Probabilidad: recordatorio 2

PFG-JLF

UAM

Estadística I, 2018-2019

Modelo: vectores aleatorios

Caso discreto.

Definimos la función de masa conjunta del vector $(X_1, \ldots, X_n)$ como la colección de números (probabilidades conjuntas)

$$\mathbf{P}(X_1 = a_1, \ldots, X_n = a_n) \geq 0,$$

donde cada $a_i \in \operatorname{sop}(X_i)$, para $i = 1, \ldots, n$, tales que

$$\sum_{a_1 \in \operatorname{sop}(X_1)} \cdots \sum_{a_n \in \operatorname{sop}(X_n)} \mathbf{P}(X_1 = a_1, \ldots, X_n = a_n) = 1.$$

Cálculo de probabilidades: la probabilidad de que $(X_1, \ldots, X_n)$ tome valores en un cierto subconjunto $A \subset \mathbb{R}^n$ viene dada por

$$\sum_{(a_1,\ldots,a_n)\in A} \mathbf{P}(X_1 = a_1, \ldots, X_n = a_n).$$

Marginales: la función de masa de, por ejemplo, la coordenada X_1, viene dada por

$$\mathbf{P}(X_1 = \alpha) = \sum_{a_2 \in \operatorname{sop}(X_2)} \cdots \sum_{a_n \in \operatorname{sop}(X_n)} \mathbf{P}(X_1 = \alpha, X_2 = a_2, \ldots, X_n = a_n)$$

para cada $\alpha \in \operatorname{sop}(X_1)$.

Cálculo de medias: Dada una función $h : \mathbb{R}^n \to \mathbb{R}$, la media de la variable aleatoria

$$Z = h(X_1, \ldots, X_n)$$

es

$$\mathbf{E}(Z) = \sum_{a_1 \in \operatorname{sop}(X_1)} \cdots \sum_{a_n \in \operatorname{sop}(X_n)} h(a_1, \ldots, a_n)\, \mathbf{P}(X_1 = a_1, \ldots, X_n = a_n)\,.$$

Caso continuo.

El vector aleatorio $(X_1, \ldots, X_n)$ se define a través de una función de densidad conjunta

$$f_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n)$$

definida en $\mathbb{R}^n$ tal que

$$\begin{aligned} &\text{(no negativa)} \quad f_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) \geq 0; \\ &\text{(integral 1)} \quad \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n)\, dx_1 \cdots dx_n = 1\,. \end{aligned}$$

Cálculo de probabilidades: la probabilidad de que $(X_1, \ldots, X_n)$ tome valores en un cierto subconjunto $A \subset \mathbb{R}^n$ viene dada por

$$\int_A f_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n)\ dx_1 \cdots dx_n.$$

Marginales: las funciones de densidad marginal de cada X_i se calculan

$$f_{X_i}(x) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{(X_1,\ldots,X_n)}(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_n) \cdot dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n$$

(se integra en todas las variables excepto la i-ésima).

Cálculo de medias: la media de la variable aleatoria

$$Z = h(X_1, \dots, X_n)$$

se calcula en este caso como

$$\mathbf{E}(Z) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(x_1, \dots, x_n)\, f_{(X_1,\dots,X_n)}(x_1, \dots, x_n)\; dx_1 \cdots dx_n.$$

Independencia

Caso discreto: las variables $(X_1, \ldots, X_n)$ son independientes si

$$\mathbf{P}(X_1 \in A_1\,, \ldots, X_n \in A_n) = \mathbf{P}(X_1 \in A_1) \cdot \mathbf{P}(X_2 \in A_2) \cdots \mathbf{P}(X_n \in A_n),$$

para cualesquiera conjuntos (de Borel) $A_1, \ldots, A_n \subset \mathbb{R}$.

Caso continuo: las variables coordenadas X_j son independientes si y sólo si la función de densidad conjunta $f_{(X_1,\ldots,X_n)}$ se factoriza como producto de las funciones de densidad de las coordenadas X_j:

$$f_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)\,,$$

para cada $(x_1, \ldots, x_n) \in \mathbb{R}^n$.

Observación: si las variables

$$(X_1, \ldots, X_n)$$

son independientes, entonces las variables coordenadas del vector

$$(Y_1, \ldots, Y_n) = (T_1(X_1), \ldots, T_n(X_n)),$$

donde $T_1, \ldots, T_n$ son funciones medibles de $\mathbb{R}$ en $\mathbb{R}$, también son independientes.

Covarianzas

Dado un vector aleatorio $(X_1, \ldots, X_n)$, la covarianza entre las variables X_i y X_j se define como sigue:

$$\operatorname{cov}(X_i, X_j) = \mathbf{E}\big[(X_i - \mathbf{E}(X_i)) \cdot (X_j - \mathbf{E}(X_j))\big] = \mathbf{E}(X_i X_j) - \mathbf{E}(X_i)\, \mathbf{E}(X_j).$$

(el caso $i = j$ es varianza)

Independencia vs covarianza 0

Si X_i y X_j son independientes, entonces $\operatorname{cov}(X_i, X_j) = 0$ (pero al revés no, en general).

Varianza de sumas

Para $1 \leq i, j \leq n$ se tiene que

$$\mathbf{V}(X_i + X_j) = \mathbf{V}(X_i) + \mathbf{V}(X_j) + 2\,\mathrm{cov}(X_i, X_j).$$

Así que, por ejemplo, si X_i y X_j son independientes (y por tanto tienen covarianza 0), entonces

$$\mathbf{V}(X_i + X_j) = \mathbf{V}(X_i) + \mathbf{V}(X_j).$$

Varianza de combinaciones lineales

En general, si $a_1, \ldots, a_n \in \mathbb{R}$, entonces la combinación lineal

$$\sum_{j=1}^{n} a_j\, X_j$$

tiene varianza

$$\begin{aligned}
\mathbf{V}\Big(\sum_{j=1}^{n} a_j X_j\Big) &= \sum_{1\le i,j\le n} a_i a_j \operatorname{cov}(X_i, X_j) \\
&= \sum_{j=1}^{n} a_j^2\, \mathbf{V}(X_j) + \sum_{1\le i\ne j\le n} a_i\, a_j \operatorname{cov}(X_i, X_j)\,.
\end{aligned}$$

Coeficientes de correlación

El coeficiente de correlación entre X_i y X_j es

$$\rho(X_i, X_j) = \frac{\mathrm{cov}(X_i, X_j)}{\sqrt{\mathbf{V}(X_i)}\sqrt{\mathbf{V}(X_j)}}.$$

- $\rho(X_i, X_j)$ está definido sólo si X_i y X_j son variables no constantes, es decir, si $\mathbf{V}(X_i) \neq 0$ y $\mathbf{V}(X_j) \neq 0$.
- Si $\rho(X_i, X_j) = 0$, se dice que X_i y X_j están incorreladas.
- X_i y X_j independientes implica incorreladas (al revés no, en general).
- $-1 \leq \rho(X, Y) \leq 1$.

Notación matricial

$$\underbrace{\mathbb{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}}_{\text{vector aleatorio}}, \qquad \underbrace{\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}}_{\text{vector de } \mathbb{R}^n}, \qquad \underbrace{\mathbf{E}(\mathbb{X}) = \begin{pmatrix} \mathbf{E}(X_1) \\ \vdots \\ \mathbf{E}(X_n) \end{pmatrix}}_{\text{vector de medias}}.$$

(Más generalmente, si $\mathbb{M} = (X_{i,j})_{i,j}$ es una *matriz* de dimensiones $n \times m$ cuyas componentes son variables aleatorias, escribiremos $\mathbf{E}(\mathbb{M})$ para referirnos a la matriz $(\mathbf{E}(X_{i,j}))_{i,j}$ de medias de esas variables.)

Si $\mathbb{X}$ es un vector aleatorio de dimensión n, si $\mathbf{b}$ es un vector de dimensión n, y si A es una matriz $n \times n$, entonces

$$\mathbf{E}(A\mathbb{X} + \mathbf{b}) = A\,\mathbf{E}(\mathbb{X}) + \mathbf{b}.$$

Matriz de covarianzas

$$\mathbf{Cov}(\mathbb{X}) = \big(\mathrm{cov}(X_i, X_j)\big)_{1\le i,j\le n}$$

(denotando $\mathbf{V}(X_i) = \mathrm{cov}(X_i, X_i)$ para las varianzas).

Matricialmente,

$$\mathbf{Cov}(\mathbb{X}) = \mathbf{E}\big((\mathbb{X} - \mathbf{E}(\mathbb{X})) \cdot (\mathbb{X} - \mathbf{E}(\mathbb{X}))^{\mathsf{T}}\big).$$

Si A es una matriz $n \times n$ y $\mathbf{b}$ es un vector de dimensión n, entonces

$$\mathbf{Cov}(A\mathbb{X} + \mathbf{b}) = A\,\mathbf{Cov}(\mathbb{X})\,A^{\mathsf{T}}.$$

Matriz de correlaciones:

$$\boldsymbol{\Sigma}(\mathbb{X}) = \big(\rho(X_i, X_j)\big)_{1\le i,j\le n},$$

- La matriz de correlaciones tiene unos en la diagonal.
- sólo está definida cuando $\mathbf{V}(X_j) \neq 0$, para $1 \le j \le n$.

- La matriz $\mathbf{\Sigma}(\mathbb{X})$ es la matriz de covarianzas del vector $\widehat{\mathbb{X}}$, cuyas componentes son

$$\widehat{X}_j = \frac{X_j}{\sqrt{\mathbf{V}(X_j)}}.$$

- Matricialmente,

$$\mathbf{\Sigma}(\mathbb{X}) = D \cdot \mathbf{Cov}(\mathbb{X}) \cdot D,$$

donde

$$D = \begin{pmatrix} \frac{1}{\sqrt{\mathbf{V}(X_1)}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sqrt{\mathbf{V}(X_n)}} \end{pmatrix}$$

Sea $\mathbb{X}$ un vector aleatorio.

Tanto la matriz de covarianzas $\mathbf{Cov}(\mathbb{X})$ como la matriz de correlaciones $\mathbf{\Sigma}(\mathbb{X})$ de $\mathbb{X}$ son

- matrices simétricas
- y semidefinidas positivas.

Basta observar que, para cualquier $\mathbf{a}^\intercal = (a_1, \dots, a_n) \in \mathbb{R}^n$,

$$\begin{aligned}\mathbf{a}^\intercal \, \mathbf{Cov}(\mathbb{X}) \, \mathbf{a} &= \sum_{j=1}^{n} a_j^2 \, \mathbf{V}(X_j) + \sum_{i \neq j} \mathrm{cov}(X_i, X_j) \, a_i \, a_j \\ &= \mathbf{V}(a_1 X_1 + \cdots + a_n X_n) \geq 0.\end{aligned}$$

Si no es definida positiva, es porque para algún $\mathbf{a}^\intercal = (a_1, \dots, a_n)$ no nulo, la variable aleatoria

$$a_1 X_1 + \cdots + a_n X_n$$

es una constante.

Para la matriz de correlaciones,

$$\mathbf{a}^\top \mathbf{\Sigma}(\mathbb{X})\, \mathbf{a} = \mathbf{b}^\top \mathbf{Cov}(\mathbb{X})\, \mathbf{b} \geq 0\,,$$

donde $b_j = \frac{a_j}{\sqrt{\mathbf{V}(X_j)}}$ para $1 \leq j \leq n$.

Funciones de densidad de transformaciones lineales

- Sea $\mathbb{X}$ un vector aleatorio, con función de densidad conjunta $f_{\mathbb{X}}(\mathbf{x})$.
- Sea M una matriz $n \times n$ invertible, y sea $\mathbf{b}$ un vector $n \times 1$.
- Sea $\mathbb{Y}$ el vector aleatorio dado por $\mathbb{Y} = M\mathbb{X} + \mathbf{b}$.

Entonces

$$f_{\mathbb{X}}(\mathbf{x}) = f_{\mathbb{Y}}(M\mathbf{x} + \mathbf{b})\,|\mathrm{det}(M)|\,, \qquad \text{para todo } \mathbf{x} \in \mathbb{R}^n\,,$$

y también

$$f_{\mathbb{Y}}(\mathbf{y}) = \frac{1}{|\mathrm{det}(M)|}\; f_{\mathbb{X}}(M^{-1}(\mathbf{y} - \mathbf{b}))\;, \qquad \text{para todo } \mathbf{y} \in \mathbb{R}^n\,.$$

Funciones de densidad de la suma

Sea (X, Y) un vector aleatorio con función de densidad conjunta $f_{(X,Y)}(x, y)$.

La variable $Z = X + Y$ tiene función densidad

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_{(X,Y)}(x, z - x)\, dx.$$

Si X e Y son independientes, entonces

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x)\, f_Y(z - x)\, dx\,.$$

Teorema central del límite

Sea X una variable aleatoria con $\mathbf{E}(X^2) < \infty$. Llamamos $\mathbf{E}(X) = \mu$ y $\mathbf{V}(X) = \sigma^2$.

Consideremos una sucesión $(X_1, X_2, \dots)$ de variables aleatorias iid.

Interesa la variable

$$S_n = \sum_{j=1}^{n} X_j.$$

Su media y varianza son

$$\mathbf{E}(S_n) = n\,\mu, \qquad \mathbf{V}(S_n) = n\,\sigma^2.$$

El teorema central del límite nos dice que, cuando $n \to \infty$,

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1).$$

La convergencia es en distribución: para cada $t \in \mathbb{R}$,

$$\mathbf{P}\Big(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \le t\Big) \longrightarrow \Phi(t) \qquad \text{cuando } n \to \infty.$$

Si la variable fuera

$$Z_n = \frac{1}{n} \sum_{j=1}^{n} X_j,$$

cuya media y varianza son

$$\mathbf{E}(Z_n) = \mu, \qquad \mathbf{V}(Z_n) = \frac{\sigma^2}{n}.$$

el resultado sería el siguiente: cuando $n \to \infty$,

$$\frac{Z_n - \mu}{\sqrt{\sigma^2/n}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1).$$

En este curso escribiremos que, para la variable promedio

$$Z_n = \frac{1}{n} \sum_{j=1}^{n} X_j,$$

se tiene que, cuando $n \to \infty$,

$$\sqrt{n}\,\big(Z_n - \mu\big) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \sigma^2).$$