Revisión de Conceptos y Notación Básica

1EST17 - Aprendizaje Estadístico I

Enver G. Tarazona

2022-09-10

Contenidos

- Vectores Aleatorios
- Matríz de covarianzas y correlaciones
- Distribución Normal Multivariada

Vector aleatorio

- Un vector aleatorio $X_{(p \times 1)}$ es un vector p-dimensional de variables aleatorias. Por ejemplo
  - Peso de depósitos de corcho en $p = 4$ direcciones (N, E, S, O).
  - Índice de alquileres en Munich: alquiler, superficie, año de construcción, ubicación, estado del baño, estado de la cocina, calefacción central, distrito.
- Función de distribución conjunta: $f(x)$.
- De la función de distribución conjunta a distribuciones marginales (y condicionales).

$$f_1(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_p) dx_2 \cdots dx_p$$

- La distribución acumulada (integrales definidas) se utiliza para calcular probabilidades.
- Independencia: $f(x_1, x_2) = f_1(x_1) \cdot f(x_2)$ y $f(x_1 \mid x_2) = f_1(x_1)$.

Momentos

Los momentos son propiedades importantes sobre la distribución de X. Revisaremos:

- E: Media de vectores aleatorios y matrices aleatorias.
- Cov: matriz de covarianza.
- Corr: Matriz de correlación.
- E y Cov de múltiples combinaciones lineales.

Los datos del depósito de corcho

- Conjunto de datos multivariados clásico de Rao (1948).
- Peso de los depósitos de corcho (centigramos) de $n = 28$ alcornoques en las $p = 4$ direcciones cardinales (N, E, S, O).

```
corkds=as.matrix(
  read.table("corkMKB.txt")
  )
dimnames(corkds)[[2]]=c("N","E","S","W")
head(corkds)
```

```
##       N  E  S  W
## [1,] 72 66 76 77
## [2,] 60 53 66 63
## [3,] 56 57 64 58
## [4,] 41 29 36 38
## [5,] 32 32 35 36
## [6,] 30 35 34 26
```

```
dim(corkds)
```

```
## [1] 28  4
```

- Aquí tenemos una muestra aleatoria de $n = 28$ alcornoques de la población y observamos un vector aleatorio dimensional $p = 4$ para cada árbol.
- Esto nos lleva a la definición de vectores aleatorios y una matriz aleatoria para alcornoques:

$$X_{(28\times 4)} = \begin{bmatrix} X_{11} & X_{12} & X_{13} & X_{14} \\ X_{21} & X_{22} & X_{23} & X_{24} \\ X_{31} & X_{32} & X_{33} & X_{34} \\ \vdots & \vdots & \ddots & \vdots \\ X_{28,1} & X_{28,2} & X_{28,3} & X_{28,4} \end{bmatrix}$$

Propiedades de Medias

- Vector aleatorio $X_{(p\times 1)}$ con vector de medias $\mu_{(p\times 1)}$:

$$X_{(p\times 1)} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}, \text{ y } \mu_{(p\times 1)} = \mathrm{E}(X) = \begin{bmatrix} \mathrm{E}(X_1) \\ \mathrm{E}(X_2) \\ \vdots \\ \mathrm{E}(X_p) \end{bmatrix} .$$

[1]

- Matríz aleatoria $X_{(n\times p)}$ y matriz aleatoria $Y_{(n\times p)}$:

$$\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y) .$$

(Regla de adición vectorial)

---

[1] Observar que $\mathrm{E}(X_j)$ es calculado a partir de la distribución marginal X_j y no contiene información sobre la relación entre X_j y X_k, $k \neq j$.

- Matriz aleatoria $X_{(n\times p)}$ y matrices de constantes A y B:

$$\mathrm{E}(AXB) = A\mathrm{E}(X)B$$

Prueba: Observar el elemento (i, j) de AXB

$$e_{ij} = \sum_{k=1}^{n} a_{ik} \sum_{l=1}^{p} X_{kl} b_{lj}$$

(donde a_{ik} y b_{lj} son elementos de A y B respectivamente), y ver que $\mathrm{E}(e_{ij})$ es el elemento (i, j) si $A\mathrm{E}(X)B$.

P:

- ¿Cuáles son los análogos univariados de las fórmulas de las dos diapositivas anteriores (que estudió en su primer curso de introducción a la estadística)?
- ¿Qué crees que sucede si miramos $\mathrm{E}(X) + d$ cuando X y d tienen la misma dimensión? ?

R:

$$\mathrm{E}(aX + b) = a\mathrm{E}(X) + b$$

La Covarianza

En los cursos introductorios de estadística se define a la covarianza como:

$$\begin{aligned} \sigma_{ij} &= \mathrm{Cov}(X_i, X_j) = \mathrm{E}[(X_i - \mu_i)(X_j - \mu_j)] \\ &= \mathrm{E}(X_i \cdot X_j) - \mu_i \mu_j \ . \end{aligned}$$

- ¿Qué es la covarianza cuando $i = j$?
- ¿Qué significa que la covarianza sea
  - negativa
  - cero
  - positiva?

Realice un gráfico de dispersión que permita demostrarlo.

Matriz Varianza-Covarianza

- Considere el vector aleatorio $X_{(p\times 1)}$ con vector de medias $\mu_{(p\times 1)}$:

$$X_{(p\times 1)} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}, \text{ y } \mu_{(p\times 1)} = \mathrm{E}(X) = \begin{bmatrix} \mathrm{E}(X_1) \\ \mathrm{E}(X_2) \\ \vdots \\ \mathrm{E}(X_p) \end{bmatrix}$$

- Matriz varianza-covarianza Σ (real y simétrica)

$$\begin{aligned} \Sigma &= \mathrm{Cov}(X) = \mathrm{E}[(X-\mu)(X-\mu)^T] \\ &= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{bmatrix} = \mathrm{E}(XX^T) - \mu\mu^T \end{aligned}$$

- Los elementos de la diagonal de Σ, $\sigma_{ii} = \sigma_i^2$, son varianzas.
- Los elementos fuera de la diagonal son covarianzas $\sigma_{ij} = \mathrm{E}[(X_i - \mu_i)(X_j - \mu_j)] = \sigma_{ji}$.
- Σ es llamada matríz de varianzas, covarianzas o varianza-covarianza y es denotada por $\mathrm{Var}(X)$ o $\mathrm{Cov}(X)$.

Ejercicio: Matriz de varianza-covarianza

Sea $X_{4\times1}$ con matriz de varianza-covarianza

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

Explique que información brinda la matriz.

Correlation matrix

Matriz de correlaciones ρ (real y simétrica)

$$\rho = \begin{bmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}\sigma_{11}}} & \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}} & \cdots & \frac{\sigma_{1p}}{\sqrt{\sigma_{11}\sigma_{pp}}} \\ \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}} & \frac{\sigma_{22}}{\sqrt{\sigma_{22}\sigma_{22}}} & \cdots & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}\sigma_{pp}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{1p}}{\sqrt{\sigma_{11}\sigma_{pp}}} & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}\sigma_{pp}}} & \cdots & \frac{\sigma_{pp}}{\sqrt{\sigma_{pp}\sigma_{pp}}} \end{bmatrix} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{12} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1p} & \rho_{2p} & \cdots & 1 \end{bmatrix}$$

$$\rho = (V^{\frac{1}{2}})^{-1}\Sigma(V^{\frac{1}{2}})^{-1}, \text{ donde } V^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{\sigma_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\sigma_{pp}} \end{bmatrix}$$

Ejercicio: Matriz de correlaciones

Sea $X_{4\times 1}$ con matriz de varianza-covarianza

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

Halle la matriz de correlaciones

R:

$$\rho = \begin{bmatrix} 1 & 0.5 & 0 & 0 \\ 0.5 & 1 & 0 & 0.5 \\ 0 & 0 & 1 & 0.5 \\ 0 & 0.5 & 0.5 & 1 \end{bmatrix}$$

Combinaciones Lineales

Considere el vector aleatorio $X_{(p\times 1)}$ con vector de medias $\mu = \mathrm{E}(X)$ y matriz varianza-covarianza $\Sigma = \mathrm{Cov}(X)$.
Las combinaciones lineales

$$Z = CX = \left[\begin{array}{c} \sum_{j=1}^{p} c_{1j} X_j \\ \sum_{j=1}^{p} c_{2j} X_j \\ \vdots \\ \sum_{j=1}^{p} c_{kj} X_j \end{array} \right]$$

tienen

$$\mathrm{E}(Z) = \mathrm{E}(CX) = C\mu$$

$$\mathrm{Cov}(Z) = \mathrm{Cov}(CX) = C\Sigma C^T$$

Ejercicio: Combinaciones Lineales

$$X = \begin{bmatrix} X_N \\ X_E \\ X_S \\ X_W \end{bmatrix}, \mu = \begin{bmatrix} \mu_N \\ \mu_E \\ \mu_S \\ \mu_W \end{bmatrix}, \text{ y } \Sigma = \begin{bmatrix} \sigma_{NN} & \sigma_{NE} & \sigma_{NS} & \sigma_{NW} \\ \sigma_{NE} & \sigma_{EE} & \sigma_{ES} & \sigma_{EW} \\ \sigma_{NS} & \sigma_{EE} & \sigma_{SS} & \sigma_{SW} \\ \sigma_{NW} & \sigma_{EW} & \sigma_{SW} & \sigma_{WW} \end{bmatrix}$$

A los científicos les gustaría comparar los siguientes tres *contrastes*: N-S, E-W y (E+W)-(N+S), y definir un nuevo vector aleatorio $Y_{(3\times 1)} = C_{(3\times 4)} X_{(4\times 1)}$ dando los tres contrastes.

- Defina C.
- Explique como hallar $\mathrm{E}(Y_1)$ y $\mathrm{Cov}(Y_1, Y_3)$.
- Use R para hallar el vector de medias, la matriz de covarianza y la matriz de correlaciones de Y, cuando el vector de medias y la matriz de covarianza para X se dan a continuación.

```
corkds <- as.matrix(read.table("corkMKB.txt"))
dimnames(corkds)[[2]] <- c("N","E","S","W")
mu=apply(corkds,2,mean)
mu
Sigma=var(corkds)
Sigma
```

```
##        N        E        S        W
## 50.53571 46.17857 49.67857 45.17857
##          N        E        S        W
## N 290.4061 223.7526 288.4378 226.2712
## E 223.7526 219.9299 229.0595 171.3743
## S 288.4378 229.0595 350.0040 259.5410
## W 226.2712 171.3743 259.5410 226.0040
```

```
(C <- matrix(c(1,0,-1,0,0,1,0,1,-1,1,-1,1),byrow=T,nrow=3))
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    1    0   -1    0
## [2,]    0    1    0    1
## [3,]   -1    1   -1    1
```

```
C %*% mu
```

```
##            [,1]
## [1,]  0.8571429
## [2,] 91.3571429
## [3,] -8.8571429
```

```
C %*% Sigma %*% t(C)
```

```
##           [,1]      [,2]       [,3]
## [1,]  63.53439 -38.57672   21.02116
## [2,] -38.57672 788.68254 -149.94180
```

La matriz de covarianza: ¿más requisitos?

Sea el vector aleatorio $X_{(p\times 1)}$ con vector de medias $\mu_{(p\times 1)}$ y matriz de covarianzas

$$\Sigma = \mathrm{Cov}(X) = \mathrm{E}[(X-\mu)(X-\mu)^T] = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{bmatrix}$$

- La matriz de covarianza es simétrica por construcción, y es común requerir que la matriz de covarianza sea semidefinida positiva. Esto significa que, para cada vector $b \neq 0$

$$b^T \Sigma b \geq 0 \ .$$

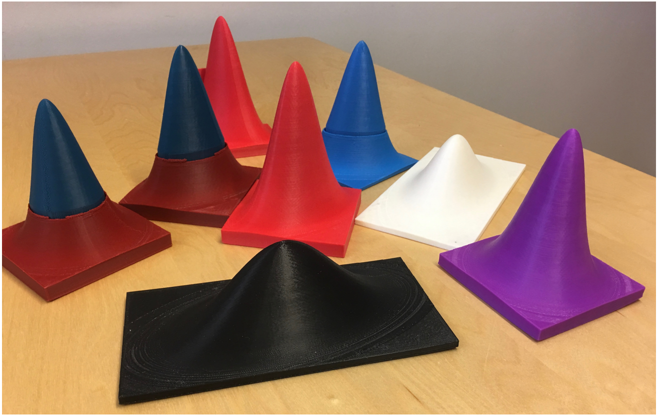
- ¿Cuál crees que sea la razón?

Pista: ¿Es posible que la varianza de la combinación lineal $Y = b^T X$ sea negativa?

La distribución normal multivariada

¿Por qué es tan popular la mvN?

- Se pueden modelar muchos fenómenos naturales utilizando esta distribución (como en el caso univariante).
- Versión multivariante del teorema del límite central: la media de la muestra será aproximadamente normal multivariada para muestras grandes.
- Buena interpretabilidad de la covarianza.
- Matemáticamente manejable.
- Fundamento de muchos modelos y métodos.



Distribuciones normales multivariadas 3D

La fdp de la distribución normal multivariada (mvN)

El vector aleatorio $X_{p\times 1}$ es normal multivariado N_p con media μ y matriz de covarianza (definida positiva) Σ. La fdp será:

$$f(x) = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\}$$

Preguntas:

- ¿Cómo se compara esto con la versión univariante?

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{\frac{1}{2\sigma^2}(x-\mu)^2\}$$

- ¿Por qué necesitamos la constante delante de exp?
- ¿Cuál es la dimensión de la parte de exp?
- ¿Qué pasa si el determinante $|\Sigma| = 0$?

Cuatro propiedades útiles del mvN

Sea $X_{(p\times 1)}$ un vector aleatorio de una $N_p(\mu, \Sigma)$.

1. Los contornos gráficos del mvN son elipsoides (se pueden mostrar mediante descomposición espectral).
2. Las combinaciones lineales de componentes de X son normales (multivariados).
3. Todos los subconjuntos de los componentes de X son normales multivariados (caso especial de los anteriores).
4. La covarianza cero implica que los componentes correspondientes se distribuyen de forma independiente.

Contornos de la distribución normal multivariada

- Los contornos de densidad constante para la distribución normal p dimensional son elipsoides definidos por x tales que
$$(x-\mu)^T \Sigma^{-1}(x-\mu) = b$$
donde $b > 0$ es una constante.
- Estos elipsoides están centrados en μ y tienen ejes en $\pm\sqrt{b\lambda_i}e_i$, donde $\Sigma e_i = \lambda_i e_i$, para $i = 1, ..., p$.
- Para ver esto es útil la descomposición espectral de la matriz de covarianza. Ver: http://www2.stat.duke.edu/~rcs46/lectures_2015/02-multivar/02-multivar.pdf
- $(x-\mu)^T \Sigma^{-1}(x-\mu)$, conocida como el cuadrado de la distancia de Mahalanobis, está distribuida como χ^2_p.

Nota:

- El volumen dentro del elipsoide de x valores que satisface

$$(x-\mu)^T\Sigma^{-1}(x-\mu) \leq \chi^2_p(\alpha)$$

tiene probabilidad $1-\alpha$.

En Clasificación, el mvN es muy importante y, a menudo, se dibujan los contornos del mvN como elipses, y esta es la razón por la que es importante conocer esta propiedad.

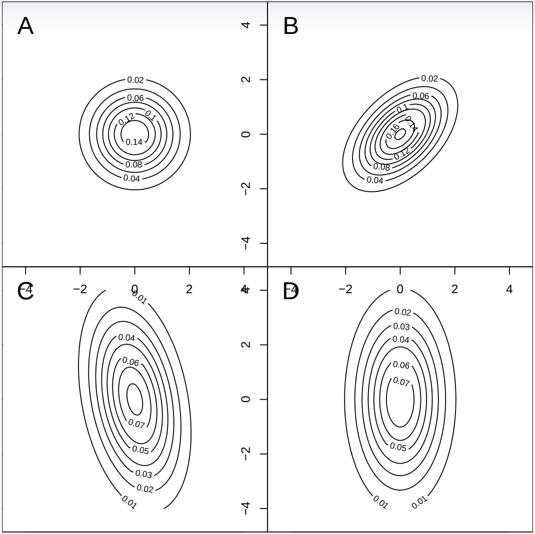
Identificar los mvN a partir de sus contornos.

Sea $\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$.

Se han generado los siguientes cuatro contornos de figuras:

- 1: $\sigma_x = 1$, $\sigma_y = 2$, $\rho = -0.3$
- 2: $\sigma_x = 1$, $\sigma_y = 1$, $\rho = 0$
- 3: $\sigma_x = 1$, $\sigma_y = 1$, $\rho = 0.5$
- 4: $\sigma_x = 1$, $\sigma_y = 2$, $\rho = 0$

** Haga coincidir las distribuciones con las figuras de la siguiente diapositiva. **



Eche un vistazo a los gráficos de contorno: ¿cuándo son círculos los contornos? ¿cuándo elipses?