EST-46114 Métodos Multivariados y Datos Categóricos

Sesion 03 - Inferencia en la Distribucion Gaussiana Multivariada - Parte 1

Juan Carlos Martinez-Ovando

Maestria en Ciencia de Datos

Objetivos

- Revisaremos el procedimiento de estimacion puntual frecuentista y bayesiano de los parmetros de la distribucion gaussiana multidimensional.
- Revisaremos como hacer prediccion con ambos enfoques inferenciales.
- Implicaciones inferenciales pertinentes en ambos escenarios.

Clase de distribuciones

Como mencionamos, cuando nos referimos al modelo gaussiano (y a cualquier modelo de probabilidad) nos estamos refiriendo en realidad a clases modelos, $\mathcal{F}$, dadas por

$$\mathcal{F} = \{F(\cdot|\theta) : F(\cdot|\theta) \text{ es una medida de probabilidad para todo } \theta\},$$

con $\theta \in \Theta$ espacio parametral y *soporte* $\mathcal{X}$.
Usualmente se supone el cumplimiento de **condiciones de regularidad.**
Observacion

- Cada valor de θ define un modelo de probabilidad $F(\cdot|\theta)$.

Clase gaussiana de distribuciones

Para $X = (X_1, \ldots, X_p)$ vector p-dimensional, suponemos que es *modelado por la clase de distribuciones gaussianas* $\mathcal{F}$ si su funcion de densidad es,

$$f(x|\boldsymbol{\mu}, \boldsymbol{\lambda}) = (2\pi)^{-p/2}|\boldsymbol{\Lambda}|^{1/2} \exp\left\{-\frac{1}{2}(x-\boldsymbol{\mu})'\Lambda(x-\boldsymbol{\mu})\right\},$$

para todo x en $\mathcal{X}$, siendo $\boldsymbol{\Lambda}$ la matrix de precision.

Datos

Consideremos ahora que contamos con n datos recolectados,

$$\text{datos} = \{x_1, \ldots, x_p\},$$

con x_js vectores p-dimensionales. Tipicamente renglones de una matriz de datos, e.g.

```
data(swiss)
swiss[1,]
```

```
##            Fertility Agriculture Examination Education Catholic
## Courtelary      80.2          17          15        12     9.96
##            Infant.Mortality
## Courtelary             22.2
```

Inferencialmente buscaremos el modelo particular en $\mathcal{F}$ que *sea mas compatible* con los datos.

Dos paradigmas

A. **Frecuentista**
B. **Bayesiano**
Ambos basados en el `principio de verosimilitud`, i.e. la compatibilidad de los datos y los modelos en $\mathcal{F}$ se mide con la funcion

$$l(\boldsymbol{\mu}, \boldsymbol{\Lambda}|\text{datos}) = f(x_1, \ldots, x_n|\boldsymbol{\mu}, \boldsymbol{\Lambda}).$$

Es decir,

$$l(\boldsymbol{\mu}, \boldsymbol{\Lambda}|\text{datos}) = \begin{cases} \prod_{i=1}^{n} f(x_i; \boldsymbol{\mu}, \boldsymbol{\Lambda}) & , \text{(frecuentista)}, \\ \prod_{i=1}^{n} f(x_i|\boldsymbol{\mu}, \boldsymbol{\Lambda}) & , \text{(bayesiano)}. \end{cases}$$

Estimacion frecuentista

El *modelo mas compatible*, bajo el enfoque frecuentista, es

$$F(\cdot|\boldsymbol{\mu}^*, \boldsymbol{\Lambda}^*),$$

tal que

$$\boldsymbol{\mu}^*, \boldsymbol{\Lambda}^* = \arg\max_{\boldsymbol{\mu}, \boldsymbol{\Lambda}} l(\boldsymbol{\mu}, \boldsymbol{\Lambda}|\text{datos}).$$

Es decir,

$$\begin{aligned} \boldsymbol{\mu}^* &= \bar{x}, \\ \boldsymbol{\Lambda}^* &= S^{-1}, \end{aligned}$$

con $S = \sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})'$.
Se obtiene de manera analoga al caso univariado (revisen las lecturas complementarias).

Estimacion bayesiana I

En el caso bayesiano, se considera que junto con los `datos` se dispone de fuentes complementarias de informacion reflejadas en la `distribucion prior` con soporte en la clase de modelos $\mathcal{F}$ (a.k.a. espacio parametral Θ), con densidad,

$$q(\boldsymbol{\mu}, \boldsymbol{\Lambda} | \text{informacion complementaria}).$$

- Como la verosimilitud, la funcion $q(\cdot)$ sirve para cuantificar **grados de acertividad** que *creamos convenientes.*
- En un momento veremos formas de especificar $q(\cdot)$.

Estimacion bayesiana II

La informacion consolidad de los `datos` y `complementaria` se **consolida** en el producto

$$l(\boldsymbol{\mu}, \boldsymbol{\Lambda}|\text{datos}) \times q(\boldsymbol{\mu}, \boldsymbol{\Lambda}|\text{complementaria}).$$

Estimacion bayesiana II

Generalmente, resulta que este producto es una medida de probabilidad para $(\boldsymbol{\mu}, \boldsymbol{\Lambda})$, con densidad dada por

$$q(\boldsymbol{\mu}, \boldsymbol{\Lambda} | \text{complementaria}) \quad \propto \quad l(\boldsymbol{\mu}, \boldsymbol{\Lambda} | \text{datos}) q(\boldsymbol{\mu}, \boldsymbol{\Lambda} | \text{complementaria}),$$

siendo constante de normalizacion,

$$\int \int l(\boldsymbol{\mu}, \boldsymbol{\Lambda} | \text{datos}) q(\boldsymbol{\mu}, \boldsymbol{\Lambda} | \text{complementaria}) \mathrm{d}\boldsymbol{\mu} \mathrm{d}\boldsymbol{\Lambda}.$$

Priors

Difusa (no informativa)

$$q(\boldsymbol{\mu}, \boldsymbol{\Lambda}) \propto |\boldsymbol{\Lambda}|^{1/2}.$$

Conjugada (parcialmente informativa)

$$q(\boldsymbol{\mu}, \boldsymbol{\Lambda}) = N(\boldsymbol{\mu}|\boldsymbol{m}_0, s_0\boldsymbol{\Lambda})W(\boldsymbol{\Lambda}|a_0, \boldsymbol{B}_0),$$

donde s_0 y a_0 son escalares positivos, $\boldsymbol{m}_0$ es un vector p-dimensional y $\boldsymbol{B}_0$ es una matriz simetrica $p \times p$-dimensional.

- $N(\boldsymbol{\mu}|\boldsymbol{m}_0, s_0\boldsymbol{\Lambda})$ - Densidad gaussiana p-dimensional
- $W(\boldsymbol{\Lambda}|a_0, \boldsymbol{B}_0)$ - Densidad Wishart p-dimensional (vean lecturas complementarias)

Los valores $(\boldsymbol{m}_0, s_0, a_0, \boldsymbol{B}_0)$ *se conocen como* **hiperparametros** *y reflejan (de alguna forma) la* `informacion complementaria` *a los datos.*

Estimacion bayesiana III

Como mencionamos, el aprendizaje bayesiano con `datos` y `complemento` se consolida con $q(\boldsymbol{\mu}, \boldsymbol{\Lambda}|\text{datos y complemento})$.

Caso conjugado:

$$q(\boldsymbol{\mu}, \boldsymbol{\Lambda}|\text{datos y complemento}) = N(\boldsymbol{\mu}|\boldsymbol{m}_n, s_n\boldsymbol{\Lambda})W(\boldsymbol{\Lambda}|a_n, \boldsymbol{B}_n),$$

con,

$$\begin{aligned} \boldsymbol{m}_n &= \frac{s_0\boldsymbol{m}_0 + n\bar{x}}{s_n}, \\ s_n &= s_0 + n, \\ a_n &= a_0 + \frac{n}{2}, \\ \boldsymbol{B}_n &= \boldsymbol{B}_0 + \frac{n}{2}\left(S + \frac{s_0}{s_n}(\bar{x} - \boldsymbol{m}_0)(\bar{x} - \boldsymbol{m}_0)'\right), \end{aligned}$$

donde,

$$S = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})'.$$

Estimacion bayesiana IV

Para μ

$$\boldsymbol{\mu}^{**} = \boldsymbol{m_n}.$$

Para $\boldsymbol{\Lambda}$

$$\boldsymbol{\Lambda}^{**} = \boldsymbol{B_n}^{-1}.$$

- *Salvo que la prior sea suficientemente distinta de la verosimilitud, $\boldsymbol{\mu}^{*}$ y $\boldsymbol{\mu}^{**}$ no deberan diferir significativamente.*
- *Lo mismo aplica a $\boldsymbol{\Lambda}$.*

Estimacion bayesiana IV

Para μ

$$\boldsymbol{\mu}^{**} = \boldsymbol{m_n}.$$

Para $\boldsymbol{\Lambda}$

$$\boldsymbol{\Lambda}^{**} = \boldsymbol{B_n}^{-1}.$$

- *Salvo que la prior sea suficientemente distinta de la verosimilitud, $\boldsymbol{\mu}^{*}$ y $\boldsymbol{\mu}^{**}$ no deberan diferir significativamente.*
- *Lo mismo aplica a $\boldsymbol{\Lambda}$.*

Frecuentista vs bayesiano I

La distincion del paradigma bayesiano reside en:

a. Capacidad de incorporar `informacion complementaria`
b. Cuantificar probabilisticamente diferentes modelos probabilisticos consolidando informacion de `datos` y `complemento`

Frecuentista vs bayesiano I

Usando `gaussian.posterior`

```
# Frecuentista
mstar <- colMeans(data)
Lstar <- solve(cov(swiss))
mstar
```

```
##        Fertility      Agriculture      Examination        Education
##         70.14255         50.65957         16.48936         10.97872
##         Catholic Infant.Mortality
##         41.14383         19.94255
```

```
# Bayesiano
m2star <- t(output[[1]])
L2star <- solve(output[[4]])
m2star
```

```
##      Fertility Agriculture Examination Education Catholic Infant.Mor
## [1,]  68.68125    49.60417    16.14583     10.75 40.28667           19
```

Frecuentista vs bayesiano II

Los estimadores de las precisiones si pueden diferir (esta es una distincion muy importante).

```
Lstar[1:4,1:4]
```

```
##             Fertility Agriculture  Examination   Education
## Fertility   0.021852286 0.003761084  0.005638070  0.019032031
## Agriculture 0.003761084 0.005075664  0.004886445  0.007566829
## Examination 0.005638070 0.004886445  0.059201951 -0.019558373
## Education   0.019032031 0.007566829 -0.019558373  0.046589397
```

```
L2star[1:4,1:4]
```

```
##                Fertility   Agriculture   Examination     Education
## Fertility    5.600515e-04 -1.257432e-05 -1.256250e-04  0.0004243553
## Agriculture -1.257432e-05  1.350104e-04  3.640363e-05  0.0001418069
## Examination -1.256250e-04  3.640363e-05  2.148985e-03 -0.0012095659
## Education    4.243553e-04  1.418069e-04 -1.209566e-03  0.0015777335
```

Implicaciones inferenciales I

Error de estimacion

Frecuentista

`Regiones (intervalos) de confianza` - **Clases de equivalencia**

Bayesiano

`Regiones de credibilidad` - **No clases de equivalencia**

[Incorporen graficas de verosimilitud y posterior marginales.]

Ejercicios

1. Revisen los calculos de la distribucion posterior de $(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ empleando $q(\cdot)$ `difusa`.
2. Consideren los datos `data(swiss)` de provincias en Suiza. La descripcion de los datos esta en esta liga.
3. Estudien la implementacion en `R` de las distribuciones Wishart, normal y Student multivariadas. Vean las siguientes ligas:

Liga Wishart
Liga normal y t

En la siguiente sesion

Revisaremos como hacer:

1. Revisaremos las implicaciones de modelacion para responder problemas inferenciales espcificos sobre $\boldsymbol{\mu}$ y sobre $\boldsymbol{\Lambda}$.

Lecturas complementarias

- Martinez-Ovando (2016) “Paradigma Bayesiano de Inferencia (Resumen de Teoria)”, *Notas de clase EST-46114*. (De momento, no presten antencion a la descripcion de los metodos de simulacion transdimensionales.) `est46114_s03_suplemento2.pdf`
- Murphy (2007) “Conjugate Bayesian analysis of the Gaussian distribution”. Mimeo. `est46114_s03_suplemento2.pdf`
- Press (2005). “Applied multivariate analysis, using Bayesian and frequentist methods of inference.” Dover Pub. (Capitulo 4.)

Table of Contents