EST-46114 Métodos Multivariados y Datos Categóricos

Sesion 02 - Distribucion Gaussiana Multivariada

Juan Carlos Martinez-Ovando

Maestria en Ciencia de Datos

Objetivos

- Entender el uso de los modelos gaussianos en el analisis de datos multivariados.
- Estudiar propiedades teorico/practicas de la distribucion gaussiana multidimensional.
- Revisar procedimientos de inferencia y prediccion empleando estos modelos.

Usos

- La distribucion gaussiana *brinda* la posisilidad de estudiar **covariabilidad**, **simetria estocastica** y **simultaneidad** en fenomenos multidimensionales.
- Provee un **marco parcimonioso** de modelacion.
- Brinda una **plataforma de extension** de modelos sofisticados.

Limitaciones

- Solamente se atiende a los **primeros dos momentos** de la variabilidad entre dimensiones multiples.

Caso Univariado

Iniciamos con la distribucion gaussiana unidimensional, acompanada por una funcion de densidad de probabilidad,

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

con soporte en $\mathcal{X} = \mathbb{R}$.

- μ es parametro de localizacion.
- λ es parametro de dispersion (**precision**, en este caso).
- El espacio parametral para (μ,λ) es

$$\mathbb{R} \times \mathbb{R}_+.$$

Caso mulvivariado

El caso multivariado es una extension del univariado para $\boldsymbol{x} = (x_1, \ldots, x_p)$, pero que toma en cuenta las **interacciones** entre las $(x_i)_{i=1}^{p}$. Esta tiene una funcion de densidad conjunta dada por,

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

con soporte en $\mathbb{R}^p$ para $\boldsymbol{x}$.

- Espacio parametral para $\boldsymbol{\mu}$ es $\mathbb{R}^p$
- Espacio parametral para $\boldsymbol{\Lambda}$ es $\mathcal{M}_p$ –espacio de todas las matrices de dimension $p \times p$ positivo definidas y simetricas–.
- Numero de **parametros independientes** en $\boldsymbol{\mu}$ es igual a p, mientras que el numero de parametros independientes en $\boldsymbol{\Lambda}$ es igual a $p(p+1)/2$.

Propiedad P1

Si $\boldsymbol{X} \sim N(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Lambda})$ entonces,
P1. Cualquier combinacion lineal de $\boldsymbol{X}$ es gaussiana multivariada, i.e. para todo A matriz de dimension $q \times p$ y c vector de dimension $q \times 1$, tales que

$$\boldsymbol{\mu}_y = A\boldsymbol{\mu} + c,$$

y

$$\boldsymbol{\Lambda}_y = A\boldsymbol{\Lambda}A',$$

se sigue que

$$\boldsymbol{Y} \sim N(\boldsymbol{y}|\boldsymbol{\mu}_y, \boldsymbol{\Lambda}_y),$$

con soporte para $\boldsymbol{y}$ dado por $\mathbb{R}^q$.

Propiedad P1

Comentarios

- Esta propiedad es util ya que muchos problemas (contraste de hipotesis, agregacion, construccion de portafolios, proyecciones, etc.) pueden expresarse como una **transformacion afin (trasnformacion lineal anterior)**.

Propiedad P2

Si $\boldsymbol{X} \sim N(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Lambda})$ entonces,
P2. Cualquier sub-conjunto de $\boldsymbol{X}$, formado por subselecciones este vector, tiene una distribucion gaussiana multivariada asociada, i.e. si

$$\boldsymbol{X}_s = (X_{s(1)}, \ldots, X_{s(q)}),$$

es un subconjunto de $\boldsymbol{X}$, con $q \leq p$, para un subconjunto de indices

$$(s(1), \ldots, s(q))$$

de

$$(1, \ldots, p).$$

Propiedad P2

Se sigue que

$$\boldsymbol{X}_s \sim N(\boldsymbol{x}_s | \boldsymbol{\mu}_s, \boldsymbol{\Lambda}_s),$$

donde

$$\boldsymbol{\mu}_s = (\mu_{s(1)}, \ldots, \mu_{s(q)}),$$

y

$$\boldsymbol{\Lambda}_s = \left(\lambda_{s(i)s(j)}\right)_{i,j=1}^{q}.$$

Comentarios

- Esta propiedad permite relaizar estudios conjuntos o marginales reducidos a partir del mismo procedimiento, mediante la implementacion del proceso de marginalizacion involucrado.

Propiedad P3

Si $\boldsymbol{X} \sim N(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Lambda})$ entonces,
P3. Si un subconjunto de $\boldsymbol{X}$ no esta correlacionado, entonces las variables de ese conjunto son *estocasticamente independientes*, i.e. si

$$(X_i, X_j),$$

son dos dimensiones de $\boldsymbol{X}$ (con $i, j < p$), tales que

$$\lambda_{i,j} = 0,$$

se sigue que

$$X_i \text{ y } X_j,$$

son independientes estocasticamente con,

$$f(x_i, x_j|\mu_i, \mu_j, \lambda_{ii}, \lambda_{jj}, \lambda_{ij}) = N(x_i|\mu_i, \lambda_{ii}) \times N(x_j|\mu_j, \lambda_{jj}).$$

Propiedad P3

Comentarios

- A partir de esta propiedad podemos **estudiar relaciones de dependencia lineal** entre dimensiones de un conjunto de datos o problemas multidimensionales.
- Consideremos que las dependencias lineales se refieren *simultaneidad* mas no a *causalidad*.

Propiedad P4

Si $\boldsymbol{X} \sim N(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Lambda})$ entonces,
P4. Si un A y B son dos matrices de dimension $(q \times p)$ tales que

$$A\boldsymbol{\Sigma}B' = \mathbf{0},$$

de dimension $q \times q$, entonces

Propiedad P4

Se sigue que las variables,

$$\boldsymbol{Y}_A = A\boldsymbol{X} \;\text{ y }\; \boldsymbol{Y}_B = B\boldsymbol{X},$$

son **estocasticamente independientes**.
Comentarios

- A partir de esta propiedad, podemos entender que las estructuras de simultaneaidad entre variables pueden modificarse a traves de operaciones algebraicas teoricas.
- Esta propiedad es crucial, porque podemos **modificar patrones subyacentes de dependencia/simetria estocastica** aplicando transformaciones lineales.

Propiedad P5

P5. (Distribuciones condicionales). Si

$$\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2),$$

donde ambos componentes son de dimension p_1 y p_2 respectivamente, (lo anterior se conoce como una particion de dos componentes de $\boldsymbol{X}$), con $p = p_1 + p_2$, se sigue:

- Los parametros del modelo tambien estan particionados de la siguiente forma,

$$\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2),$$

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_{11} & \boldsymbol{\Lambda}_{12} \\ \boldsymbol{\Lambda}_{21} & \boldsymbol{\Lambda}_{22} \end{pmatrix}.$$

Propiedad P5

- La *distribucion marginal* de $\boldsymbol{X}_i$ (para $i = 1, 2$) es

$$\boldsymbol{X}_i \sim N(\boldsymbol{x}_i | \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i).$$

Propiedad P6

Si $\boldsymbol{X}$ es particionado en $\boldsymbol{X}_1$ y $\boldsymbol{X}_2$, se sigue:
P6. La distribucion condicional de $\boldsymbol{X}_2$ dado $\boldsymbol{X}_1 = \boldsymbol{x}_1$ es gaussiana multivariada, con

$$\boldsymbol{X}_2 | \boldsymbol{X}_1 = \boldsymbol{x}_1 \sim N\left(\boldsymbol{x}_2 | \boldsymbol{\mu}_{2|1}, \boldsymbol{\Lambda}_{2|1}\right),$$

donde

$$\boldsymbol{\mu}_{2|1} = \boldsymbol{\mu}_2 + \boldsymbol{\Lambda}_{21}\boldsymbol{\Lambda}_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1),$$

y

$$\boldsymbol{\Lambda}_{2|1} = \boldsymbol{\Lambda}_{22} - \boldsymbol{\Lambda}_{21}\boldsymbol{\Lambda}_{11}^{-1}\boldsymbol{\Lambda}_{12}.$$

Propiedad P6

Comentarios

- Esta propiedad es de **simetria estocastica**, en el sentido que la distribucion condicional de $\boldsymbol{X}_1 | \boldsymbol{X}_2 = \boldsymbol{x}_2$ es analoga.
- Esta propiedad es fundamental para entender el modelo de regresion multiple.

Descomposicion I

Estructura espectral.- Una matrix P de dimension $p \times p$ es **ortogonal** si

$$PP' = P'P = I.$$

P7. Una matrix A simetrica de dimension $p \times p$ puede reexpresarse como

$$A = P\Lambda P',$$

donde

- $\Lambda = diag\{\lambda_1, \ldots, \lambda_p\}$ de *eigenvalores*
- P es una matriz ortogonal de eigenvectores columna.

Descomposicion I

Intuicion

- La matriz de *eigenvalores* representa *rotaciones* de coordenadas.
- La matriz de *eigenvalores* representan *contracciones* o *expansiones* de coordenadas.

Descomposicion II

Acerca de determinantes.- Bajo la descomposicion espectral de $\mathbf{\Lambda}^-1$ se sigue

$$\det(\mathbf{\Lambda}^{-1}) = \prod_{j=1}^{p} \phi_j.$$

Esto es analogo a calcular los determinantes de $\mathbf{\Lambda}$, que estan determinados por

$$\det(\mathbf{\Lambda}) = \prod_{j=1}^{p} \phi_j^{-1},$$

por ser $\mathbf{\Lambda}$ una matriz positivo definida y simetrica.

Descomposicion II

Intiucion

- Los **eigenvalores** de $\mathbf{\Lambda}^{-1}$ nos indican el grado de variabilidad de cada dimension (por separado) en $\boldsymbol{X}$
- Los valores ϕ_j nos indican que tan expandida/contraida es la distribucion en esta dimension
- El factor $\det(\mathbf{\Lambda}^{-1})$ puede verse como un factor de expansion/contraccion de la distribucion gaussiana conjunta.

Visualizacion

La **grafica de contornos** nos permiten visualizar distribuciones gaussianas de dos dimensiones (i.e. es una representacion de graficas en 3D).
Los *contornos* se obtienen como *rebanadas* de la densidad conjunta en 2D, y representan **areas cuantilicas** del soporte $\mathcal{X}$ para un nivel $0 < c < 1$.
Las lineas representan la coleccion de valores de $\boldsymbol{x}$ que comparten el mismo nivel de $N(\boldsymbol{x}|\mu, \Lambda)$.
Calculo.- Los contornos se definen en terminos de Λ y P como **elipsoides**

$$||\boldsymbol{\Lambda}^{1/2}(\boldsymbol{x} - \boldsymbol{\mu})|| = c^2,$$

siendo centradas en

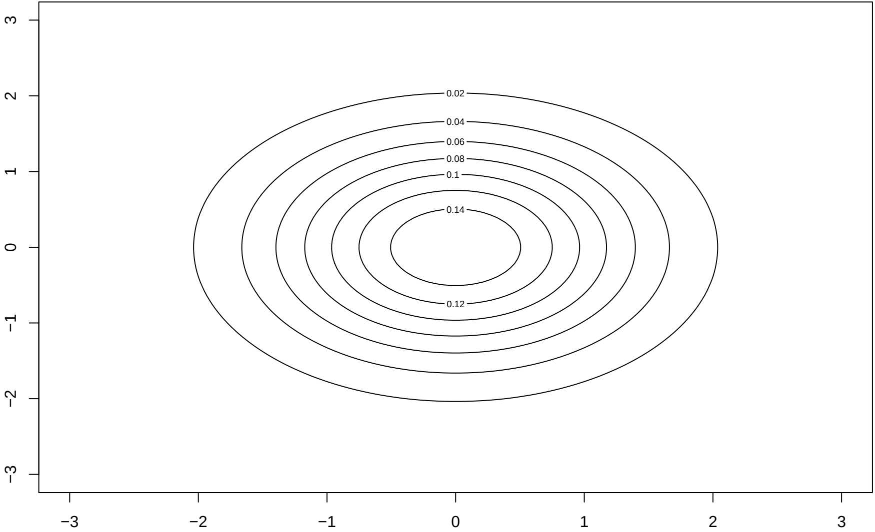
$$\boldsymbol{\mu},$$

y con ejes en la direccion

$$\sqrt{\lambda_j p_j}.$$

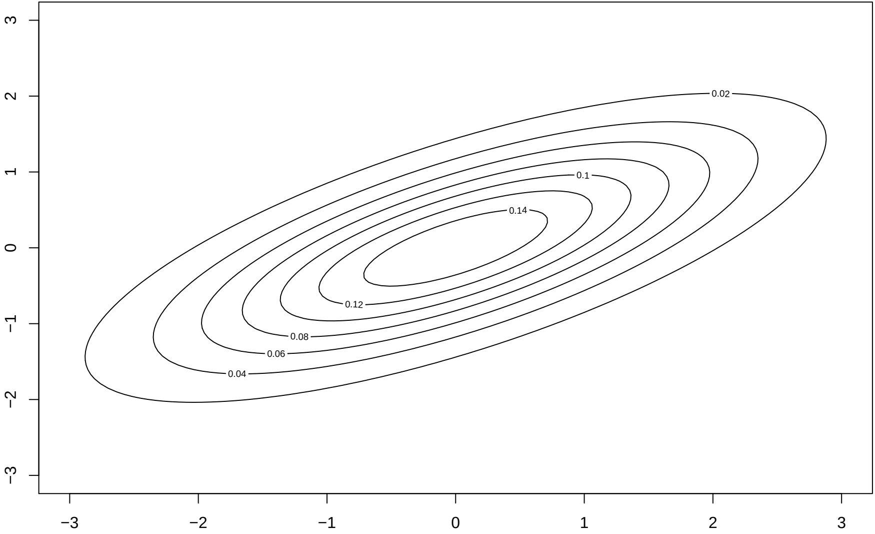
Visualizacion I

Caso: Independencia estocastica/simetria



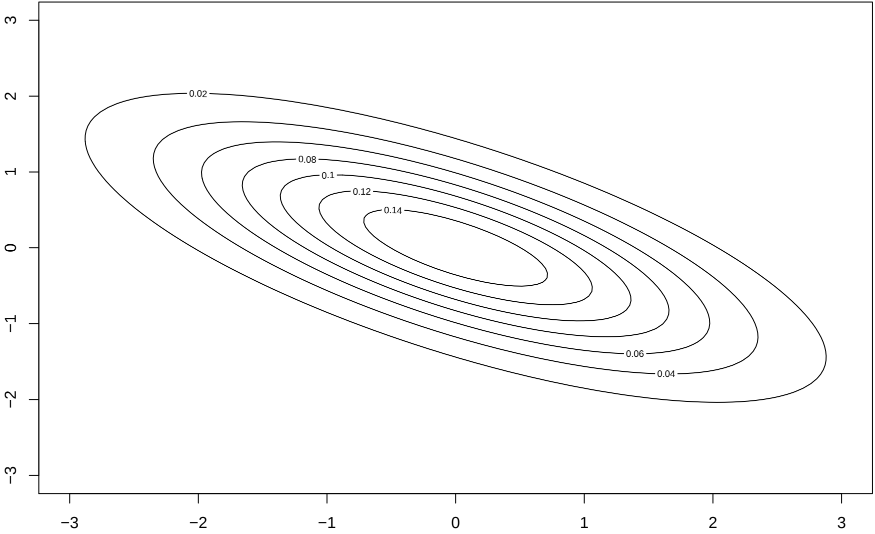
Visualizacion II

Caso: Dependencia positiva.



Visualizacion III

Caso: Dependencia negativa.



Verosimilitud

La funcion de verosimilitud para (μ, Λ) condicional en un conjunto de datos, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ se calcula como la densidad conjunta de los n datos, **bajo el supuesto de independencia en las observaciones**, de la forma

$$\begin{aligned} l(\mu, \Lambda | \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) &= \prod_i N(\boldsymbol{x}_i | \mu, \Lambda) \\ &\propto |\Lambda|^{n/2} \exp\left\{ -\frac{1}{2} \sum_i (x_i - \mu)' \Lambda (x_i - \mu) \right\} \\ &\propto |\Lambda|^{n/2} \exp\left\{ -\frac{1}{2} tr \left(\Lambda^{-1} (S + dd') \right) \right\}, \end{aligned}$$

donde $d = \bar{x} - \mu$ y S es la matriz de covarianzas muestral.

Ejercicio

La ultima linea de la ecuacion anterior es un ejercicio a casa. Consideren la complesion del cuadrado que esta en el documento suplementario de esta sesion (adjunto, escrito a mano).

Estimacion I

Caso: Estimacion frecuentista.
A partir de las ecuaciones anteriores, podemos encontrar el **estimador maximo verosilil** de μ y Λ, dados por

$$\hat{\mu} = \bar{x},$$

y

$$\hat{\Lambda} = \left(\sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})' \right)^{-1}.$$

El ultimo estimador corresponde a la matriz inversa de covarianzas muestral.

Comentarios

La distribucion gaussiana multidimensional es una buena herramienta para:

- Describir la aleatoriedad/desconocimiento en fenomenos multidimensionales.
- Su buen uso descansa en el supuesto de **simetria eliptica**.
- En la practica, la **simetria eliptica** puede corroborarse contrastando *graficas de dispersion de datos* con las curvas de nivel de **un modelo gaussiano multidimensional.**
- A partir de esto, muchos procesos inferenciales (incluyendo prediccion, descomposiciones expectrales, contraste de hipotesis) pueden plantearse con esto.

Ejercicios

1. Consideren los datos `data(swiss)` de provincias en Suiza.
2. Calculen el EMV de μ y Λ en la distribucion gaussiana.
3. Identifiquen la estructura de *simetria eliptica* y grafiquen tales relaciones en parejas de variables.

La siguiente sesion

Revisaremos como hacer:

1. Inferencia frecuentista (en parametros y transformaciones).
2. Inferencia bayesiana (en parametros y trasnformaciones).
3. Prediccion.
4. Modificaciones de distribuciones.

Lecturas complementarias

- Complecion de cuadrados en la distribucion gaussiana.
  est46114_s02_suplemento1.pdf

Lecturas extendidas

- Dawid, P.A. (1981) "Matrix-Variate Distribution Theory: Notational Considerations and a Bayesian Application", *Biometrika*. `est46114_s02_suplemento2.pdf`

Table of Contents