} \Sigma_{Rj},$$

maximiza la logverosimilitud en cada paso del algoritmo.

Podemos dar un algoritmo análogo que itere sobre la matriz de precisión.

---

**Algorithm 2** Descenso de coordenadas para la estimación de  $D$  por máxima verosimilitud

---

**Entrada:**  $S, G = (V, E)$  grafo de independencia, tolerancia

**Salida:** Estimación de máxima verosimilitud de  $D$

1:  $\tilde{D} \leftarrow I_n$

2: **repeat**

3:     **for**  $(i, j) \in E$  **do**

$$\underset{D \in \mathbb{S}_{\succ 0}^n}{\text{maximizar}} \quad \log |D| - \text{tr}(DS)$$

$$\text{sujeto a } D_{uv} = \tilde{D}_{uv} \quad \forall (u, v) \neq (i, j), (i, i), (j, j), (j, i).$$

4:     **end for**

5:      $\tilde{D} \leftarrow D$

6: **until**  $\|D - \tilde{D}\|_1 < \text{tolerancia}$

---

# Solución cerrada para el paso de optimización iterativa

La solución cerrada de la maximización

$$\underset{D \in \mathbb{S}_{>0}^n}{\text{maximizar}} \quad \log |D| - \text{tr}(DS)$$

$$\text{sueto a } D_{uv} = \tilde{D}_{uv} \quad \forall (u, v) \neq (i, j), (i, i), (j, j), (j, i).$$

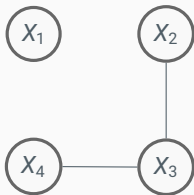
## Solución cerrada

Usando de nuevo el complemento de Schur obtenemos

$$D_{AA} = S_{AA}^{-1} + D_{AR} D_{BB}^{-1} D_{RA}$$

donde  $A = \{i, j\}$  y  $R = V \setminus \{i, j\}$ .

Podemos plantear el problema de la existencia del estimador de máxima verosimilitud como la completación positiva definida de la matriz parcial  $S_G$ , resultado de aplicar la proyección del grafo a la matriz  $S$ .



## Ejemplo

Dado el grafo y la siguiente matriz de covarianzas muestral  $S$  su proyección será la matriz parcial

$$S = \begin{bmatrix} 1 & 0 & 3 & 1 \\ 0 & 1 & 2 & 0 \\ 3 & 2 & 1 & 3 \\ 1 & 0 & 3 & 1 \end{bmatrix} \Rightarrow S_G = \begin{bmatrix} 1 & * & * & * \\ * & 1 & 2 & * \\ * & 2 & 1 & 3 \\ * & * & 3 & 1 \end{bmatrix}$$

## Definición

El umbral de observaciones del estimador de máxima verosimilitud dado un grafo  $G$ ,  $UMV(G)$  se define como el número mínimo de observaciones  $N$  tal que el estimador de máxima verosimilitud de  $\Sigma$  existe con probabilidad uno.

Una condición necesaria para la existencia del estimador es que todas las submatrices de  $S_G$  completamente definidas sean definidas positivas.

## Teorema

Si el número de observaciones  $N$  es mayor que el número de variables  $n$  entonces el estimador de máxima verosimilitud existe con probabilidad uno, por tanto

$$UMV(G) \leq n$$

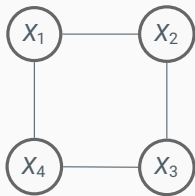
Nos interesa dar una cota inferior para el umbral de observaciones.

## Definición

Dado un grafo  $G$  un clique es un subgrafo completo de  $G$ , denotamos por  $\omega(G)$  el tamaño del mayor clique de  $G$ .

## Teorema

$$UMV(G) \geq \omega(G)$$



## Ejemplo

Veamos que  $UMV(G) \neq 1$ . Dada una muestra  $X_1 = (1, 2, 3, 4)$ , la proyección de la matriz de covarianzas muestral será

$$S_G = \begin{bmatrix} 1 & 2 & x_{13} & 4 \\ 2 & 4 & 6 & x_{24} \\ x_{13} & 6 & 9 & 12 \\ 4 & x_{24} & 12 & 16 \end{bmatrix}$$

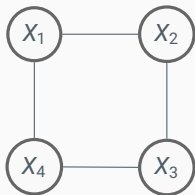


## Definición

Dado un grafo  $G$  un clique es un subgrafo completo de  $G$ , denotamos por  $\omega(G)$  el tamaño del mayor clique de  $G$ .

## Teorema

$$UMV(G) \geq \omega(G)$$



## Ejemplo

La submatriz en azul tiene determinante 0  $\implies$  no es definida positiva.

$$S_G = \begin{bmatrix} 1 & 2 & x_{13} & 4 \\ 2 & 4 & 6 & x_{24} \\ x_{13} & 6 & 9 & 12 \\ 4 & x_{24} & 12 & 16 \end{bmatrix}$$

# Cota superior Umbral de observaciones

## Definición

Un grafo  $G$  es cordal si cada camino de tamaño 4 o mayor tiene una triangulación.

## Teorema

Si el grafo  $G$  es cordal se tiene  $UMV(G) = \omega(G)$ .

## Definición

La cobertura cordal de un grafo  $G = (V, E)$  es el grafo cordal  $G^+ = (V, E^+)$  que cumple  $E \subset E^+$ . La cobertura cordal mínima de  $G$  es la cobertura cordal  $G^+$  con menor  $\omega(G^+)$ .

Usando estos resultados podemos dar una cota superior para el umbral de observaciones.

## Teorema

Sea  $G$  grafo y  $G^*$  cobertura cordal mínima  $UMV(G) \leq \omega(G^*)$ .

# Ejemplo de cotas Umbral de observaciones

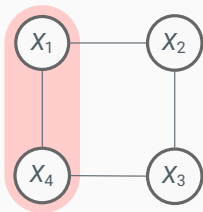


Figura: Grafo original  $G$

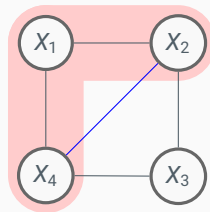


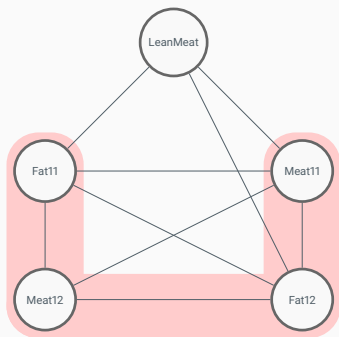
Figura: cobertura cordal mínima  $G^*$  de  $G$

## Ejemplo

Los cliques están marcados en rojo en ambos grafos, por lo que

$$2 = \omega(G) \leq \text{UMV}(G) \leq \omega(G^*) = 3$$

Tomamos el conjunto de datos *carcass* del paquete *gRbase*, con variables *Fat11*, *Meat11*, *Fat12*, *Meat12*, *LeanMeat*. Su grafo de independencia es



## Resultado

$G$  cordal y  $\omega(G) = 4 \Rightarrow \text{UMV}(G) = 4$ .

Usando la función *ggmfit* obtenemos la estimación de máxima verosimilitud de la matriz de precisión.

	Fat11	Meat11	Fat12	Meat12	LeanMeat
Fat11	0.368	0.020	-0.261	-0.045	0.090
Meat11	0.020	0.128	-0.065	-0.084	-0.046
Fat12	-0.261	-0.065	0.480	0.047	0.143
Meat12	-0.045	-0.084	0.047	0.079	0.000
LeanMeat	0.090	-0.046	0.143	0.000	0.251

1. WHITTAKER, J. *Graphical Models in Applied Multivariate Statistics*. Wiley, Chichester, 1990.
2. GRONE, R., JOHNSON, C., SÁ, E., WOLKOWICZ, H. *Positive definite completions of partial Hermitian matrices*. *Linear Algebra and Its Applications*, vol. 58, 109-124. 1984.
3. UHLER, C. *Geometry of maximum likelihood estimation in Gaussian graphical models*. *The Annals of Statistics*, vol. 40, no. 1, 2012.
4. BOYD, S., VANDENBERGHE, L. *Convex optimization*. Cambridge University Press, Cambridge, 2009.
5. HØJSGAARD, S., EDWARDS, D., LAURITZEN, S. *Graphical Models with R* Springer, 2012.