ESTADÍSTICA MULTIVARIANTE

UGR, GRADO EN MATEMÁTICAS

Curso Académico 2023-2024

José Miguel Angulo Ibáñez (*jmangulo@ugr.es*)

Departamento de Estadística e Investigación Operativa
Universidad de Granada

- **TEMA 1. Distribución Normal Multivariante**

  Fundamentación probabilística (recordatorio)

- $(\Omega, \mathcal{F}, P)$, espacio de probabilidad
- $\mathbf{X} = (X_1, \ldots, X_p)'$, vector aleatorio; *i. e.*, aplicación medible

$$\begin{aligned} \mathbf{X} : (\Omega, \mathcal{F}) &\longrightarrow (\mathbb{R}^p, \mathcal{B}^p) \\ \omega &\longmapsto \mathbf{X}(\omega) = (X_1(\omega), \ldots, X_p(\omega))', \end{aligned}$$

$$\forall B \in \mathcal{B}^p, \;\; \mathbf{X}^{-1}(B) := \{\omega \in \Omega : \mathbf{X}(\omega) \in B\} \in \mathcal{F}$$

- $P_{\mathbf{X}}$, distribución de probabilidad inducida por $\mathbf{X}$ sobre $(\mathbb{R}^p, \mathcal{B}^p)$:

$$P_{\mathbf{X}}[B] := P[\mathbf{X}^{-1}(B)], \text{ para cada } B \in \mathcal{B}^p$$

Función de distribución

- Se define la *función de distribución* asociada a $P_{\mathbf{X}}$ como

$$F_{\mathbf{X}}(\mathbf{x}) = P_{\mathbf{X}}\left[X_1 \leq x_1, \ldots, X_p \leq x_p\right], \ \ \forall \mathbf{x} = (x_1, \ldots, x_p) \in \mathbb{R}^p$$

- Si existe $f_{\mathbf{X}}$ integrable (en el sentido de Lebesgue) t.q.

$$F_{\mathbf{X}}(\mathbf{x}) = \int_{-\infty}^{x_p} \cdots \int_{-\infty}^{x_1} f_{\mathbf{X}}(u_1, \ldots, u_p) du_1 \ldots du_p, \ \ \forall \mathbf{x} \in \mathbb{R}^p,$$

  $f_{\mathbf{X}}$ se denomina *función de densidad* de la distribución, estando definida salvo conjuntos de medida de Lebesgue nula. (Se dice entonces que la distribución de $\mathbf{X}$ es 'continua')

  Si $f_{\mathbf{X}}$ es una función continua, se puede escribir

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\partial^p}{\partial x_1 \ldots \partial x_p} F_{\mathbf{X}}(\mathbf{x}), \ \ \forall \mathbf{x} \in \mathbb{R}^p$$

  (en otro caso, se podrá escribir la expresión únicamente para los puntos de continuidad de $f_{\mathbf{X}}$) y $f_{\mathbf{X}}$ estará definida de forma unívoca bajo esa condición

Independencia

- En general, para un conjunto producto de la forma $B = B_1 \times \cdots \times B_p$, con $B_j \in \mathcal{B}$ ($j = 1, \ldots, p$), se tendrá que

$$P_{\mathbf{X}}[B] \neq P_{X_1}[B_1] \cdot \ldots \cdot P_{X_p}[B_p]$$

- Cuando se da la igualdad para todo conjunto producto $B = B_1 \times \cdots \times B \in \mathcal{B}$, se dice que las variables $X_1, \ldots, X_p$, componentes del vector aleatorio $\mathbf{X}$, son *independientes* entre sí (o *mutuamente independientes*).

- Equivalentemente, esto ocurre cuando

$$F_{\mathbf{X}}(\mathbf{x}) = F_{X_1}(x_1) \cdot \ldots \cdot F_{X_p}(x_p), \quad \forall \mathbf{x} = (x_1, \ldots, x_p)' \in \mathbb{R}^p,$$

  y en el caso continuo, si y solo si

$$f_{\mathbf{X}}(\mathbf{x}) = f_{X_1}(x_1) \cdot \ldots \cdot f_{X_p}(x_p), \quad \forall \mathbf{x} = (x_1, \ldots, x_p)' \in \mathbb{R}^p,$$

- De forma más general, se habla de independencia entre cualquier número finito de vectores aleatorios (por ejemplo, entre subvectores de un vector aleatorio dado), bajo condiciones de factorización similares en términos de los factores correspondientes a éstos.

Función característica

- Dado un vector aleatorio $\mathbf{X} = (X_1, \ldots, X_p)'$, se define su *función característica* como

$$\begin{aligned}\phi_{\mathbf{X}}(\mathbf{t}) &= E\left[e^{i\mathbf{t}'\mathbf{X}}\right] \\ &= \int_{\Omega} e^{i\mathbf{t}'\mathbf{X}(\omega)} P(d\omega) = \int_{\mathbb{R}^p} e^{i\mathbf{t}'\mathbf{x}} P_{\mathbf{X}}(d\mathbf{x}),\end{aligned}$$

  $\forall \mathbf{t} = (t_1, \ldots, t_p)' \in \mathbb{R}^p$

- EXISTENCIA Y UNICIDAD (TEOREMA): La función característica de un vector aleatorio siempre existe, para cualquier valor del argumento, y determina de forma única su distribución.

- La f.c. tiene gran utilidad como herramienta en la resolución de aspectos analíticos sobre distribuciones de vectores aleatorios.

Dos resultados importantes:

- **Resultado 1**: Las componentes del vector aleatorio $\mathbf{X} = (X_1, \ldots, X_p)'$ son independientes si y solo si su función característica se factoriza de la forma

$$\phi_{\mathbf{X}}(\mathbf{t}) = E\left[e^{i\mathbf{t}'\mathbf{X}}\right] = \prod_{k=1}^{p} E\left[e^{it_k X_k}\right] = \phi_{X_1}(t_1) \cdot \ldots \cdot \phi_{X_p}(t_p),$$

  $\forall \mathbf{t} = (t_1, \ldots, t_p)' \in \mathbb{R}^p$ (*i. e.*, como el producto de las funciones características de las distribuciones univariantes marginales)

- **Resultado 2**: Si las componentes del vector aleatorio $\mathbf{X} = (X_1, \ldots, X_p)'$ son independientes, entonces la función característica de la variable aleatoria

$$Y = \sum_{k=1}^{p} X_k$$

  se factoriza como

$$\phi_Y(t) = E\left[e^{itY}\right] = \prod_{k=1}^{p} E\left[e^{itX_k}\right] = \phi_{X_1}(t) \cdot \ldots \cdot \phi_{X_p}(t), \ \ \forall t \in \mathbb{R}$$

- Sea $\mathbf{X} = (X_1, \ldots, X_p)'$ un vector aleatorio p-dimensional.

  Sea $\mathbf{Y} = (Y_1, \ldots, Y_q)'$ un vector aleatorio q-dimensional definido como

  $$\mathbf{Y} = B\mathbf{X} + \mathbf{b}, \quad \text{con} \quad \begin{array}{l} B \text{ matriz } q \times p \text{ (cte.)} \\ \mathbf{b} \text{ vector } q \times 1 \text{ (cte.)} \end{array}$$

  Entonces, la f.c. de $\mathbf{Y}$ se obtiene, para cada $\mathbf{t} = (t_1, \ldots, t_q)' \in \mathbb{R}^q$, como

  $$\phi_{\mathbf{Y}}(\mathbf{t}) = e^{i\mathbf{t}'\mathbf{b}} \phi_{\mathbf{X}}(B'\mathbf{t})$$ (□ Probar)

  En particular, si el vector aleatorio $\mathbf{X}$ se particiona en dos subvectores como $\mathbf{X} = \begin{pmatrix} \mathbf{X}_{(1)} \\ \mathbf{X}_{(2)} \end{pmatrix}$, con $\mathbf{X}_{(1)}$ de dimensión $k \times 1$ y $\mathbf{X}_{(2)}$ de dimensión $(p-k) \times 1$, se tiene que

  $$\phi_{\mathbf{X}_{(1)}}(\mathbf{t}_{(1)}) = \phi_{\mathbf{X}} \begin{pmatrix} \mathbf{t}_{(1)} \\ \mathbf{0}_{(2)} \end{pmatrix},$$ (□ Probar)

  con $\mathbf{t}_{(1)} = (t_1, \ldots, t_k)' \in \mathbb{R}^k$ y $\mathbf{0}_{(2)} = (0, \ldots, 0)' \in \mathbb{R}^{p-k}$

Función característica y función de densidad (caso continuo)

- La función de densidad de un vector aleatorio $\mathbf{X} = (X_1, \ldots, X_p)'$ con distribución continua se puede obtener como la transformada de Fourier de la función característica:

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^p} \int_{\mathbb{R}^p} e^{-i\mathbf{t}'\mathbf{x}} \phi_{\mathbf{X}}(\mathbf{t}) d\mathbf{t}$$

Recordemos que, en este caso, la función característica también se escribe como

$$\phi_{\mathbf{X}}(\mathbf{t}) = \int_{\mathbb{R}^p} e^{i\mathbf{t}'\mathbf{x}} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}$$

(Es decir, $f_{\mathbf{X}}(\cdot)$ y $\phi_{\mathbf{X}}(\cdot)$ constituyen un 'par de transformadas de Fourier')

Observaciones finales

- Desde el punto de vista puramente estadístico, especialmente en el contexto de la Inferencia Estadística, interesa acerca de un vector aleatorio, $\mathbf{X}$, el conocimiento sobre su distribución, $P_{\mathbf{X}}$, puesto que los sucesos de interés vendrán expresados en términos de condiciones relativas a los valores que toma $\mathbf{X}$.
- Son importantes, en particular, familias de distribuciones multivariantes (*modelos*) que, de forma flexible, puedan dar una representación exacta o aproximada de la variabilidad o el comportamiento de poblaciones o fenómenos objeto de investigaciones aplicadas.
- En este contexto, la *distribución normal multivariante* (DNM) constituye un modelo central en el desarrollo y aplicación de la Estadística Multivariante.