eman ta zabal zazu

Universidad del País Vasco — Euskal Herriko Unibertsitatea

TAREA 1

EJERCICIOS

Por la teoría vista en clase sabes que cuando $\vec{X}$ es normal multivariante, la distribución de $\vec{X}_1|\vec{X}_2 = \vec{x}_2$ viene dada por

$$N(\vec{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\vec{x}_2 - \vec{\mu}_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}). \tag{1}$$

Los ejercicios que siguen son de manipulación y tienen por objeto que compruebes empíricamente algunas de las cosas estudiadas en teoría.

1. Genera 500 observaciones normales multivariantes con vector de medias $\vec{\mu} = (\vec{\mu}_1 \quad \vec{\mu}_2)' = (3 \quad 4)'$ y matriz de covarianzas
$$\Sigma = \begin{pmatrix} 1 & 0{,}8 \\ 0{,}8 & 1 \end{pmatrix}. \tag{2}$$

2. Estima máximo-verosímilmente el vector de medias y la matriz de covarianzas. Con un tamaño de muestra $N = 500$ debes obtener estimaciones bastante ajustadas.

3. Como has generado artificialmente las observaciones, sabes cual es su función de densidad teórica. Calcúlala para cada observación, divide el rango de valores que obtengas en tres intervalos (los cien mayores, los doscientos siguientes y los doscientos menores) y representa las observaciones correspondientes. ¿Qué obtienes?

4. ¿Que orientación (paralela a los ejes, SW-NE, NW-SE, ...) tiene cualquiera de las nubes de puntos representadas en el apartado anterior?

5. Selecciona un valor de $\vec{X}_2$ (por ejemplo, 0.8) y calcula la media y varianza teóricas de $\vec{X}_1$ para observaciones con $\vec{X}_2 = 0{,}8$.

6. Siendo la distribución de $\vec{X}_2$ continua, sólo por autentica casualidad obtendrías alguna de tus 500 observaciones con $\vec{x}_2 = 0{,}8$. Por tanto, no puedes estimar la media obtenida en el apartado anterior de modo empírico. Pero si puedes calcular la media aritmética de valores de $\vec{X}_1$ en observaciones con $\vec{X}_2 \approx 0{,}8$ (por ejemplo, entre 0.7 y 0.9); no debería separarse mucho del valor teórico obtenido en el apartado anterior.

7. Idem. con la varianza de $\vec{X}_1|\vec{X}_2 = 0{,}8$.

8. Genera ahora 500 observaciones normales trivariantes, $\vec{X} = (X_1, X_2, X_3)'$ con vector de medias $\vec{0}$ y matriz de covarianzas unidad. Obtén un nuevo vector de variables normales multivariantes (Y_1, Y_2, Y_3) a partir de $\vec{X}$ del siguiente modo:
$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 3 \\ 1 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}. \tag{3}$$
Por simple inspección ¿cuál dirías que es la correlación de Y_1 con Y_2? ¿Y la correlación *parcial* controlado el efecto de X_3?

9. Verifica tu intuición en el apartado anterior calculando los valores teóricos de las correlaciones indicadas.

10. Comprueba, a continuación, que las estimaciones de las respectivas correlaciones son razonablemente aproximadas a sus valores teóricos.

11. *(optativo)* Computa los valores de una función de densidad normal bivariante en una malla de puntos (x, y). Dibuja sus contornos de igual densidad para los casos de las siguientes matrices:

$$\Sigma = \begin{pmatrix} 1 & 0{,}8 \\ 0{,}8 & 1 \end{pmatrix}; \qquad \Sigma = \begin{pmatrix} 1 & -0{,}8 \\ -0{,}8 & 1 \end{pmatrix}; \qquad \Sigma = \begin{pmatrix} 1 & 0{,}1 \\ 0{,}1 & 1 \end{pmatrix}.$$

Observa las formas y orientaciones de los ejes de simetría y comenta los resultados.

AYUDAS, SUGERENCIAS Y COMPLEMENTOS

1. Para generar observaciones normales multivariantes con vector de medias y matriz de covarianzas arbitraria, sólo tienes que generar normales multivariantes con vector de medias $\vec{0}$ y matriz de covarianzas unidad, y hacer la oportuna transformación lineal $\vec{Y} = \Sigma^{1/2}\vec{X} + \vec{\mu}_1$. Si trabajas con S-PLUS y consultas la ayuda *on line* de la función `rnorm` verás un ejemplo de función `rmultinorm`. Escribe la tuya propia, aunque sea menos eficiente, si no estás seguro de entender lo que `rmultinorm` hace.

2. Para dibujar contornos de igual densidad, te será de utilidad la función de R `contour` (`persp` es una función relacionada si quisieras una gráfica en perspectiva). Habrás de computar la función de densidad de una normal bivariante sobre una malla de puntos: puedes emplear la función siguiente como modelo:

```
ndes2 <- function(x,y,m,sigma) {
    f <- outer(x,y)
    invsigma <- solve(sigma)
    raizdetsigma <- sqrt(det(sigma))
    for (i in 1:length(x)) {
      for (j in 1:length(y)) {
        v <- c(x[i],y[j]) - m
        f[i,j] <- exp(-0.5*v %*% invsigma %*% v) / (2*pi*raizdetsigma)
      }
    }
    return(f)
 }
```

3. Cualquiera de los manuales citados en la bibliografía del programa te servirá: [7], [4], [6], [5], [1], [2], [3], . . .

Referencias

[1] T.W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley, 1984.

[2] C.M. Cuadras. *Métodos de Análisis Multivariante*. Eunibar, Barcelona, 1981.

[3] W.R. Dillon and M. Goldstein. *Multivariate Analysis: Methods and Applications*. Wiley, New York, 1984.

[4] J.D. Jobson. *Applied Multivariate Data Analysis, vol. II*. Springer Verlag, New York, 1991. Signatura: 519.237 JOB.

[5] R.A. Johnson and D.W. Wichern. *Applied Multivariate Statistical Analysis*. Prentice Hall, 1992.

[6] A.C. Rencher. *Methods of Multivariate Analysis*. Wiley, 1995.

[7] A.C. Rencher. *Multivariate Statistical Inference and Applications*. Wiley, 1998.