

Teoría de la Probabilidad

Rafael González López

8 de febrero de 2018

Capítulo 1

Variables aleatorias multidimensionales

1.1. Introducción

Definición 1.1.1 Sean $(\Omega_1, \mathcal{A}_1), \dots, (\Omega_n, \mathcal{A}_n)$, se define el **σ -álgebra producto** como

$$\sigma(\mathcal{C}) \equiv \mathcal{A}_1 \times \dots \times \mathcal{A}_n$$

Donde $\mathcal{C} = \{A_1 \times \dots \times A_n : A_i \in \mathcal{A}_i \forall i\}$.

Definición 1.1.2 Definimos el **σ -álgebra de Borel en \mathbb{R}^n** como, sea el siguiente conjunto: $\mathcal{C} = \{B_1 \times \dots \times B_n : B_i \in \mathcal{B}(\mathbb{R})\}$,

$$\mathcal{B}(\mathbb{R}^n) = \sigma(\mathcal{C})$$

Definición 1.1.3 Sean $(\Omega, \mathcal{A}, \mathcal{P})$, $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ y sea $X : \Omega \rightarrow \mathbb{R}^n$. Se dirá que X es un **vector aleatorio real** si

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A} \quad \forall B \in \mathcal{B}(\mathbb{R}^n)$$

Proposición 1.1.4 Sean (Ω, \mathcal{A}) , $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ y $X : \Omega \rightarrow \mathbb{R}^n$. Entonces la función X es vector aleatorio real si y sólo si

$$X^{-1}((-\infty, x]) \in \mathcal{A} \quad \forall x \in \mathbb{R}^n$$

Demostración 1.1.5 Si X es vector aleatoria real, entonces $X^{-1}(B) \in \mathcal{A} \quad \forall B \in \mathcal{B}(\mathbb{R}^n)$, luego también será cierto para los conjuntos de la forma $(-\infty, x]$. Para el recíproco consideramos $\mathcal{C} = \{(-\infty, x] \mid x \in \mathbb{R}^n\}$. Entonces, $X^{-1}(\mathcal{C}) \in \mathcal{A}$ y $\sigma(X^{-1}(\mathcal{C})) \subset \mathcal{A}$. Como $\sigma(\mathcal{C})$ es precisamente $\mathcal{B}(\mathbb{R}^n)$, tenemos el resultado.

Proposición 1.1.6 Sean (Ω, \mathcal{A}) , $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ y $X : \Omega \rightarrow \mathbb{R}^n$. Entonces la función X es vector aleatorio real si y sólo si cada una de sus funciones componentes es variable aleatoria real.

Demostración 1.1.7 Supongamos que X es vector aleatorio. Consideremos pues

$$\begin{aligned} X_i^{-1}(-\infty, x_i] &= \{\omega \in \Omega : X_i(\omega) \in (-\infty, x_i]\} = \\ &= \{\omega \in \Omega : X_1(\omega) \in \mathbb{R}, \dots, X_i(\omega) \in (-\infty, x_i], \dots, X_n(\omega) \in \mathbb{R}\} = \\ &= \{\omega \in \Omega : X(\omega) \in (-\infty, b]\} = X^{-1}(-\infty, b] \in \mathcal{A} \end{aligned}$$

Donde $b = (\infty, \dots, x_i, \dots, \infty)$. Para el recíproco, supongamos que X_i son variable aleatoria real.

$$\begin{aligned} X^{-1}(-\infty, x] &= \{\omega \in \Omega : X(\omega) \in (-\infty, x]\} = \\ &= \{\omega \in \Omega : X(\omega) \in (-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_n]\} = \\ &= \{\omega \in \Omega : X_1(\omega) \in (-\infty, x_1], X_2(\omega) \in (-\infty, x_2], \dots, X_n(\omega) \in (-\infty, x_n]\} = \\ &= \bigcap_{i=1}^n \{\omega \in \Omega : X_i(\omega) \in (-\infty, x_i]\} = \bigcap_{i=1}^n X_i^{-1}(-\infty, x_i] \in \mathcal{A} \end{aligned}$$

Proposición 1.1.8 Sean (Ω, \mathcal{A}, P) , $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, X vector aleatorio real y $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$ una función medible-Borel. Entonces $Y = g(X)$ es vector aleatorio real.

Demostración 1.1.9 Sea $B \in \mathcal{B}(\mathbb{R}^n)$, por ser g medible-Borel se tiene que $B' = g^{-1}(B) \in \mathcal{B}(\mathbb{R}^n)$. Se tiene pues que

$$Y^{-1}(B) = (g \circ X)^{-1}(B) = X^{-1}(g^{-1}(B)) = X^{-1}(B') \in \mathcal{A}$$

Definición 1.1.10 Sean (Ω, \mathcal{A}, P) , $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, X vector aleatorio real, definimos la función **probabilidad inducida** P_X como

$$P_X : \mathcal{B}(\mathbb{R}^n) \rightarrow \mathbb{R}_{\geq 0}$$

$$P_X(B) = P[X^{-1}(B)]$$

Nota 1.1.11 Se tienen la siguientes igualdades

$$P_X(B) = P_X(B_1 \times \dots \times B_n) = P(\{\omega \in \Omega : X(\omega) \in B_1 \times \dots \times B_n\}) =$$

$$= P(X_1(\omega) \in B_1, \dots, X_n(\omega) \in B_n) = P[x_1 \in B_1, \dots, x_n \in B_n]$$

Definición 1.1.12 Se define la **distribución marginal** como la ley de probabilidad de un cierto conjunto de variables aleatorias haciendo caso omiso del resto de variables. Ejemplo:

$$P_X(B) = P[x_1 \in B_1, x_2 \in \mathbb{R}, \dots, x_n \in \mathbb{R}] = P_{X_1}(B_1)$$

$$P_X(B) = P[x_1 \in B_1, x_2 \in B_2, x_3 \in \mathbb{R}, \dots, x_n \in \mathbb{R}] = P_{X_1, X_2}(B_1 \times B_2)$$

Definición 1.1.13 Se define la **función de distribución** como la única función que cumple que

$$F : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$$

$$F(x) = P_X(-\infty, x] = P_X[(-\infty, x_1] \times \dots \times (-\infty, x_n]] =$$

$$= P(X^{-1}(-\infty, x]) = P[\{\omega \in \Omega : X(\omega) \in (-\infty, x]\}] =$$

$$= P[\{\omega \in \Omega : X_1(\omega) \in (-\infty, x_1], \dots, X_n(\omega) \in (-\infty, x_n]\}] =$$

$$= P[X_1(\omega) \leq x_1, \dots, X_n(\omega) \leq x_n] = P[X_1 \leq x_1, \dots, X_n \leq x_n]$$

Nota 1.1.14 Sean $a = (a_1, \dots, a_n)$ y $b = (b_1, \dots, b_n)$. Decimos que $a \leq b$ si $a_i \leq b_i \forall i$.

Definición 1.1.15 Sea $g : \mathbb{R}^n \rightarrow \mathbb{R}$, definimos el **operador diferencia** como

$$b_i \geq a_i \quad \Delta_{b_i, a_i} g(x_1, \dots, x_n) = g(x_1, \dots, b_i, \dots, x_n) - g(x_1, \dots, a_i, \dots, x_n)$$

$$h_i \geq 0 \quad \Delta_{x_i + h_i, x_i} g(x_1, \dots, x_n) = g(x_1, \dots, x_i - 1, x_i + h_i, x_{i+1}, \dots, x_n) - g(x_1, \dots, x_n) = \Delta_{h_i}$$

$$\Delta_h g(x) = \Delta_{h_1, \dots, h_n} g(x_1, \dots, x_n) = \Delta_{h_1} \dots \Delta_{h_n} g(x_1, \dots, x_n)$$

Proposición 1.1.16 Sea (Ω, \mathcal{A}, P) , X vector aleatorio n -dimensional y F función de distribución. Se cumplen

1. $\Delta_h F = \Delta_{h_1, \dots, h_n} F \geq 0$
2. $\lim_{x_i \rightarrow +\infty \forall i} F(x_1, \dots, x_n) = 1$
3. $\lim_{x_i \rightarrow -\infty} F(x_1, \dots, x_n) = 0 \forall i$
4. F es continua por la derecha.

Demostración 1.1.17 Demostramos la proposición para el caso bidimensional. En el primer caso consideramos:

$$\begin{aligned}
\Delta_h F(x, y) &= \Delta_{h_1} \Delta_{h_2} F(x, y) = \Delta_{h_1} [F(x, y + h_2) - F(x, y)] = \\
&= (F(x + h_1, y + h_2) - F(x, y + h_2)) - (F(x + h_1, y) - F(x, y)) = \\
&= P[X \leq x + h_1, Y \leq y + h_2] - P[X \leq x, Y \leq y + h_2] - \\
&\quad - P[X \leq x + h_1, Y \leq y] + P[X \leq x, Y \leq y] = \\
&= (P_X[(-\infty, x + h_1] \times (-\infty, y + h_2]] - P_X[(-\infty, x] \times (-\infty, y + h_2])) - \\
&\quad - (P_X[(-\infty, x + h_1] \times (-\infty, y]] - P_X[(-\infty, x] \times (-\infty, y])) = \\
&= P_X[(x, x + h_1] \times (-\infty, y + h_2]] - P_X[(x, x + h_1] \times (-\infty, y]] = \\
&= P_X[(x, x + h_1] \times (y, y + h_2]] = P[x < X \leq x + h_1, y < Y \leq y + h_2] \geq 0
\end{aligned}$$

Para demostrar la segunda propiedad, consideramos $F(x_n, y_n)$ con $(x_n, y_n) \rightarrow (+\infty, +\infty)$

$$F(x_n, y_n) = P_X[(-\infty, x_n] \times (-\infty, y_n]] = P_X(R_n)$$

Como R_n es una sucesión creciente $\lim R_n = \bigcup_n R_n = \mathbb{R}^2$.

$$\lim_{\substack{x_n \rightarrow +\infty \\ y_n \rightarrow +\infty}} F(x_n, y_n) = \lim_{\substack{x_n \rightarrow +\infty \\ y_n \rightarrow +\infty}} P_X(R_n) = P_X \left(\lim_{\substack{x_n \rightarrow +\infty \\ y_n \rightarrow +\infty}} R_n \right) = P_X(\mathbb{R}^2) = 1$$

En el tercer caso, probaremos para la primera variable. El resto de casos será análogo. Considero ahora el rectángulo $R_n = (-\infty, x_n] \times (-\infty, y]$ con $x_n \downarrow -\infty$, por lo que R_n es una sucesión decreciente de forma que

$$\lim R_n = \bigcap_n R_n = \emptyset$$

$$\lim_{x_n \rightarrow -\infty} F(x_n, y) = \lim_{x_n \rightarrow -\infty} P_X(R_n) = P_X \left(\lim_{x_n \rightarrow -\infty} R_n \right) = P_X(\emptyset) = 0$$

Para finalizar, consideramos $x_n \downarrow a$ e $y_n \downarrow b$ y el rectángulo $R_n = (-\infty, x_n] \times (-\infty, y_n]$

$$\lim_{\substack{x_n \rightarrow a \\ y_n \rightarrow b}} F(x_n, y_n) = \lim_{\substack{x_n \rightarrow a \\ y_n \rightarrow b}} P_X(R_n) = P_X \left(\lim_{\substack{x_n \rightarrow a \\ y_n \rightarrow b}} R_n \right) = P_X[(-\infty, a] \times (-\infty, b]] = F(a, b)$$

Definición 1.1.18 *Cualquier función que cumpla las cuatro propiedades anteriores definida de \mathbb{R}^n en $\mathbb{R}_{\geq 0}$ se considerará función de distribución.*

Propiedades 1.1.19 *Veamos algunas propiedades sobre la continuidad de las funciones de distribución*

1. La condición necesaria y suficiente para que $F(x)$ sea continua en x es que se cumpla que $P[X = x] = 0$.
2. Sean F_1 y F_2 dos funciones de distribución y \mathcal{C} un conjunto denso en \mathbb{R}^n . Si $F_1(x) = F_2(x) \forall x \in \mathcal{C}$, se tiene entonces que $F_1(x) = F_2(x) \forall x \in \mathbb{R}^n$.
3. Una función de distribución tiene a lo sumo un conjunto numerable de discontinuidades.
4. Sea F_1 y F_2 funciones de distribución con sendos conjuntos de puntos de continuidad \mathcal{C}_1 y \mathcal{C}_2 . Si para todo $x \in \mathcal{C}_1 \cap \mathcal{C}_2$ $F_1(x) = F_2(x)$, entonces $F_1 \equiv F_2$.

Demostración 1.1.20 *Veamos las demostraciones*

1. Sea $x \in \mathbb{R}^n$

$$\begin{aligned} \lim_{x \rightarrow x^+} F(x) &= F(x^+) = F(x) = P_X(-\infty, x] = P_X[(-\infty, x) \cup \{x\}] = \\ &= P_X(-\infty, x) + P_X(\{x\}) = F(x^-) + P_X(\{x\}) \end{aligned}$$

Como $P_X(\{x\}) = P(\{\omega \in \Omega \mid X(\omega) = x\}) = P[X = x]$ y los límites laterales deben coincidir, $P[X = x]$ tiene que ser nulo.

2. Sea $x \in \mathbb{R}^n$, existe $\{x_n\} \subset \mathcal{C}$ tal que $\{x_n\} \rightarrow x$.

$$F_1(x) = \lim_{x_n \rightarrow x^+} F_1(x_n) = \lim_{x_n \rightarrow x^+} F_2(x_n) = F_2(x)$$

3. Sea $D = \{x \in \mathbb{R} \mid F(x) - F(x^-) = p(x) > 0\}$ y $D_n = \{x \in \mathbb{R}^n \mid \frac{1}{n+1} < p(x) \leq \frac{1}{n}\}$. Naturalmente D es la unión numerable de D_n . Sean $x_1, \dots, x_k \in D_n$ con $a \leq x_1^- < x_1 \dots < x_k \leq b$.

$$\begin{aligned} \frac{k}{n+1} &< \sum_{j=1}^k p(x_j) = \sum_{j=1}^k [F(x_j) - F(x_j^-)] \leq \\ &\leq \sum_{j=1}^k [F(x_j) - F(x_j^-)] + \sum_{j=1}^k [F(x_j^-) - F(x_{j-1})] = \\ &= \sum_{j=1}^k [F(x_j) - F(x_{j-1})] = F(x_k) - F(x_1) \leq F(b) - F(a) \leq 1 \end{aligned}$$

Por tanto, $k \leq n$ y D_n admite a lo más n elementos. Resulta pues que D es numerable al ser unión de conjuntos finitos y/o numerables.

4. Se tiene a partir de la igualdad

$$\begin{aligned} \mathbb{R}^n - (\mathcal{C}_1 \cap \mathcal{C}_2) &= \mathbb{R}^n \cap \overline{(\mathcal{C}_1 \cap \mathcal{C}_2)} = \mathbb{R}^n \cap (\overline{\mathcal{C}_1} \cup \overline{\mathcal{C}_2}) = \\ &= (\mathbb{R}^n \cap \overline{\mathcal{C}_1}) \cup (\mathbb{R}^n \cap \overline{\mathcal{C}_2}) = D_{F_1} \cup D_{F_2} \end{aligned}$$

La unión de los conjuntos de discontinuidades es numerable, luego $\mathcal{C}_1 \cap \mathcal{C}_2$ es denso en \mathbb{R} .

Nota 1.1.21 Puede ocurrir que un subespacio tenga probabilidad nula. Por ejemplo, que en \mathbb{R}^2 una curva o una superficie en \mathbb{R}^3 tengan probabilidad nula.

Teorema 1.1.22 Sea F una función de distribución, puede descomponerse de la siguiente forma

$$F(x) = \alpha_1 F_d + \alpha_2 F_{ac} + \alpha_3 F_s$$

Con $0 \leq \alpha_i \leq 1$ y $\alpha_1 + \alpha_2 + \alpha_3 = 1$.

Nota 1.1.23 La distribución marginal de una variable X_1 se consigue

$$F_{X_1}(x_1) = P[X_1 \leq x_1] = P_{X_1}(-\infty, x_1] = P_X[(-\infty, x_1] \times \mathbb{R} \times \dots \times \mathbb{R}]$$

1.2. Modelos de distribución

1.2.1. Modelo discreto

Definición 1.2.1 Sea (Ω, \mathcal{A}, P) , un vector aleatorio X n -dimensional y F su función de distribución, se dirá que un vector aleatorio es **discreto** si su función de distribución es una función discreta. Es decir que si D es el conjunto de los puntos de salto de la función de distribución, entonces se verifica que $P(D) = 1$.

$$P[X = x_i] = P[X_1 = x_{i1}, \dots, X_n = x_{in}] > 0$$

Con $x_i \in D$. De la definición se sigue la función ha de verificar que $\sum_{x_i} P[X = x_i] = 1$.

Proposición 1.2.2 En el caso de las distribuciones discretas podemos escribir la función de distribución de la siguiente manera

$$F(y_1, \dots, y_n) = P[X_1 \leq y_1, \dots, X_n \leq y_n] = \sum_{x_i \leq y_i} P[X_1 = x_1, \dots, X_n = x_n]$$

De la misma forma, podemos escribir la función de probabilidad marginal (para más de una variable es análogo)

$$P_{x_1}[X_1 = x_{i1}] = \sum_{x_{i2}, \dots, x_{in}} P[X_1 = x_{i1}, \dots, X_n = x_{in}]$$

1.2.2. Modelo absolutamente continuo

Definición 1.2.3 Sea (Ω, \mathcal{A}, P) , un vector aleatorio X n -dimensional y F su función de distribución, se dirá que un vector aleatorio es **absolutamente continuo** si su función de distribución es una función absolutamente continua. Es decir, existe una función $f : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$ que verifica

1. Existe $f(x) \geq 0 \forall x \in \mathbb{R}^n$
2. $\int_{\mathbb{R}^n} f(x) dx = 1$
3. $F(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f(t_1, \dots, t_n) dt_1 \dots dt_n$.

Proposición 1.2.4 En el caso de las distribuciones discretas podemos escribir la función de probabilidad marginal de la siguiente manera

$$f_1(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_2 \dots, dx_n$$

Obviamente, para obtener su función de distribución se obtiene integrando la función de densidad.

1.2.3. Distribuciones condicionadas

Definición 1.2.5 Sea un espacio probabilístico (Ω, \mathcal{A}, P) , y un vector aleatorio bidimensional $(X, Y)'$ y $A \in \mathcal{A}$ con $P(A) > 0$. Se define la **probabilidad condicionada** de X a A de B como

$$P_{X|A}(B) = \frac{P[X^{-1}(B) \cap A]}{P(A)}$$

Tomando B como los conjuntos de la forma $(-\infty, x]$ obtenemos la **distribución condicionada**

$$F_{X|A}(x) = \frac{P[X(\omega) \leq x, x \in A]}{P(A)} = P_{X|A}((-\infty, x])$$

Condicionadas discretas

Definición 1.2.6 Sea $D = \{x_n\}$ de forma que $P(D) = 1$, podemos definir la función de probabilidad condicionada discreta

$$P[X = x|Y = y] = \frac{P[X = x, Y = y]}{P[Y = y]} \geq 0$$

Obviamente se cumple que

$$\sum_x P[X = x|Y = y] = \sum_x \frac{P[X = x, Y = y]}{P[Y = y]} = \frac{1}{P[Y = y]} \sum_x P[X = x, Y = y] = 1$$

Además, en el caso de una v.a. discreta multidimensional, se tiene que, para cualquier conjunto de índices

$$P[X_{i_1} = x_{i_1}, \dots, X_{i_k} = x_{i_k} | X_{i_{k+1}} = x_{i_{k+1}}, \dots, X_{i_n} = x_{i_n}] = \frac{P[X_{i_1} = x_{i_1}, \dots, X_{i_k} = x_{i_k}]}{P[X_{i_{k+1}} = x_{i_{k+1}}, \dots, X_{i_n} = x_{i_n}]}$$

Proposición 1.2.7 Podemos escribir la función de distribución de la siguiente manera

$$F_{X|Y}(z) = P[X \leq z | Y = y] = \sum_{x \leq z} \frac{P[X = x, Y = y]}{P[Y = y]}$$

Condicionadas absolutamente continuas

Definición 1.2.8 Como la función de probabilidad de una variable continua se anula en los conjuntos unitarios, buscamos un intervalo lo más pequeño posible donde podamos definirla

$$\begin{aligned} F_{X|Y}(x) &= \lim_{h \rightarrow 0} \frac{P[X \leq x, y - h < Y \leq y + h]}{P[y - h < Y \leq y + h]} = \\ &= \lim_{h \rightarrow 0} \frac{P[X \leq x, Y \leq y + h] - P[X \leq x, Y \leq y - h]}{P[Y \leq y + h] - P[Y \leq y - h]} = \\ &= \lim_{h \rightarrow 0} \frac{F(x, y + h) - F(x, y - h)}{F_Y(y + h) - F_Y(y - h)} = \lim_{h \rightarrow 0} \frac{\frac{F(x, y + h) - F(x, y - h)}{2h}}{\frac{F_Y(y + h) - F_Y(y - h)}{2h}} = \frac{\frac{\partial}{\partial y} F(x, y)}{\frac{\partial}{\partial y} F_Y(y)} = \\ &= \frac{\frac{\partial}{\partial y} \int_{-\infty}^x \int_{-\infty}^y f(t, s) dt ds}{f(y)} = \frac{\int_{-\infty}^x \frac{\partial}{\partial y} \int_{-\infty}^y f(t, s) dt ds}{f(y)} = \\ &= \frac{\int_{-\infty}^x f(t, y) dt}{f(y)} = \int_{-\infty}^x \frac{f(t, y)}{f(y)} dt = \int_{-\infty}^x f(x|y) \end{aligned}$$

Que tiende a 1 cuando $x \rightarrow \infty$. Si tenemos $f(x_1, \dots, x_k | x_{k+1}, \dots, x_n)$ y queremos la marginal de esa función con respecto a x_1 :

$$f(x_1, \dots, x_k | x_{k+1}, \dots, x_n) = \frac{f(x_1, \dots, x_k)}{f(x_{k+1}, \dots, x_n)}$$

1.3. Independencia

Definición 1.3.1 Sea un espacio probabilístico (Ω, \mathcal{A}, P) y sean X_1, \dots, X_n variables aleatorias reales definidas sobre ese espacio. Se dirán que X_1, \dots, X_n son **mutuamente independientes** o simplemente independientes si verifican que $\forall B_j \in \mathcal{B}(\mathbb{R})$ para todo $0 \leq j \leq k \leq n$:

$$P[X_{i_1}^{-1}(B_1), \dots, X_{i_k}^{-1}(B_k)] = \prod_{j=1}^k P[X_{i_j}^{-1}(B_j)] = P[x_{i_1} \in B_1, \dots, x_{i_k} \in B_k]$$

Proposición 1.3.2 Sea un espacio probabilístico (Ω, \mathcal{A}, P) y sean X_1, \dots, X_n variables aleatorias reales definidas sobre ese espacio. La condición necesaria y suficiente para que dichas variables sean independientes es que

$$P[X_1^{-1}(B_1), \dots, X_n^{-1}(B_n)] = \prod_{j=1}^n P[X_j^{-1}(B_j)]$$

Demostración 1.3.3 Es claro que si se cumple que las variables son independientes, se debe cumplir la igualdad anterior. Pasemos a ver que es condición suficiente. Si es cierto que

$$P[X_1^{-1}(B_1), \dots, X_n^{-1}(B_n)] = \prod_{j=1}^n P[X_j^{-1}(B_j)]$$

Tomando $B_j = \mathbb{R}$ si $j \geq k+1$, pues

$$\begin{aligned} P[X_{i_1}^{-1}(B_1), \dots, X_{i_k}^{-1}(B_k)] &= P[X_{i_1}^{-1}(B_1), \dots, X_{i_k}^{-1}(B_k), X_{i_{k+1}}^{-1}(\mathbb{R}), \dots, X_{i_n}^{-1}(\mathbb{R})] = \\ &= \prod_{j=1}^k P[X_{i_j}^{-1}(B_j)] \cdot \prod_{j=k+1}^n P[X_{i_j}^{-1}(\mathbb{R})] = \\ &= \prod_{j=1}^k P[X_{i_j}^{-1}(B_j)] \cdot \prod_{j=k+1}^n P[\Omega] = \prod_{j=1}^k P[X_{i_j}^{-1}(B_j)] \end{aligned}$$

Proposición 1.3.4 Sea un espacio probabilístico (Ω, \mathcal{A}, P) , sean X_1, \dots, X_n variables aleatorias reales independientes definidas sobre ese espacio y sean g_1, \dots, g_n funciones medibles-Borel. Entonces $g_1(X_1), \dots, g_n(X_n)$ son independientes.

Proposición 1.3.5 Sea un espacio probabilístico (Ω, \mathcal{A}, P) , sean X_1, \dots, X_n variables aleatorias reales definidas sobre ese espacio. La condición necesaria y suficiente para que dichas variables sean mutuamente independientes es que

$$F(x_1, \dots, x_n) = F_1(x_1) \dots F_n(x_n) \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n$$

Demostración 1.3.6 Para ver que es condición necesaria, basta aplicar la independencia a los conjuntos $B_i = (-\infty, x_i] \forall i = 1, \dots, n$.

Proposición 1.3.7 Sea un espacio probabilístico (Ω, \mathcal{A}, P) , sean X_1, \dots, X_n variables aleatorias reales discretas definidas sobre ese espacio. La condición necesaria y suficiente para que dichas variables sean mutuamente independientes es que

$$P[X_1 = x_1, \dots, X_n = x_n] = \prod_{j=1}^n P[X_j = x_j]$$

Demostración 1.3.8 Supongamos que X_1, \dots, X_n son independientes. En este caso la implicación es trivial a partir de la definición de independencia. Veamos el recíproco. Si se verifica que F es independiente para puntos, entonces:

$$\begin{aligned} F(x_1, \dots, x_n) &= \sum_{y_i \leq x_i} P[X_1 = y_1, \dots, X_n = y_n] = \sum_{y_i \leq x_i} P[X_1 = y_1] \dots P[X_n = y_n] = \\ &= \sum_{y_i \leq x_i} P[X_1 = y_1] \dots \sum_{y_n \leq x_n} P[X_n = y_n] = \prod_{i=1}^n P[X_i \leq x_i] \end{aligned}$$

Proposición 1.3.9 Sea un espacio probabilístico (Ω, \mathcal{A}, P) , sean X_1, \dots, X_n variables aleatorias reales continuas definidas sobre ese espacio. La condición necesaria y suficiente para que dichas variables sean mutuamente independientes es que

$$f(x_1, \dots, x_n) = f_1(x_1) \dots f_n(x_n)$$

Demostración 1.3.10 Aplicando en el Teorema de Radon-Nikodym

$$\begin{aligned} F(x_1, \dots, x_n) &= F_1(x_1) \dots F_n(x_n) \iff \\ \int \dots \int_{-\infty}^x f(t) dt &= \int_{-\infty}^{x_1} f_1(t_1) dt_1 \dots \int_{-\infty}^{x_n} f_n(t_n) dt_n = \int \dots \int_{-\infty}^x f_1(t_1) \dots f_n(t_n) dt \\ &\iff f(t_1, \dots, t_n) = f_1(t_1) \dots f_n(t_n) \end{aligned}$$

Capítulo 2

Características de vectores aleatorios

Definición 2.0.1 *El vector de medias de un vector aleatorio X es aquel cuyas componentes son las esperanzas de cada componente de X , siempre que estas existen.*

$$E[X] = \begin{pmatrix} E[X_1] \\ \vdots \\ E[X_n] \end{pmatrix}$$

Proposición 2.0.2 *Sea un espacio probabilístico (Ω, \mathcal{A}, P) , X un vector aleatorio n dimensional, $g : \mathbb{R}^n \rightarrow \mathbb{R}$ una función medible-Borel. Se verifica entonces que*

$$E[g(X)] = E[g(x_1, \dots, x_n)] = \begin{cases} \sum_{x_1, \dots, x_n} g(x_1, \dots, x_n) P[X_1 = x_1, \dots, X_n = x_n] \\ \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} g(x) f(x) dx \end{cases}$$

Definición 2.0.3 *Sea un espacio probabilístico (Ω, \mathcal{A}, P) , se definen los **momentos respecto al origen** de un vector aleatorio n dimensional X de orden $r \in \mathbb{R}^n$*

$$\mu_{r_1, \dots, r_n} = E[X^r] = E[X_1^{r_1} \dots X_n^{r_n}]$$

Si $r = e_i$ entonces $\mu_r = E[X_i] = \mu_i$.

Proposición 2.0.4 Sea un espacio probabilístico (Ω, \mathcal{A}, P) , X un vector aleatorio n dimensional cuyas variables componentes son independientes. En tal caso, la esperanza del producto es el producto de las esperanzas.

$$E[X_1 \dots X_n] = E[X_1] \dots E[X_n]$$

Demostración 2.0.5 La demostración se sigue a partir de la definición de independencia

$$\begin{aligned} E[X_1 \dots X_n] &= \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} x_1 \dots x_n f(x) dx = \\ &= \int_{-\infty}^{+\infty} x_1 f(x_1) dx_1 \dots \int_{-\infty}^{+\infty} x_n f(x_n) dx_n = E[X_1] \dots E[X_n] \end{aligned}$$

Proposición 2.0.6 (Desigualdad de Cauchy-Schwarz) Sea un espacio probabilístico (Ω, \mathcal{A}, P) , $(X, Y)'$ un vector aleatorio entonces

$$E[XY]^2 \leq E[X^2] \cdot E[Y^2]$$

Demostración 2.0.7

$$\begin{aligned} 0 &\leq E[(\alpha X + \beta Y)^2] = E[\alpha^2 X^2 + \beta^2 Y^2 + 2\alpha\beta XY] = \alpha^2 E[X^2] + \beta^2 E[Y^2] + 2\alpha\beta E[XY] = \\ &= \frac{E[XY]^2}{E[X^2]^2} E[X^2] + E[Y^2] - 2 \frac{E[XY]^2}{E[X^2]} = \frac{E[XY]^2}{E[X^2]} + E[Y^2] - 2 \frac{E[XY]^2}{E[X^2]} \end{aligned}$$

Hemos tomado $\beta = 1$ y $\alpha = -\frac{E[XY]}{E[X^2]}$. Despejando se obtiene.

Definición 2.0.8 Sea un espacio probabilístico (Ω, \mathcal{A}, P) , se definen los **momentos centrales** de un vector aleatorio n dimensional X de orden $r \in \mathbb{R}^n$

$$\alpha_{r_1, \dots, r_n} = E[(X - \mu)^r] = E[(X_1 - \mu_1)^{r_1} \dots (X_n - \mu_n)^{r_n}]$$

Si $r = 2e_i$ entonces $\alpha_r = E[(X_i - \mu_i)^2] = V[X_i]$.

Definición 2.0.9 La **covarianza** es una medida de la relación lineal entre dos variables

$$Cov(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X] \cdot E[Y]$$

Cuando $\sigma_{ij} = 0$ se dice que las variables están **incorreladas**. Si $r = e_i + e_j$ entonces $\alpha_r = Cov(X_i, X_j)$.

Definición 2.0.10 La *correlación o coeficiente de correlación lineal* es una medida adimensional de la relación lineal entre dos variables

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j} \quad \rho_{ij}^2 = \frac{\sigma_{ij}^2}{\sigma_i^2 \sigma_j^2}$$

Por la desigualdad CBS, $\rho^2 \in [0, 1]$, $\rho \in [-1, 1]$ y tiene el mismo signo que la covarianza.

Definición 2.0.11 Sea un espacio probabilístico (Ω, \mathcal{A}, P) , X un vector aleatorio n dimensional. Se define la **varianza** de X como

$$Var(X) = E[(X - \mu)(X - \mu)'] = \begin{pmatrix} V[X_1] & Cov(X_1, X_2) & \cdots & Cov(X_1, X_n) \\ Cov(X_1, X_2) & V[X_2] & \cdots & Cov(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_1, X_n) & Cov(X_2, X_n) & \cdots & V[X_n] \end{pmatrix}$$

También se suele denominar matriz de varianza y covarianza y representarse por Σ . Si todas las componentes fueran independientes, la matriz sería diagonal.

Proposición 2.0.12 Sea un espacio probabilístico (Ω, \mathcal{A}, P) , X un vector aleatorio n dimensional. Entonces la matriz de varianza y covarianza cumple que es una matriz simétrica y además semidefinida positiva.

Demostración 2.0.13 La simetría se tiene por la propia construcción de la matriz. Veamos cuál es la forma cuadrática que genera la matriz. Consideramos $X = (X_1, \dots, X_n)'$ y $\mu = (\mu_1, \dots, \mu_n)'$, $c = (e_1, \dots, e_n)'$, $Y = c'X$.

$$\begin{aligned} 0 \leq V[Y] &= E[(c'X - c'\mu)^2] = E \left[\left(\sum_{i=1}^n c_i (X_i - \mu_i) \right)^2 \right] = \\ &= E \left[\sum_{i,j} c_i c_j (X_i - \mu_i)(X_j - \mu_j) \right] = \sum_{i,j} c_i c_j E[(X_i - \mu_i)(X_j - \mu_j)] = \sum_{i,j} c_i c_j \sigma_{i,j} = c' \Sigma c \end{aligned}$$

2.1. Función generatriz de momentos

Definición 2.1.1 Sea un espacio probabilístico (Ω, \mathcal{A}, P) , X un vector aleatorio n dimensional y sea $t = (t_1, \dots, t_n)'$. Se define entonces la esperanza como

$$M_X : \mathbb{R}^n \rightarrow \mathbb{R}$$

$$M_X(t) \mapsto E[e^{t'X}]$$

Siempre que esté bien definida en un entorno del origen.

Proposición 2.1.2 Sea un espacio probabilístico (Ω, \mathcal{A}, P) , X un vector aleatorio n dimensional. Si su función generatriz existe, esta determina de forma biunívoca a la función de distribución del vector aleatorio.

Corolario 2.1.3 Consideremos el espacio probabilístico (Ω, \mathcal{A}, P) y el vector bidimensional $Z = (X, Y)'$ para el cual suponemos que existe su función generatriz de momentos $M_Z(t)$. Entonces esta determina de forma única las distribuciones marginales de las componentes del vector.

Demostración 2.1.4 En este caso, la función generatriz de momentos puede escribirse como

$$M_Z(t) = E[e^{t'Z}] = E[e^{t_1X + t_2Y}] = \sum_{x,y} e^{t_1x + t_2y} P[X = x, Y = y] = M_Z(t_1, t_2)$$

Si tomamos el vector $(t_1, 0)'$ se tiene que

$$\begin{aligned} M_Z(t_1, 0) &= \sum_{x,y} e^{t_1x} P[X = x, Y = y] = \sum_x e^{t_1x} \sum_y P[X = x, Y = y] = \\ &= \sum_x e^{t_1x} P[X = x] = M_X(t_1) \end{aligned}$$

Como la función generatriz determina de forma unívoca a la función de distribución, basta trabajar análogamente para la variable Y para tener el resultado.

Corolario 2.1.5 Consideremos el espacio probabilístico (Ω, \mathcal{A}, P) y el vector n -dimensional $X = (X_1, \dots, X_n)'$ para el cual suponemos que existe su función generatriz de momentos $M_X(t)$. Entonces la condición necesaria y suficiente para que las componentes del vector sean independientes es que la función generatriz de momentos sea el producto las funciones generatrices de momento de las marginales.

Demostración 2.1.6 Supongamos que X_1, \dots, X_n son independientes. Entonces

$$M_X(t) = E[e^{t'X}] = E[e^{t_1 X_1 + \dots + t_n X_n}] = E[e^{t_1 X_1} \dots e^{t_n X_n}] = \prod_{i=1}^n E[e^{t_i X_i}] = \prod_{i=1}^n M_{X_i}(t_i)$$

Corolario 2.1.7 Consideremos el espacio probabilístico (Ω, \mathcal{A}, P) , sean X_1, \dots, X_n variables aleatorias reales independientes para las cuales existe su función generatriz de momentos. Sea también $T = \sum_{i=1}^n X_i$, entonces se verifica que

$$M_T(t) = \prod_{i=1}^n M_{X_i}(t)$$

Demostración 2.1.8

$$M_T(t) = E[e^{t \sum_i X_i}] = E\left[\prod_{i=1}^n e^{t X_i}\right] = \prod_{i=1}^n E[e^{t X_i}] = \prod_{i=1}^n M_{X_i}(t)$$

Ejemplo 2.1.9 Consideremos el espacio probabilístico (Ω, \mathcal{A}, P) , sean $X_1, \dots, X_n \in Be(p)$, entonces $M_{X_i}(t) = pe^t + q \forall i = 1, \dots, n$. Sea $T = \sum_{i=1}^n X_i$

$$M_T(t) = (pe^t + q)^n \rightarrow Bi(n, p)$$

Ejemplo 2.1.10 Si $X_1 \in Bi(n_1, p)$ y $X_2 \in Bi(n_2, p)$ entonces $L = X_1 + X_2$ cumple que $M_L(t) = (pe^t + q)^{n_1+n_2}$, $Bi(n_1 + n_2, p)$. Se suele decir que la distribución binomial es reproductiva respecto al parámetro n .

Nota 2.1.11 La distribución normal es reproductiva respecto de la varianza y la media. La distribución de Poisson también es reproductiva.

Corolario 2.1.12 Consideremos el espacio probabilístico (Ω, \mathcal{A}, P) y el vector n -dimensional $X = (X_1, \dots, X_n)'$ para el cual suponemos que existe su función generatriz de momentos $M_X(t)$ entonces

$$E[X_1^{r_1} \dots X_n^{r_n}] = \frac{\partial^{r_1 + \dots + r_n} M_X(t)}{\partial t_1^{r_1} \dots \partial t_n^{r_n}}$$

2.2. Estadísticos de orden

Definición 2.2.1 Consideremos el espacio probabilístico (Ω, \mathcal{A}, P) , sean X_1, \dots, X_n variables aleatorias reales de las que se toman muestras x_1, \dots, x_n , las cuales se ordenan de menor a mayor, quedando $x_{(1)} \leq \dots \leq x_{(n)}$. Se define el estadístico de orden k $X_{(k)}$ como la variable aleatoria que toma el valor $x_{(k)}$ en cada posible sucesión de posibles valores de las variables aleatorias. Se verifica obviamente que

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$$

Proposición 2.2.2 Consideremos el espacio probabilístico (Ω, \mathcal{A}, P) , sean X_1, \dots, X_n variables aleatorias reales independientes de forma que todas tienen la misma distribución $F(x) = P[X \leq x]$. Calculemos $F_{i:n}(x)$

$$\begin{aligned} F_{i:n}(x) &= P[X_{i:n} \leq x] = P[\text{al menos } i \text{ menores o iguales que } x] = \\ &= \sum_{k=i}^n \binom{n}{k} F(x)^k (1 - F(x))^{n-k} \end{aligned}$$

Donde la probabilidad de que i de ellos sean menores o iguales que x y que $n-i$ no lo sean es $F(x)^i (1 - F(x))^{n-i}$

Proposición 2.2.3 Consideremos el espacio probabilístico (Ω, \mathcal{A}, P) , sean X_1, \dots, X_n variables aleatorias reales independientes absolutamente continuas de forma que todas tienen la misma distribución $F(x) = P[X \leq x]$. Calculemos $f_{i:n}(x)$

$$\begin{aligned} f_{i:n}(x) &= \lim_{h \rightarrow 0} \frac{F_{i:n}(x+h) - F_{i:n}(x)}{h} = \lim_{h \rightarrow 0} \frac{P[x < X_{i:n} \leq x+h]}{h} = \\ &= \lim_{h \rightarrow 0} \binom{n}{i-1, 1, n-i} (F(x))^{i-1} \frac{P[x < X_{i:n} \leq x+h]}{h} (1 - F(x+h))^{n-i} = \\ &= \binom{n}{i-1, 1, n-i} (F(x))^{i-1} f(x) (1 - F(x))^{n-i} \end{aligned}$$

Corolario 2.2.4 Si tenemos el caso particular del máximo, $X_{(n)} = \max(X_1, \dots, X_n)$, entonces:

$$\begin{aligned} F_{X_{(n)}}(x) &= P[X_{(n)} \leq x] = P[\max(X_1, \dots, X_n) \leq x] = \\ &= P[X_1 \leq x, \dots, X_n \leq x] = P[X_1 \leq x] \cdots P[X_n \leq x] = (F(x))^n \end{aligned}$$

En el caso de que las variables aleatorias sean además absolutamente continuas

$$f_{X_{(n)}} = n(F(x))^{n-1}f(x)$$

Capítulo 3

Esperanza condicionada, correlación y regresión

3.1. Esperanza condicionada

Definición 3.1.1 Consideremos el espacio probabilístico (Ω, \mathcal{A}, P) y un vector aleatorio $(X, Y)'$, entonces la **esperanza condicionada de Y a X** es una variable aleatoria cuyos valores vienen dados por

$$E[Y|X = x_n] = \sum_{y_m} y_m P[Y = y_m|X = x_n] \quad E[Y|X = x] = \int_{-\infty}^{+\infty} y f(y|x) dy$$

Propiedades 3.1.2 Veamos algunas propiedades que podemos extraer a partir de la definición.

1. Si $P[Y = c] = 1$ entonces $E[Y|X] = c$
2. Sea $a + bY$ entonces $E[a + bY|X] = a + bE[Y|X]$
3. $P[Y \geq 0] = 1$ entonces $E[Y|X] \geq 0$
4. Sean Y_1, Y_2 v.a.r. y $(X, Y_1)', (X, Y_2)'$ de forma que $P[Y_1 \geq Y_2] = 1$, entonces $E[Y_1|X] \geq E[Y_2|X]$.
5. Si $(X, Y)'$ son independientes, entonces $E[Y|X] = E[Y]$.
6. Sea $\varphi(X)$ medible Borel, entonces $E[\varphi(X)|X] = \varphi(x)$

7. Sea $\varphi(X)$ medible Borel, entonces $E[\varphi(X)Y|X] = \varphi(x)E[Y|X]$

8. Sea $h(Y)$ una función medible Borel para la cual existe $E[h(Y)]$, entonces $E[E[h(Y)|X]] = E[h(Y)]$

$$\begin{aligned} E[E[h(Y)|X]] &= E\left[\int_{-\infty}^{+\infty} h(y)f(y|x)dy\right] = \int_{-\infty}^{+\infty} f(x)dx \int_{-\infty}^{+\infty} h(y)f(y|x)dy = \\ &= \int_{-\infty}^{+\infty} h(y)dy \int_{-\infty}^{+\infty} f(y|x)f(x)dx = \int_{-\infty}^{+\infty} h(y)dy \int_{-\infty}^{+\infty} f(x,y)dx = E[h(Y)] \end{aligned}$$

3.2. Variancia condicionada

Definición 3.2.1 Consideremos el espacio probabilístico (Ω, \mathcal{A}, P) y un vector aleatorio $(X, Y)'$, entonces definimos la **varianza condicionada** como

$$\begin{aligned} V[Y|X] &= E[(Y - E[Y|X])^2|X] = E[Y^2|X] + E[(E[Y|X])^2|X] - 2E[YE[Y|X]|X] = \\ &= E[Y^2|X] + (E[Y|X])^2 - 2E[Y|X]E[Y|X] = E[Y^2|X] - (E[Y|X])^2 \end{aligned}$$

Teorema 3.2.2 (Teorema de Madow) Consideremos el espacio probabilístico (Ω, \mathcal{A}, P) y un vector aleatorio $(X, Y)'$ y supongamos que existe la varianza de Y , entonces

$$V[Y] = E[V[Y|X]] + V[E[Y|X]]$$

Demostración 3.2.3

$$\begin{aligned} V[Y] &= E[Y^2] - E[Y]^2 = E[E[Y^2|X]] - E[E[Y|X]]^2 = \\ &= E[E[Y^2|X]] - E[E[Y|X]^2] + E[E[Y|X]^2] - E[E[Y|X]]^2 = \\ &= E[E[Y^2|X] - E[Y|X]^2] + E[E[Y|X]^2] - E[E[Y|X]]^2 = E[V[Y|X]] + V[E[Y|X]] \end{aligned}$$

Corolario 3.2.4 Consideremos el espacio probabilístico (Ω, \mathcal{A}, P) y un vector aleatorio $(X, Y)'$, entonces

$$V[Y] \geq V[E[Y|X]] \quad 0 \leq \frac{V[E[Y|X]]}{V[Y]} \leq 1$$

Al segundo cociente le llamaremos **razón de correlación de Y condicionada a X** y se denota $\theta_{Y|X}^2$.

3.3. Regresión

3.3.1. Regresión no lineal

Definición 3.3.1 Consideremos el espacio probabilístico (Ω, \mathcal{A}, P) y un vector aleatorio $(X, Y)'$. Tratamos de buscar una función $h(X)$ que cumpla que

$$Y = h(X) + e_h$$

De forma que $E[e_h^2]$ sea mínimo, es decir

$$\begin{aligned} \min_h(E[e_h^2]) &= \min_h(E[(Y - h(X))^2]) \\ E[(Y - h(X))^2] &= E[(Y - E[Y|X] + E[Y|X] - h(X))^2] = \\ &= E[(Y - E[Y|X])^2] + E[(E[Y|X] - h(X))^2] + 2E[(Y - E[Y|X])(E[Y|X] - h(X))] = \\ &= E[(Y - E[Y|X])^2] + E[(E[Y|X] - h(X))^2] + 2E[(E[Y|X] - h(X))E[(Y - E[Y|X])|X]] = \\ &= E[(Y - E[Y|X])^2] + E[(E[Y|X] - h(X))^2] \end{aligned}$$

Que se minimiza cuando $h(X) = E[Y|X]$. Una medida del error es el **residuo de regresión** $e_{Y|X} = Y - E[Y|X]$. Lo ideal sería que el error, que es una variable aleatoria, tuviese esperanza 0 y una desviación lo más pequeña posible.

$$E[e_{Y|X}] = E[Y - E[Y|X]] = E[E[(Y - E[Y|X])|X]] = 0$$

$$\begin{aligned} V[e_{Y|X}] &= E[e_{Y|X}^2] = E[(Y - E[Y|X])^2] = E[E[(Y - E[Y|X])^2|X]] = \\ &= E[V[Y|X]] = V[Y] - V[E[Y|X]] = V[Y] \left(1 - \frac{V[E[Y|X]]}{V[Y]}\right) \end{aligned}$$

$$\text{Cov}(e_{Y|X}, g(x)) = E[e_{Y|X}g(x)] = E[g(x)(Y - E[Y|X])] = E[E[g(x)(Y - E[Y|X])|X]] = 0$$

3.3.2. Regresión lineal

Definición 3.3.2 *Mejor aproximación dentro de las funciones lineales. La recta de regresión. Sea (X, Y) queremos aproximar Y por una recta. Utilizaremos el método de los mínimos cuadrados para hacer mínimo el error.*

$$\begin{aligned}\min_{a,b} E[e_L^2] &= \min_{a,b} E[(Y - (a + bX))^2] \\ \phi(a, b) &= E[Y^2] + a^2 + b^2 E[X^2] + 2abE[X] - 2aE[Y] - 2bE[XY] \\ \frac{\partial \phi(a, b)}{\partial a} &= 2a + 2bE[X] - 2E[Y] = 0 \\ \frac{\partial \phi(a, b)}{\partial b} &= 2bE[X^2] + 2aE[X] - 2E[XY] = 0 \\ a &= E[Y] - bE[X] \quad b = \frac{Cov(X, Y)}{V[X]}\end{aligned}$$

Denominamos \mathbf{b} como coeficiente de regresión lineal y \mathbf{a} ordenada en el origen.

$$r_Y := y = E[Y] + \frac{Cov(X, Y)}{V[X]}(x - E[X])$$

Propiedades 3.3.3 *Veamos algunas propiedades*

1. *Existen otros modelos que se pueden ajustar al modelo lineal.*
2. *El coeficiente de regresión es la pendiente de la recta y tiene el mismo signo que la pendiente de la recta de regresión de X sobre Y .*
3. *Estas dos rectas se cortan en el $(E[X], E[Y])$.*
4. *Si $\sigma_{xy} > 0$ las variables crecen o decrecen simultáneamente, mientras que si $\sigma_{xy} < 0$ una crece mientras otra decrece.*
5. *Las rectas coinciden si hay variación funcional.*

$$\frac{\sigma_{XY}}{\sigma_X^2} = \frac{\sigma_Y^2}{\sigma_{XY}} \Rightarrow \frac{\sigma_{XY}^2}{\sigma_Y^2 \sigma_X^2} = \rho^2 = 1$$

6. *Si las variables son incorreladas $\rho = 0$, si hay correlación negativa $\rho = -1$ y si la hay positiva $\rho = 1$*

Proposición 3.3.4 *Analicemos ahora las propiedades de los residuos lineales. Comencemos viendo que la esperanza es nula.*

$$E[\varepsilon_L] = E[Y - (a + bX)] = E[Y - E[Y] + bE[X] - bX] = 0$$

Veamos ahora que $V[\varepsilon_L] = \sigma_Y^2(1 - \rho^2)$

$$\begin{aligned} V[\varepsilon_L] &= E[(Y - (a + bX))^2] = E[((Y - E[Y]) - b(x - E[X]))^2] = \\ &= E[(Y - E[Y])^2 + b^2(x - E[X])^2 - 2b(Y - E[Y])(x - E[X])] = \sigma_Y^2 + b^2\sigma_X^2 - 2b\sigma_{XY} = \\ &= \sigma_Y^2 + \frac{\sigma_{XY}^2}{\sigma_X^2} - 2\frac{\sigma_{XY}^2}{\sigma_X^2} = \sigma_Y^2(1 - \frac{\sigma_{XY}^2}{\sigma_X^2\sigma_Y^2}) = \sigma_Y^2(1 - \rho^2) \end{aligned}$$

Si $\rho^2 = 1$, el la varianza y el propio error son nulos. Si $\rho^2 = 0$, no la recta no aporta ninguna información.

Proposición 3.3.5 *Se tiene que*

$$\begin{aligned} V[\varepsilon_{Y|X}] &= V[Y](1 - \sigma_{Y|X}^2) & V[\varepsilon_L] &= V[Y](1 - \rho^2) & V[\varepsilon_{Y|X}] &\leq V[\varepsilon_L] \\ 0 &\leq \rho^2 \leq \sigma_{Y|X}^2 \leq 1 \end{aligned}$$

Si $\rho^2 = \sigma_{Y|X}^2$ trabajaríamos con la recta de regresión.

Capítulo 4

Distribuciones multivariantes

4.1. Distribución multinomial

Definición 4.1.1 Un *experimento multinomial* es un experimento aleatorio en el cual son posibles k resultados A_1, \dots, A_k que son además mutuamente excluyentes y exhaustivos. La suma de sus probabilidades tiene que ser, por tanto, 1 y las ejecuciones del experimento son independientes unas de otras. Puede verse como una generalización del experimento de Bernoulli. La distribución multinomial se presenta cuando realizamos n experimentos multinomiales. Recibimos entonces una variable X_i , que es el número de veces que se observa A_i al realizar n veces el experimento multinomial. Además $\sum_i X_i = n$.

$$P[X_1 = x_1, \dots, X_k = x_k] = \frac{n!}{x_1! \dots x_k!} \prod_{i=1}^k p_i^{x_i}$$

Si tenemos cuanto valen $k-1$ variables aleatorias, determinamos cuanto vale la restante. Por tanto, nuestro vector aleatorio tendrá $k-1$ componentes $X = (X_1, \dots, X_{k-1})'$. $X \in M(n, p_1, \dots, p_{k-1})$.

$$P[X_1 = x_1, \dots, X_{k-1} = x_{k-1}] = \frac{n!}{x_1! \dots (n - \sum_{i=1}^{k-1} x_i)!} \prod_{i=1}^{k-1} p_i^{x_i} \left(1 - \sum_{i=1}^{k-1} p_i\right)^{n - \sum_{i=1}^{k-1} x_i}$$
$$0 \leq x_i \leq n \quad i = 1, \dots, k-1 \quad 0 \leq n - \sum_{i=1}^{k-1} x_i \leq n$$

Veamos que función de probabilidad. Es mayor o igual que 0. Veamos que la suma da 1.

$$\begin{aligned}
\sum P[X = x] &= \sum P[X_1 = x_1, \dots, X_{k-1} = x_{k-1}] = \\
&= \sum \frac{n!}{x_1! \dots (n - \sum_{i=1}^{k-1} x_i)} \prod_{i=1}^{k-1} p_i^{x_i} \left(1 - \sum_{i=1}^{k-1} p_i\right)^{n - \sum_{i=1}^{k-1} x_i} = \\
&= (p_1 + p_2 + \dots + p_{k-1} + (1 - \sum_{i=1}^{k-1} p_i))^n = (1)^n = 1
\end{aligned}$$

Veamos ahora cuales son las marginales a través de la función generatriz de momentos.

$$\begin{aligned}
M_x(t) &= E[e^{t'X}] = E[e^{t_1 X_1 + \dots + t_{k-1} X_{k-1}}] = \\
&= \sum \frac{n!}{x_1! \dots (n - \sum_{i=1}^{k-1} x_i)} \prod_{i=1}^{k-1} (e^{t_i} p_i)^{x_i} \left(1 - \sum_{i=1}^{k-1} p_i\right)^{n - \sum_{i=1}^{k-1} x_i} = \\
&= (p_1 e^{t_1} + \dots + p_{k-1} e^{t_{k-1}} + (1 - \sum_{i=1}^{k-1} p_i))^n
\end{aligned}$$

Por ejemplo, si calculamos la marginal de X_1 utilizamos el vector $(t_1, 0, \dots, 0)$. Entonces $M_{X_1}(t_1) = (p_1 e^{t_1} + (1 - p_1))^n$, luego $X_1 \in \text{Bi}(n, p_1)$. Este resultado se puede generalizar al resto.

$$\begin{aligned}
E[X] &= (E[X_1] \dots E[X_{k-1}]) \quad E[X_i] = np_i \\
\sigma_{ii} &= \sigma_i^2 = \text{var}(X_i) = np_i(1 - p_i) \\
\text{Cov}(X_i, X_j) &= E[(X_i - E[X_i])(X_j - E[X_j])] = E[X_i X_j] - E[X_i]E[X_j] = \\
&= \frac{\partial^2 M_x(t)}{\partial t_i \partial t_j} \Big|_{t=0} - E[X_i]E[X_j] = n(n-1)p_i p_j - n^2 p_i p_j = -np_i p_j
\end{aligned}$$

Podemos ahora calcular las distribuciones condicionadas

$$\begin{aligned}
P[X_2 = x_2 | X_1 = x_1] &= \frac{P[X_2 = x_2, X_1 = x_1]}{P[X_1 = x_1]} = \\
&= \frac{\frac{n!}{x_1! x_2! (n - (x_1 + x_2))!} p_1^{x_1} p_2^{x_2} (1 - (p_1 + p_2))^{n - (x_1 + x_2)}}{\frac{n!}{x_1! (n - x_1)!} p_1^{x_1} (1 - p_1)^{n - x_1}} = \\
&= \frac{(n - x_1)!}{x_2! ((n - x_1) - x_2)!} \frac{p_2^{x_2} ((1 - p_1) - p_2)^{(n - x_1) - x_2}}{(1 - p_1)^{n - x_1 - x_2 + x_2}} = \\
&= \frac{(n - x_1)!}{x_2! ((n - x_1) - x_2)!} \left(\frac{p_2}{1 - p_1}\right)^{x_2} \left(1 - \frac{p_2}{1 - p_1}\right)^{(n - x_1) - x_2}
\end{aligned}$$

Por tanto, $X_2|X_1 \in B_i(n - x_1, \frac{p_2}{1-p_1})$ y su esperanza es $E[X_2|X_1] = (n - x_1)\frac{p_2}{1-p_1}$, que es variable aleatoria. Dado que es una recta, podemos decir que $\sigma_{X_2|X_1}^2 = \rho^2 = \frac{n^2 p_1^2 p_2^2}{np_1(1-p_1)np_2(1-p_2)} = \frac{p_1 p_2}{(1-p_1)(1-p_2)}$. Como el signo coeficiente correlación del signo de la covarianza, tomamos la raíz negativa a la hora de calcularlo.

4.2. Distribución normal n-variante

Definición 4.2.1 Sea $Z_i \in N(0, 1)$, $f(z_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z_i^2}$, $E[Z_i] = 0$ y $M_{Z_i}(t) = e^{\frac{1}{2}t^2}$. Consideramos el vector $Z = (Z_1, \dots, Z_n)'$. Z_i independientes. Se tiene entonces que

$$f(z) = \prod_i^n f_i(z_i) = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}z'z} \quad Z \in N_n(0, I_n)$$

$$E[Z] = 0 \quad \Sigma_Z = I$$

Veamos la función generatriz de momentos de la variable Z

$$M_Z(t) = E[e^{t'z}] = E\left[\prod_{i=1}^n e^{t_i z_i}\right] = \prod_{i=1}^n E[e^{t_i z_i}] = \prod_{i=1}^n e^{\frac{t_i^2}{2}} = e^{\frac{1}{2}t't}$$

Sea $X = \mu + AZ$ con $\mu \in \mathbb{R}^n$, $Z \in N_n(0, I)$ y $A \in \mathbb{R}^{n \times n}$, $|A| \neq 0$, entonces

$$E[X] = \mu \quad \Sigma_X = E[(X - \mu)(X - \mu)'] = E[AZZ'A'] = AE[ZZ']A' = AA'$$

$$|\Sigma_X| = |A||A'| = |A|^2 \Rightarrow |A| = |\Sigma_X|^{1/2}$$

$$J = \frac{1}{|A|} = \frac{1}{|\Sigma_X|^{1/2}} \Rightarrow f(x) = \frac{1}{(\sqrt{2\pi})^n |A|} e^{-\frac{1}{2}(x-\mu)'(A^{-1})'A^{-1}(x-\mu)}$$

Luego $X \in N_n(\mu, \Sigma_X)$. Calculamos la función generatriz

$$M_X(t) = E[e^{t'x}] = E[e^{t'(\mu + Az)}] = e^{t'\mu} E[e^{t'Az}] = e^{t'\mu} E\left[\prod_{i=1}^n e^{t_i A z_i}\right] =$$

$$= e^{t'\mu} \prod_{i=1}^n E[e^{t_i A z_i}] = e^{t'\mu + \frac{1}{2}t'A'At}$$

Si $|A| \neq 0$, la distribución es no singular y podemos calcular su densidad y la función generatriz de momentos. Si $|A| = 0$, la distribución es singular y no podemos calcular la función de densidad.

Pasemos ahora a calcular las marginales usando la función generatriz de momentos

$$M_{X_i}(t_i) = M_x(0, \dots, t_i, \dots, 0) = e^{t_i \mu_i + \frac{1}{2} t_i^2 \sigma_i^2}$$

Luego $X_i \in N(\mu_i, \sigma_i^2)$.

Corolario 4.2.2 La distribución normal es reproductiva, de forma que $X_i \in N(\mu_i, \sigma_i^2)$ independientes entonces

$$Y = \sum_{i=1}^n X_i \in N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$$

Demostración 4.2.3 Probemos el caso $i = 2$. Sea $t \in \mathbb{R}$.

$$M_Y(t) = E[e^{ty}] = E[e^{t(x_1+x_2)}] = E[e^{tx_1}]E[e^{tx_2}] = e^{t(\mu_1+\mu_2)+\frac{1}{2}t^2(\sigma_1^2+\sigma_2^2)}$$

Proposición 4.2.4 Sea X un vector aleatorio n -dimensional, la condición necesaria y suficiente para que X siga una distribución normal n -dimensional $X \in N_n(\mu, \Sigma)$ es que $\forall c \in \mathbb{R}^n$ $c'X \in N_1(c'\mu, c'\Sigma c)$.

Demostración 4.2.5 Si $X \in N_n(\mu, \Sigma)$ y sea $Y = c'X$, calculemos su distribución:

$$M_Y(s) = E[e^{sy}] = E[e^{sc'x}] = e^{s(c'\mu) + \frac{1}{2}s^2(c'\Sigma c)}$$

El recíproco es inmediato.

Definición 4.2.6 Veamos a continuación las distribuciones condicionadas. Sea un vector bidimensional $(X, Y)'$ con

$$\begin{aligned} E[(X, Y)'] &= (\mu_1, \mu_2)' \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \quad |A|^2 = \sigma_1^2 \sigma_2^2 (1 - p^2) \\ f(x, y) &= \frac{1}{(\sqrt{2\pi})^2 [\sigma_1^2 \sigma_2^2 (1 - p^2)]^{1/2}} \exp\left(-\frac{(x - \mu_1, y - \mu_2)}{2\sigma_1^2 \sigma_2^2 (1 - p^2)} \begin{pmatrix} \sigma_1^2 & -\sigma_{12} \\ -\sigma_{12} & \sigma_2^2 \end{pmatrix} \begin{pmatrix} x - \mu_1 \\ y - \mu_2 \end{pmatrix}\right) \\ &= \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_2(1-p^2)^{1/2}} e^{\frac{-1}{2\sigma_2^2(1-p^2)}(Y-(\mu_2+p\frac{\sigma_2}{\sigma_1}(x-\mu_1)))^2} \end{aligned}$$

De esta forma, es sencillo ver que

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}$$

$$f(y|x) = \frac{f(x, y)}{f(x)} = \frac{1}{\sqrt{2\pi}\sigma_2(1-p^2)^{1/2}} e^{\frac{-1}{2\sigma_2^2(1-p^2)}(Y-(\mu_2+p\frac{\sigma_2}{\sigma_1}(x-\mu_1)))^2}$$

Definición 4.2.7 Veamos finalmente la regresión. Por lo anterior, $Y|X \in N_1(\mu_2+p\frac{\sigma_2}{\sigma_1}(x-\mu_1), \sigma_2^2(1-p)^2)$. Luego

$$E[Y|X] = \mu_2 + p\frac{\sigma_2}{\sigma_1}(x - \mu_1)$$

Como es una recta, es de hecho la recta de regresión. Se cumple por tanto que $\sigma_{Y|X}^2 = \rho^2$.

Proposición 4.2.8 Si $(X, Y)' \in N_2((\mu_1, \mu_2)', \Sigma)$ son incorreladas, entonces son independientes. El recíproco también es cierto.