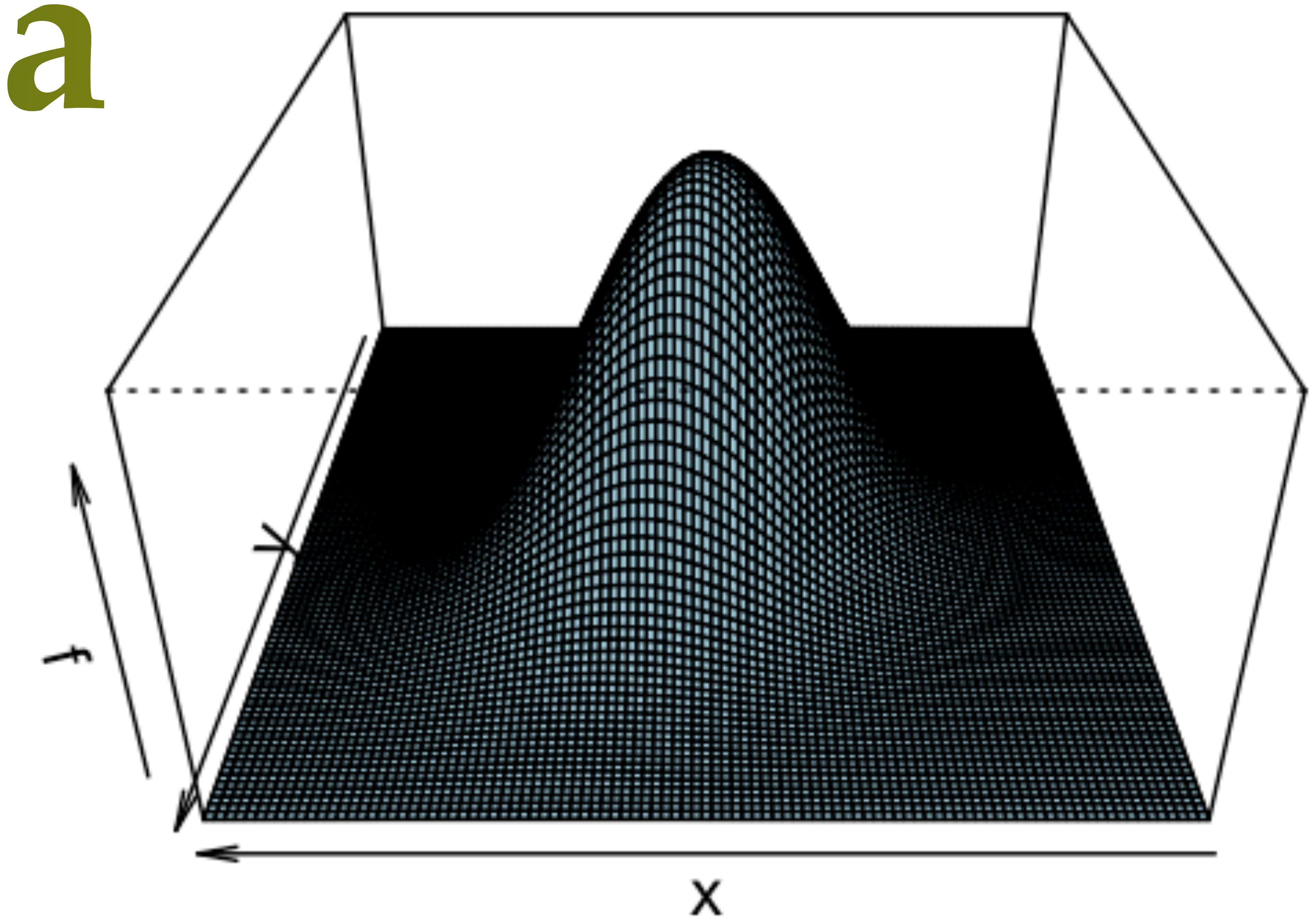
Distribución normal multivariada



Distribución normal multivariada

Definición 1

Se dice que $\mathbf{x}_{p\times 1}$ un vector aleatorio sigue una distribución normal multivariada no singular y denotado por $\mathbf{x} \sim N_p\left(\mu, \Sigma\right)$, si su densidad está dada por:

$$f(\mathbf{x}) = \frac{1}{|2\pi\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\right],$$

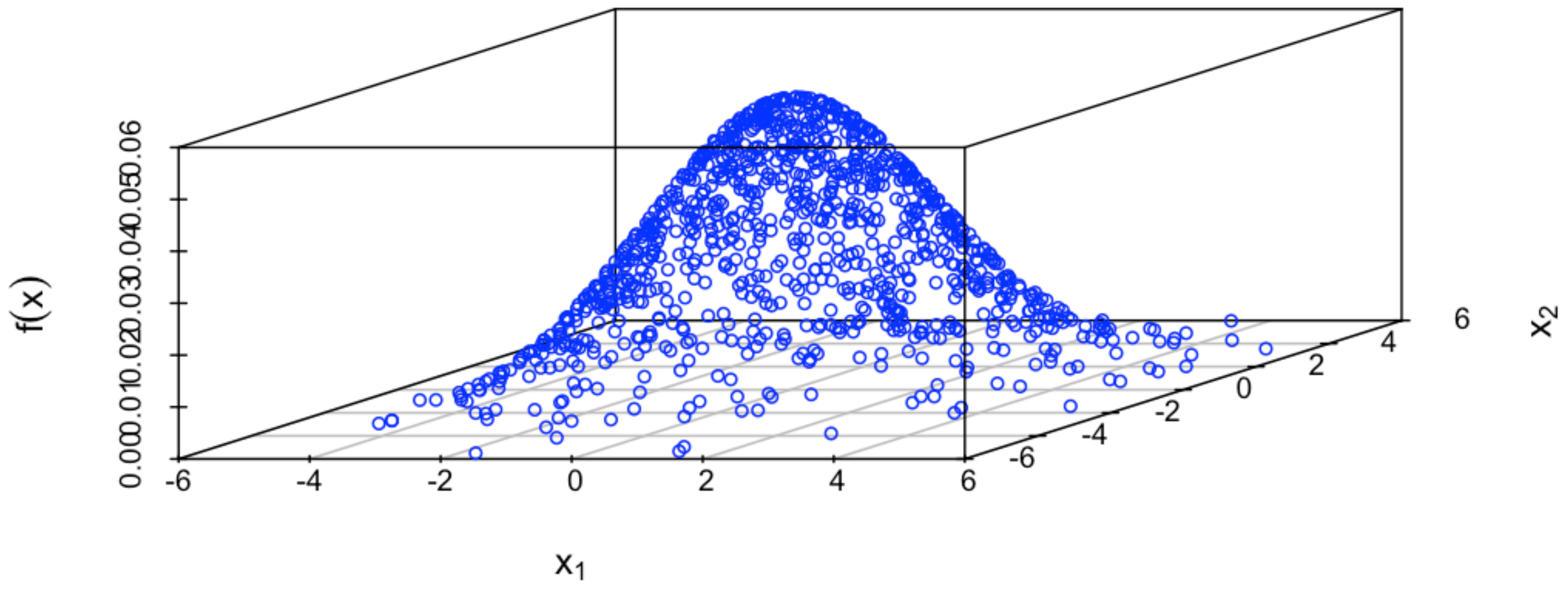
donde

- $\mathbb{E}(\mathbf{x}) = \mu$

- $\mathrm{Var}(\mathbf{x}) = \Sigma > 0$ (positiva definida)

Diagrama de dispersión

Para vectores bivariados se puede obtener un diagrama de dispersión 3D con la librería scatterplot3d



Distribución normal multivariada

Observación 1

Si $\text{ran}(\Sigma) = k < p$ se define la densidad de la distribución normal multivariada singular como

$$f(\mathbf{x}) = \frac{(2\pi)^{-\frac{k}{2}}}{(\lambda_1 \cdots \lambda_k)^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-} (\mathbf{x}-\mu)\right],$$

donde

- $\mathbf{x}$ vive en el híper-plano $\mathbf{N}'(\mathbf{x}-\mu)$ y $\mathbf{N}$ es una matriz de tamaño $p \times (p-k)$ tal que:

    1. $\mathbf{N}^T\Sigma = \mathbf{0}$
    2. $\mathbf{N}^T\mathbf{N} = \mathbf{I_{p-k}}$

- Σ^{-} es la inversa generalizada y $\lambda_1, \ldots, \lambda_k$ son los eigenvalores diferentes de cero.

Caracterizaciones

Definición 2

Decimos que el vector aleatorio $\mathbf{x}_{p\times 1}$ sigue una distribución normal p-variada si y solo si $\mathbf{a}^T\mathbf{x}$ tiene una distribución normal univariada para todos los vectores p-variados (no triviales) $\mathbf{a}$

Proposición 1

Sea $\mathbf{x}$ un vector normal p-variado. Si se define a $\mathbf{y}_{q\times 1} = \mathbf{A}_{q\times p}\mathbf{x} + \mathbf{b}_{q\times 1}$ entonces el vector aleatorio $\mathbf{y}$ tiene una distribución normal q-variada con esperanza y varianza dadas por:

$$\mathbb{E}(\mathbf{y}) = \mathbf{A}\mu + \mathbf{b} \qquad \text{Var}(\mathbf{y}) = \mathbf{A}\Sigma\mathbf{A}^{\mathbf{T}}$$

Caracterizaciones

Corolario 1

Sean $\mathbf{x} \sim N_p(\mathbf{0}, \mathbf{I}_p)$, $\mu_{p\times 1}$ un vector real y $\Sigma_{p\times p}$ una matriz semidefinda positiva y simétrica. Entonces $\mathbf{y} = \Sigma^{\frac{1}{2}}\mathbf{x} + \mu \sim N_p(\mu, \Sigma)$

Corolario 2

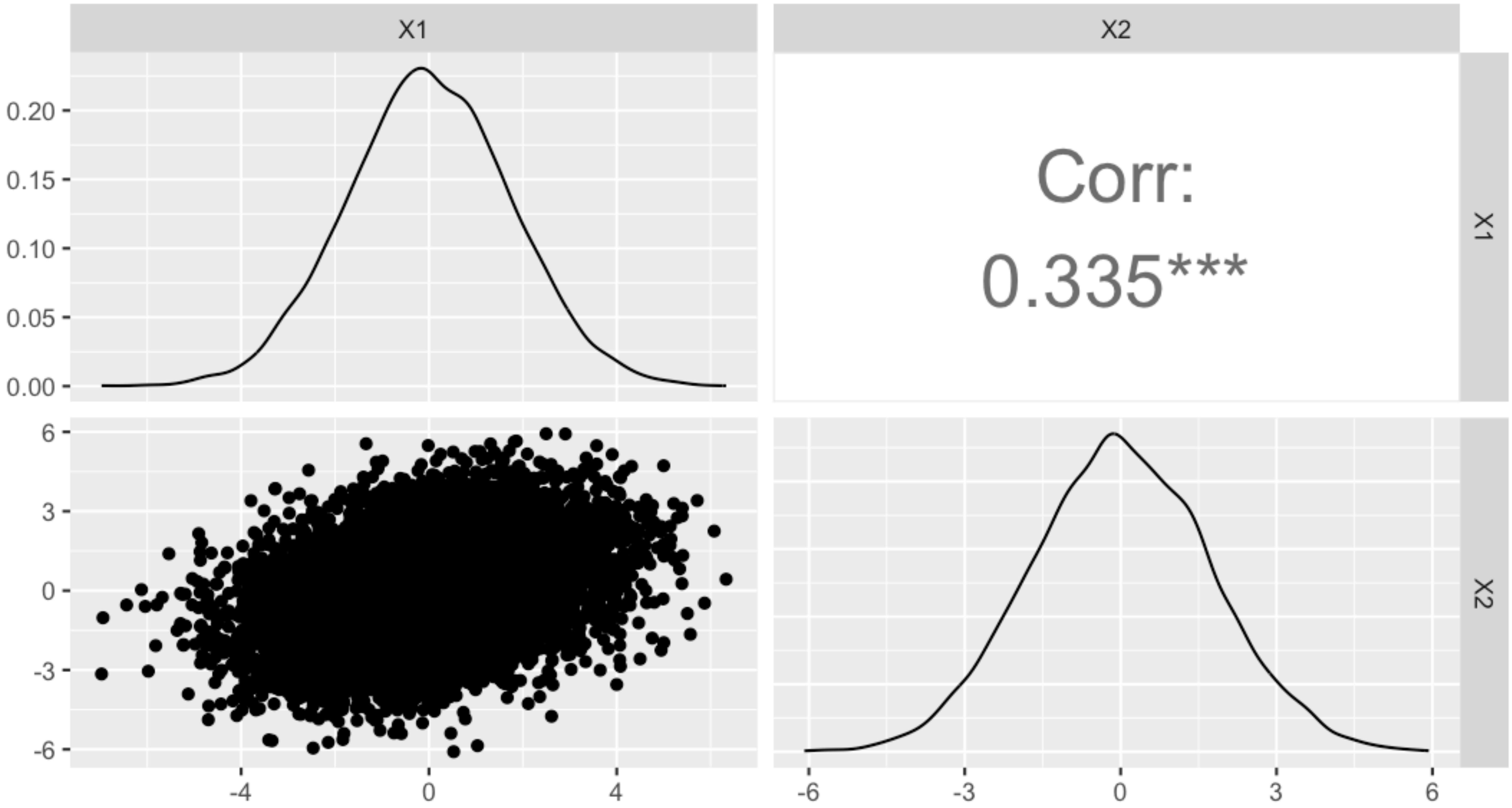
Sean $\mathbf{x} \sim N_p(\mu, \Sigma)$ con $\Sigma > 0$ y $\mathbf{y} = \Sigma^{-\frac{1}{2}}(\mathbf{x} - \mu)$, donde $\Sigma^{-\frac{1}{2}}$ es la matriz raíz cuadrada de Σ^{-1}. Entonces, $y_1, y_2, \ldots, y_p$ son variables aleatorias iid $N(0,1)$.

- En **R** la librería expm proporciona la función requerida para obtener $\Sigma^{-\frac{1}{2}}$ con sqrtm

Simulación

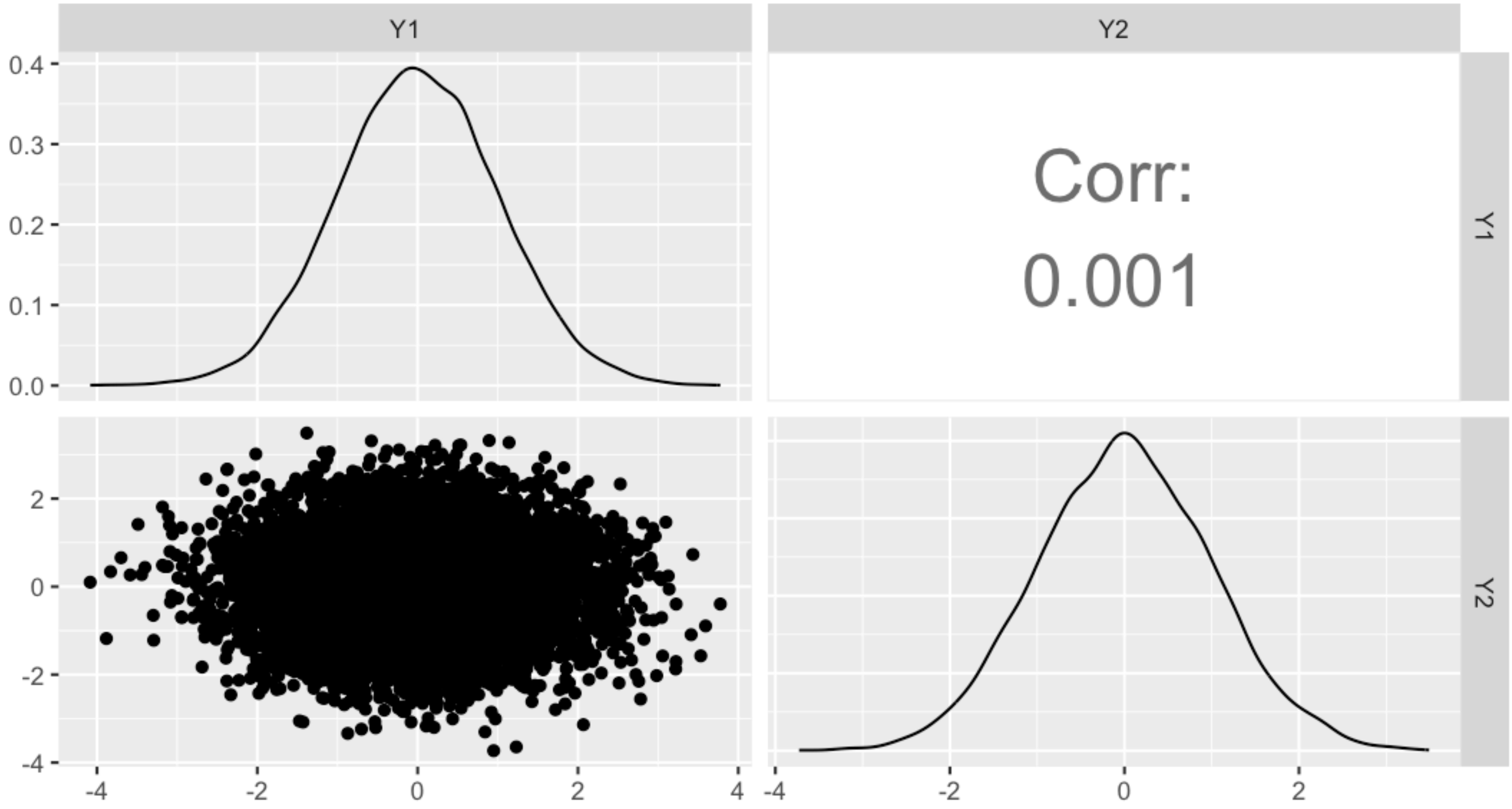
- ¿Cómo se ven las densidades univariadas y las curvas de nivel de $\mathbf{x} \sim N_2(\mu, \Sigma)$? Donde

$$\mu = \begin{pmatrix} 3 \\ 3 \end{pmatrix} \qquad \Sigma = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$$



Simulación

- ¿Qué pasa si hacemos $\mathbf{y} = \Sigma^{-\frac{1}{2}}(\mathbf{x} - \mu)$?

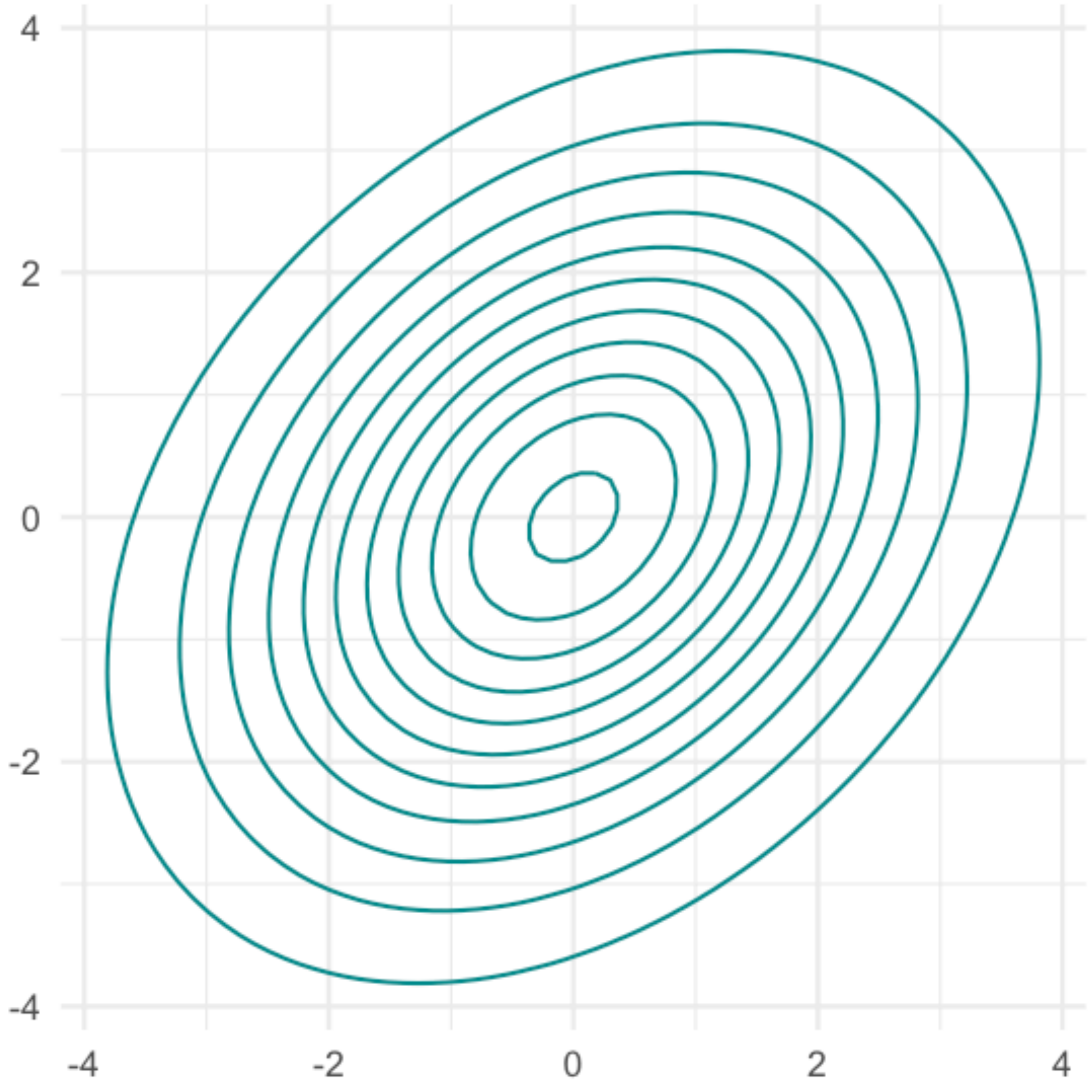


Propiedades

Observación 2

La distribución normal multivariada tiene densidad constante en elipses (elipsoides)

$$(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)=k$$



Propiedades

Proposición 2

Sea $\mathbf{x} \sim N_p(\mu, \Sigma)$ entonces, $U = (\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu) \sim \chi^2_p$.

Observación 3

Se puede fácilmente evaluar la probabilidad de que $\mathbf{x}$ este en un elipsoide, i.e.

$$\mathbb{P}\left[(\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu) < k\right]$$

Propiedades

Proposición 3

Sea $\mathbf{x} \sim N_p(\mu, \Sigma)$, entonces los coeficientes de asimetría y curtosis están dados respectivamente por,

$$\beta_{1,p} = \mathbb{E}\left[\left((\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{y}-\mu)\right)^3\right] = 0$$

$$\beta_{2,p} = \mathbb{E}\left[\left((\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\right)^2\right] = p(p+2)$$

Proposición 4

Sea $\mathbf{x} \sim N_p(\mu, \Sigma)$, entonces la función característica de $\mathbf{x}$ está dada por,

$$\phi(\mathbf{t}) = \exp\left(i\mathbf{t}^T\mu - \frac{1}{2}\mathbf{t}^T\Sigma\mathbf{t}\right)$$

Propiedades

Proposición 5

Sea $\mathbf{x} \sim N_p(\mu, \Sigma)$ y considere la partición

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \end{pmatrix} \qquad \mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

donde $\mathbf{x}^{(1)}$ es de dimensión k y $\mathbf{x}^{(2)}$ es de dimensión $p-k$, entonces:

1. $\mathbf{x}^{(1)} \sim N_k\left(\mu^{(1)}, \Sigma_{11}\right)$
2. $\mathbf{x}^{(1)}$ y $\mathbf{x}^{(2)}$ son independientes si y solo si $\mathrm{Cov}\left(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}\right) = \Sigma_{12} = \mathbf{O}$
3. $\mathbf{x}^T\Sigma^{-1}\mathbf{x} \sim \chi_p^2\left(\mu^T\Sigma^{-1}\mu\right)$
4. $\mathbf{x}^{(2)} \,|\, \mathbf{x}^{(1)} \sim N_{p-k}\left(\mu^{(2)} + \Sigma_{21}\Sigma_{11}^{-1}[\mathbf{x}^{(1)} - \mu^{(1)}], \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right)$

Pruebas de normalidad

- Todas las distribuciones univariadas son normales
  - qqplot
  - histogramas
  - Pruebas de normalidad (e.g. Anderson-Darling, Shapiro-Wilk, Lilliefors, etc.)

- $(\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu) \sim \chi^2_p$
  - qqplot

- Prueba de Mardia (1970) basada en los coeficientes de asimetría y curtosis multivariados

- Otras pruebas (e.g. Henze-Zirkler (1990), Royston (1982))

- En **R:** librería MVN

Convergencia

Teorema 1 (Teorema Central del Límite)

Sean $\mathbf{x}_n = (x_{n1}, \ldots, x_{np})$ una colección de vectores aleatorios independientes e idénticamente distribuidos, con vector de medias μ y matriz (finita) de covarianza Σ. Entonces,

$$\sqrt{n}\,(\bar{\mathbf{x}} - \mu) \to N_p\left(\mathbf{0}_p, \Sigma\right)$$

Teorema 2 (Teorema de Cramér-Wold)

Para $\mathbf{x}_n = (x_{n1}, \ldots, x_{np})$ y $\mathbf{x} = (x_1, \ldots, x_p)$ dos vectores aleatorios y $\mathbf{t} \in \mathbb{R}^p$, entonces

$$\mathbf{x}_n \xrightarrow{d} \mathbf{x} \qquad \Leftrightarrow \qquad \sum_{i=1}^{p} t_i x_{ni} \xrightarrow{d} \sum_{i=1}^{p} t_i x_i$$

Distribución Wishart

Distribución Wishart

Definición 3

Sea $\mathbf{M}_{p\times p}$ una matriz simétrica de variables aleatorias, tal que $\mathbb{P}(\mathbf{M} > 0) = 1$, y sea $\Sigma_{p\times p}$ una matriz definida positiva. Si $n \in \mathbb{N}$, tal que $n \geq p$, entonces $\mathbf{M}_{p\times p}$ tiene una distribución Wishart, $\mathbf{M} \sim W_p(n, \Sigma)$, no singular con n grados de libertad si la función de densidad de los $\dfrac{p(p+1)}{2}$ distintos elementos de $\mathbf{M}_{p\times p}$ está dada por:

$$f(m_{11}, m_{12}, \ldots, m_{pp}) = c^{-1}\,|\,\mathbf{M}\,|^{(n-p-1)/2}\,\mathrm{etr}\left(-\frac{\Sigma^{-1}\mathbf{M}}{2}\right)$$

donde

- etr es el operador $\exp^{\text{trace}}$
- $c = 2^{\frac{np}{2}}\,|\,\Sigma\,|^{\frac{n}{2}}\,\Gamma_p\left(\dfrac{n}{2}\right)$ y $\Gamma_p(\,\cdot\,)$ la función gamma multivariada