Estadística Aplicada III

Distribución normal multivariada
Distribuciones muestrales

Jorge de la Vega Góngora

Departamento de Estadística,
Instituto Tecnológico Autónomo de México

Semana 4

Resumen

Distribuciones muestrales

- Hemos visto que los estimadores máximo verosímiles para $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$ de una muestra $\mathbf{x}_1, \ldots, \mathbf{x}_n \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ están dados por:
  - $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}} \sim \mathcal{N}_p\left(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma}\right)$
  - $\hat{\boldsymbol{\Sigma}} = \mathbf{S}_n = \left(\frac{n-1}{n}\right)\mathbf{S}$.
  - La distribución de la forma cuadrática $(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})$ es $\chi^2_{(p)}$.
- Todavía tenemos que revisar las siguientes ideas:
  - ¿Cuál es la relación entre $\bar{\mathbf{x}}$ y $\mathbf{S}$?
  - ¿Cómo se distribuye $\frac{n-1}{n}\mathbf{S}$?
  - ¿Cómo se distribuye $(\mathbf{x} - \bar{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{x} - \bar{\mathbf{x}})$, la forma cuadrática de la distancia estadística sustituyendo los parámetros por sus estimadores?

Distribución muestral de **S** I

- En el caso univariado,cabe recordar que se tienen las siguientes distribuciones muestrales:
  - $\bar{X} \sim \mathcal{N}\left(\mu, \sigma^2\right)$
  - $(n-1)s^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2 \sim \sigma^2\chi^2_{(n-1)}$
  - $\bar{X}$ y s^2 son independientes.
- Podemos notar que podemos reescribir, si $Z_i \sim \mathcal{N}(0,1), i = 1, \ldots, n-1$ independientes, entonces $(n-1)s^2 \stackrel{d}{=} (\sigma Z_1)^2 + \cdots (\sigma Z_{n-1})^2 \sim \chi^2_{n-1}$. Esta forma es la que nos permitirá extender la definición al caso multivariado.

Distribución muestral de **S** II

Distribución Wishart

- La distribución muestral de la matriz de varianzas y covarianzas de una normal multivariada se llama la *distribución de Wishart* y se define como la distribución de la suma de productos independientes de vectores normales:

$$\mathcal{W}_m(\cdot|\mathbf{\Sigma}) \stackrel{d}{=} \sum_{i=1}^{m} \mathbf{Z}_i\mathbf{Z}_i'$$

  donde $\mathbf{Z}_i \sim \mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma})$ y $\mathbf{Z}_i \perp\!\!\!\perp \mathbf{Z}_j$.
- La densidad Wishart con m grados de libertad para una matriz $\mathbf{A}_{p\times p}$ definida positiva, cuando $n > p$, está dada por:

$$w_m(\mathbf{A}|\mathbf{\Sigma}) = \frac{|\mathbf{A}|^{\frac{m-p-1}{2}} e^{-tr[\mathbf{A}\mathbf{\Sigma}^{-1}]/2}}{2^{pm/2}\pi^{p(p-1)/4}|\mathbf{\Sigma}|^{m/2}\prod_{i=1}^{p}\Gamma\left(\frac{1}{2}(m+1-i)\right)}$$

Entonces $n\hat{\mathbf{\Sigma}} = (n-1)\mathbf{S} \sim \mathcal{W}_{n-1}((n-1)\mathbf{S}|\mathbf{\Sigma})$. Por último, se puede probar que $\bar{\mathbf{x}}$ y $\mathbf{S}$ son independientes.

Propiedades de la distribución Wishart

- La distribución Wishart es la generalización multivariada de la distribución χ^2.
- Las siguientes propiedades de la distribución Wishart serán utilizadas en algunos casos más adelante:
  1. Si $\mathbf{S}_1 \sim \mathcal{W}_n(\mathbf{S}_1|\boldsymbol{\Sigma})$ y $\mathbf{S}_2 \sim \mathcal{W}_m(\mathbf{S}_2|\boldsymbol{\Sigma})$ independientes, entonces
  $$\mathbf{S} = \mathbf{S}_1 + \mathbf{S}_2 \sim \mathcal{W}_{n+m}(\mathbf{S}|\boldsymbol{\Sigma}).$$
  2. Si $\mathbf{S} \sim \mathcal{W}_n(\mathbf{S}|\boldsymbol{\Sigma})$ y $\mathbf{C}_{l\times p} \neq \mathbf{0}$ entonces
  $$\mathbf{B} = \mathbf{CSC}' \sim \mathcal{W}_n(\mathbf{B}|\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}').$$
  3. La densidad Wishart existe cuando $n > p$.

Teorema del límite central multivariado

Teorema del límite central

Sea $\mathbf{X}_1, \ldots, \mathbf{X}_n$ una muestra aleatoria de una población con media $\boldsymbol{\mu}$ y covarianza finita no singular $\boldsymbol{\Sigma}$. Entonces

$$\sqrt{n}(\bar{\mathbf{X}} - \boldsymbol{\mu}) \dot{\sim} \mathcal{N}_p\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$$

para tamaños de muestra n grandes y $n >> p$.

- Como consecuencia del teorema del límite central, se tiene que $n(\bar{\mathbf{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu}) \dot{\sim} \chi^2_{(p)}$ para $n - p$ grande.
- Notar que la aproximación es para cualquier distribución, no necesariamente normal.

Distribución T^2 de Hotelling I

- Cabe recordar que en el caso univariado, si $X \sim \mathcal{N}\left(\mu_0, 1\right)$ independiente de $V \sim \chi^2_{(p)}$, entonces
$$t = \frac{X - \mu_0}{\sqrt{(V/n)}} \sim t_{(n)}$$
y adicionalmente, sabemos que $t^2_{(n)} = n(X - \mu_o)(s^2)^{-1}(X - \mu_o) \sim F_{1,p}$, donde $F_{u,v}$ es la distribución F de Fisher, con u y v grados de libertad.
- La generalización de la distribución t al caso multivariado, corresponde a la distribución de una forma cuadrática.

Distribución T^2 de Hotelling

Para una muestra aleatoria $\mathbf{x}_1, \ldots, \mathbf{x}_n \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$,y si $\boldsymbol{\mu}_0$ es un vector fijo, la estadística

$$T^2 = n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0) \sim \frac{(n-1)p}{n-p} F_{p,n-p}.$$

Esta estadística es llamada T^2 *de Hotelling*. Esta estadística combina una normal multivariada con una matriz Wishart, y es análoga a la distribución t univariada elevada al cuadrado, que coincide con una distribución F.

- Para la estadística T^2 de Hotelling, una de sus principales características es que es invariante bajo transformaciones afines $\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{d}$, donde $\mathbf{C}$ no sea singular. Esto es fácil de demostrar considerando las propiedades de la media y varianza de $\mathbf{y}$ a partir de la de $\mathbf{x}$.

Ejemplos I

1. Con la matriz de datos $\mathbf{X} = \begin{bmatrix} 2 & 12 \\ 8 & 9 \\ 6 & 9 \\ 8 & 10 \end{bmatrix}$, calculen T^2 para la prueba $H_0 : \boldsymbol{\mu}' = (7, 11)$, suponiendo que los datos son muestra de una distribución binormal.
2. ¿Cuál es la distribución de T^2?
3. ¿Cómo se distribuye $3\mathbf{S}$?

Solución.

Ejemplos II

```
X <- matrix(c(2,8,6,8,12,9,9,10),nrow=4)
n <- dim(X)[1]
mu <- c(7,11)
(S <- var(X))

          [,1]      [,2]
[1,]  8.000000 -3.333333
[2,] -3.333333  2.000000

(Sinv <- solve(S))

          [,1]      [,2]
[1,] 0.4090909 0.6818182
[2,] 0.6818182 1.6363636

(xbar <- colMeans(X))

[1]  6 10

(T2 <- as.numeric(n*(t(xbar-mu) %*% Sinv %*% (xbar-mu))))

[1] 13.63636
```

La distribución de $T^2 \sim \frac{2(4-1)}{4-2} F_{2,4-2} = 3F_{2,2}$. La distribución de la matriz $3\mathbf{S}$ es Wishart con 3 grados de libertad.

$\square$

Ejemplos III

4. Verifiquen que pasa con T^2 si cada caso de la matriz $\mathbf{X}$ se sustituye por $\mathbf{C}\mathbf{x}_i + \mathbf{d}$, donde $\mathbf{C} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ y $\mathbf{d} = (1,1)'$.

Solución.

```
C <- matrix(c(1,1,-1,1),nrow=2)
Y <- X %*% t(C) + c(1,1)
ybar <- colMeans(Y)
S2 <- var(Y)
muy <- C%*%mu + c(1,1)
T3 <- as.numeric(n*(t(ybar-muy) %*% solve(S2) %*% (ybar-muy)))
T3

[1] 13.63636
```

Vimos que la distribución no cambia bajo transformaciones afines, por las propiedades de las medias y las varianzas.

□

Requerimientos de optimización matricial

- Durante el curso en diversas partes utilizaremos los siguientes resultados de optimización de cocientes de formas cuadráticas, que parten en todos los casos de la desigualdad de Cauchy-Schwarz. Revisaremos los resultados para tenerlos presentes.

Teorema (Desigualdad de Cauchy-Scwharz)

Sean $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$. Entonces

$$(\mathbf{x}'\mathbf{y})^2 \leq (\mathbf{x}'\mathbf{x})(\mathbf{y}'\mathbf{y})$$

y la igualdad se cumple si y sólo si $\mathbf{x} = c\mathbf{y}$ para alguna $c \in \mathbb{R}$

- El primer paso es extender esta desigualdad para incorporar una matriz definida positiva.

Teorema (Desigualdad de Cauchy-Scwharz extendida)

Sean $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$ y $\mathbf{B}_{p\times p} > 0$ (definida positiva). Entonces:

$$(\mathbf{x}'\mathbf{y})^2 \leq (\mathbf{x}'\mathbf{B}\mathbf{x})(\mathbf{y}'\mathbf{B}^{-1}\mathbf{y})$$

y la igualdad se da si y sólo si $\mathbf{x} = c\mathbf{B}^{-1}\mathbf{y}$ para alguna $c \in \mathbb{R}$.

Demostración.

Básicamente tenemos que escribir $\mathbf{B} = \mathbf{B}^{1/2}\mathbf{B}^{1/2}$ usando la descomposición espectral. Entonces

$$\mathbf{x}'\mathbf{y} = \mathbf{x}'\mathbf{I}\mathbf{y} = \mathbf{x}'\mathbf{B}^{1/2}\mathbf{B}^{-1/2}\mathbf{y} = (\mathbf{B}^{1/2}\mathbf{x})'(\mathbf{B}^{-1/2}\mathbf{y})$$

Un recordatorio rápido III

y el resultado sigue por la desigualdad de Cauchy-Schwarz.

□

- El siguiente lema es el primer resultado sobre optimización del cociente de formas cuadráticas.

Lema

Sea $\mathbf{B}_{p\times p} > 0$ y $\mathbf{a} \in \mathbb{R}^p$. Entonces para todo vector $\mathbf{x} \neq \mathbf{0}$ en $\mathbb{R}^p$

$$\max_{\mathbf{x}\neq\mathbf{0}} \frac{(\mathbf{x}'\mathbf{a})^2}{\mathbf{x}'\mathbf{B}\mathbf{x}} = \mathbf{a}'\mathbf{B}^{-1}\mathbf{a}$$

donde el máximo se alcanza cuando $\mathbf{x} = c\mathbf{B}^{-1}\mathbf{a}$ para alguna $c \neq 0$.

Demostración.

Un recordatorio rápido IV

Por el teorema extendido de Cauchy-Schwarz $(\mathbf{x}'\mathbf{a})^2 \leq (\mathbf{x}'\mathbf{B}\mathbf{x})(\mathbf{a}'\mathbf{B}^{-1}\mathbf{a})$. Además como $\mathbf{x} \neq \mathbf{0}$ y $\mathbf{B} > \mathbf{0}$, $\mathbf{x}'\mathbf{B}\mathbf{x} > 0$. Entonces:

$$\frac{(\mathbf{x}'\mathbf{a})^2}{\mathbf{x}'\mathbf{B}\mathbf{x}} \leq \mathbf{a}'\mathbf{B}^{-1}\mathbf{a}$$

Como la igualdad se alcanza con $\mathbf{x} = c\mathbf{B}^{-1}\mathbf{a}$, está claro que:

$$\max_{\mathbf{x}\neq\mathbf{0}} \frac{(\mathbf{x}'\mathbf{a})^2}{\mathbf{x}'\mathbf{B}\mathbf{x}} = \mathbf{a}'\mathbf{B}^{-1}\mathbf{a}$$

$\square$

Un recordatorio rápido V

- Por ultimo, el siguiente teorema relaciona los óptimos con los eigenvalores.

Teorema (Maximización de formas cuadráticas para puntos en la esfera unitaria)

Sea $\mathbf{B}_{p\times p} > \mathbf{0}$ con eigenvalores $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ y correspondientes eigenvectores normalizados $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_p\}$. Entonces:

$$\max_{\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}'\mathbf{B}\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \lambda_1 \qquad \text{que se alcanza cuando } \mathbf{x} = \mathbf{e}_1$$

$$\min_{\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}'\mathbf{B}\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \lambda_p \qquad \text{que se alcanza cuando } \mathbf{x} = \mathbf{e}_p$$

$$\max_{\mathbf{x}\perp\mathbf{e}_1,\ldots\mathbf{e}_k} \frac{\mathbf{x}'\mathbf{B}\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \lambda_{k+1} \qquad \text{que se alcanza cuando } \mathbf{x} = \mathbf{e}_{k+1}, \quad k = 1, 2, \ldots, p-1$$

Un recordatorio rápido VI

Demostración.

Supongamos que $\mathbf{B} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'$ es la descomposición espectral de tal manera que $\mathbf{B}^{1/2} = \mathbf{P}\Lambda^{1/2}\mathbf{P}'$, donde la matriz $\mathbf{P}$ cumple con la condición $\mathbf{P}'\mathbf{P} = \mathbf{P}\mathbf{P}' = \mathbf{I}$. Sea $\mathbf{y} = \mathbf{P}'\mathbf{x}$. Entonces, si $\mathbf{x} \neq \mathbf{0}, \mathbf{y} \neq \mathbf{0}$ y

$$\begin{aligned}
\frac{\mathbf{x}'\mathbf{B}\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \frac{\mathbf{x}'\mathbf{B}^{1/2}\mathbf{B}^{1/2}\mathbf{x}}{\mathbf{x}'\mathbf{P}\mathbf{P}'\mathbf{x}} &= \frac{\mathbf{x}'\mathbf{P}\Lambda^{1/2}\mathbf{P}'\mathbf{P}\Lambda^{1/2}\mathbf{P}'\mathbf{x}}{\mathbf{y}'\mathbf{y}} \\
&= \frac{\mathbf{y}'\Lambda\mathbf{y}}{\mathbf{y}'\mathbf{y}} \\
&= \frac{\sum_{i=1}^{p}\lambda_i y_i^2}{\sum_{i=1}^{p} y_i^2} \\
&\leq \lambda_1 \frac{\sum_{i=1}^{p} y_i^2}{\sum_{i=1}^{p} y_i^2} = \lambda_1
\end{aligned}$$

Si se toma $\mathbf{x} = \mathbf{e}_1$ entonces $\mathbf{y} = \mathbf{P}'\mathbf{e}_1 = \mathbf{e}_1$, entonces $\frac{\mathbf{e}_1'\mathbf{B}\mathbf{e}_1}{\mathbf{e}_1'\mathbf{e}} = \mathbf{e}_1\mathbf{B}\mathbf{e}_1 = \lambda_1$.

Por otra parte, $\mathbf{y} = \mathbf{P}'\mathbf{x} \Rightarrow \mathbf{P}\mathbf{y} = \mathbf{x} = \sum_{i=1}^{p} y_i \mathbf{e}_i$. Si $\mathbf{x} \perp \mathbf{e}_1, \ldots \mathbf{e}_k$ entonces

$$0 = \mathbf{e}_i'\mathbf{x} = y_1\mathbf{e}_1'\mathbf{e}_1 + y_2\mathbf{e}_2'\mathbf{e}_2 + \cdots + y_p\mathbf{e}_p'\mathbf{e}_p = y_i, \quad i \leq k$$

Por lo tanto el cociente se puede escribir como:

$$\frac{\mathbf{x}'\mathbf{B}\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \frac{\sum_{i=k+1}^{p} \lambda_i y_i^2}{\sum_{i=k+1}^{p} y_i^2}$$

Si tomamos $y_{k+1} = 1$ y el resto de las $y_{k+i} = 0$ para $i = 2, \ldots, p$, se obtiene el máximo establecido.

$\square$

Ejemplo. []

Sea $\mathbf{B} = \begin{pmatrix} 2.2 & 0.4 \\ 0.4 & 2.8 \end{pmatrix}$. consideremos como función de $\mathbf{x}$ la forma cuadrática $g(\mathbf{x}) = \frac{\mathbf{x}'\mathbf{B}\mathbf{x}}{\mathbf{x}'\mathbf{x}}$.

Un recordatorio rápido VIII

Los eigenvalores de **B** son $\lambda_1 = 3$ y $\lambda_2 = 2$ y la matriz **P** $= \begin{pmatrix} 0.4472136 & -0.8944272 \\ 0.8944272 & 0.4472136 \end{pmatrix}$ es la matriz de eigenvectores. Las direcciones hacia donde se maximiza y minimiza respectivamente la función $g(\mathbf{x})$ son precisamente las columnas de **P**, que se muestran en la gráfica con las flechas.

Adicionalmente se generan una muestra de puntos para calcular el valor de la función g y ver que precisamente en la dirección de esos dos vectores es donde se tiene el valor más alto y el más bajo respectivamente.

Un recordatorio rápido IX

```r
options(scipen=5) #control de expansión en notación científica
set.seed(100)
B <- matrix(c(2.2,.4,.4,2.8),nrow=2) #matriz ejemplo

#Define la función cociente forma cuadrática
g <- function(x,Q){ as.numeric((t(x) %*% Q %*% x)/sum(x*x))}

#genera algunas direcciones al azr en el cuadro [-2,2]^2
vectores <- cbind(runif(100,-2,2),runif(100,-2,2))
#calcula la función cuadrática en esas direcciones
z <- apply(vectores, 1, g, Q = B)
z
```

```
  [1] 2.865464 2.605354 2.702268 2.481084 2.405223 2.725723 2.985216 2.261115
  [9] 2.719030 2.121670 2.988370 2.825763 2.999635 2.991407 2.997557 2.471272
 [17] 2.221621 2.268857 2.999396 2.997854 2.488158 2.000051 2.024597 2.502406
 [25] 2.998946 2.979134 2.998698 2.947164 2.914689 2.997919 2.834432 2.842367
 [33] 2.265531 2.195287 2.566529 2.173319 2.273766 2.546125 2.143509 2.015677
 [41] 2.106951 2.139769 2.185648 2.955779 2.554457 2.768516 2.991348 2.157182
 [49] 2.993137 2.988717 2.176551 2.223113 2.972192 2.982185 2.941364 2.184734
 [57] 2.074136 2.120140 2.620404 2.018893 2.701188 2.612524 2.166557 2.344150
 [65] 2.894450 2.995829 2.868977 2.566814 2.107218 2.998822 2.781370 2.625233
 [73] 2.943850 2.770820 2.998349 2.217282 2.961285 2.682076 2.111565 2.043335
 [81] 2.926591 2.998736 2.902600 2.525851 2.761411 2.350119 2.275301 2.988704
 [89] 2.904123 2.980375 2.949500 2.392441 2.999654 2.704568 2.554010 2.998603
 [97] 2.762132 2.040846 2.828235 2.917087
```

```r
#Ahora calcula los valores propios, grafica las direcciones de los vectores propios y muestra las direcciones simuladas
```

Un recordatorio rápido X

```r
lambdas <- eigen(B)$values
        P <- eigen(B)$vectors
        par(pty = "s") #gráfica cuadrada
        plot(c(-2,2), c(-2,2), type = "n", xlim = c(-2,2), ylim = c(-2,2), xlab = "", ylab = "",
             main = "gráfica de la función g(x) y su valor") #gráfica vacía
        abline(h=0)
        abline(v=0)
        arrows(0,0,P[1,1],P[2,1], length = 0.1, lwd = 2)
        arrows(0,0,P[1,2],P[2,2], length = 0.1, lwd = 2)
        arrows(0,0,vectores[,1],vectores[,2], length=0.1, pch = 16, cex = 0.5, lty=3, col = "gray2")
        text(vectores, labels = round(z,3), adj=c(0,-0.2), cex = 0.7, col = "red")
```

Un recordatorio rápido XI