

Dado un espacio muestral $\mathcal{S}$, diremos que $\mathbf{X} = (X_1, X_2, \dots, X_k)$ es un vector aleatorio de dimension k si cada una de sus componentes es una variable aleatoria $X_i : \mathcal{S} \rightarrow \mathbb{R}$, para $i = 1 \dots, k$. Notemos que el vector aleatorio es una función $\mathbf{X} : \mathcal{S} \rightarrow \mathbb{R}^k$.

Para comenzar, trabajaremos con vectores aleatorios bidimensionales, es decir, haciendo $k = 2$. Utilizaremos indistintamente (X_1, X_2) o (X, Y) .

Función de distribución acumulada de un vector

Vectores Aleatorios

Vectores Aleatorios

Vectores Discretos

Vectores Aleatorios
Continuos

Considere un espacio muestral $\mathcal{S}$ y una probabilidad P definida en $\mathcal{S}$. Dado un vector aleatoria $(X, Y) : \mathcal{S} \rightarrow \mathbb{R}^2$, definimos la función de distribución (acumulada) conjunta asociada a (X, Y) , mediante la fórmula

$$F_{XY}(x, y) := P(X \leq x, Y \leq y) , \text{ para } (x, y) \in \mathbb{R}^2 .$$

Como haría para obtener la función de distribución acumulada de la variable aleatoria X conociendo la del vector (X, Y) ?

Un vector aleatorio $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_k)$ se dice discreto si cada una de sus componentes X_i lo es. Notemos que si $\mathbb{I}m_i$ denota la imagen de la i -ésima componente, tenemos que la imagen del vector es

$$\mathbb{I}m_{\mathbf{X}} = \mathbb{I}m_1 \times \dots \times \mathbb{I}m_k .$$

Función de probabilidad puntual conjunta

Vectores Aleatorios

Vectores Aleatorios

Vectores Discretos

Vectores Aleatorios
Continuos

Vectores aleatorios discretos, función de probabilidad puntual
CONJUNTA

$$p_{X,Y}(x,y) = P(X = x \cap Y = y) .$$

1- $0 \leq p_{X,Y}(x, y) \leq 1.$

2-

$$\sum_{y_j \in \mathbb{I}m_Y} \sum_{x_i \in \mathbb{I}m_X} p_{XY}(x_i, y_j) = \sum_{x_i \in \mathbb{I}m_X} \sum_{y_j \in \mathbb{I}m_Y} p_{XY}(x_i, y_j) = 1 .$$

3- Si A es un subconjunto de $\mathbb{R}^2$, tenemos que

$$P\left((X, Y) \in A\right) = \sum_{(x_i, y_j) \in A} p_{XY}(x_i, y_j) .$$

Observemos que conocer la función de probabilidad puntual conjunta nos dice mucho mas que conocer las puntuales individuales.

Como calculamos las puntuales marginales conociendo la puntual conjunta? EL proximo lema responde a esta pregunta.

Marginalización

Sea (X, Y) un vector aleatorio con función de probabilidad puntual conjunta dada por $p_{(X,Y)}(x_i, y_j)$. Tenemos entonces que la función de probabilidad puntual de X se obtiene mediante

$$p_X(x_i) = \sum_j p(x_i, y_j) .$$

Análogamente, tenemos que

$$p_Y(y_j) = \sum_i p(x_i, y_j) .$$

Ejercicio

En la siguiente tabla se presenta la función de probabilidad conjunta del vector aleatorio discreto (X, Y) :

$Y \setminus X$	1	2	3	4
1	0.10	0.05	0.02	0.02
2	0.05	0.20	0.05	0.02
3	0.02	0.05	0.20	0.04
4	0.02	0.02	0.04	0.10

- a) Hallar las funciones de probabilidad marginal de X y de Y : $p_X(x)$ y $p_Y(y)$ respectivamente.
- b) Calcular la probabilidad de que tanto X como Y sean menores que 3.
- c) Calcular la probabilidad de que X sea par e Y sea impar.

Recordemos que una variable aleatoria se dice continua cuando existe una función de densidad que nos permite calcular la probabilidad de que X este en A integrando sobre A la función de densidad:

$$P(X \in A) = \int_A f_X(x) dx ,$$

donde f_X verifica $f_X(x) \geq 0$ y $\int_{-\infty}^{\infty} f_X(x) dx = 1$.

Vectores continuos - Definición

Vectores Aleatorios

Vectores Aleatorios

Vectores Discretos

Vectores Aleatorios
Continuos

Diremos que el vector (X, Y) es continuo si existe una función $f_{XY}(x, y)$ de densidad , satisfaciendo que

$$P\left((X, Y) \in A\right) = \int \int_A f_{XY}(x, y) \, dx \, dy ,$$

para $A \subset \mathbb{R}^2$. En tal caso, diremos que f_{XY} es la función de densidad asociada al vector (X, Y) .

Notemos que si f_{XY} es la densidad del vector (X, Y) , tenemos que

$$F_{XY}(a, b) = P\left((X, Y) \in (-\infty, a] \times (-\infty, b]\right) = \int_{-\infty}^a \int_{-\infty}^b f_{XY}(x, y) \, dx \, dy$$

Derivando, tenemos entonces que

$$f_{XY}(a, b) = \frac{\partial^2}{\partial a \partial b} F_{XY}(a, b) .$$

Observemos que

$$P(X \in B) = P\left((X, Y) \in B \times (-\infty, \infty)\right) = \int_B \int_{-\infty}^{\infty} f_{XY}(x, y) \, dy \, dx .$$

Tenemos entonces una formula para la función de densidad marginal a partir de la densidad conjunta.

Marginalización - Caso Continuo

Vectores Aleatorios

Vectores Aleatorios

Vectores Discretos

Vectores Aleatorios
Continuos

Sea (X, Y) un vector aleatorio continuo con función de densidad conjunta dada por f_{XY} , entonces tenemos que las funciones de densidad marginales vienen dadas por

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy ,$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dx ,$$

Teorema

Vectores Aleatorios

Vectores Aleatorios

Vectores Discretos

Vectores Aleatorios
Continuos

Sea $g(X_1, X_2, \dots, X_n)$ un vector aleatorio n -dimensional.

Sea $g: \mathbb{R}^n \rightarrow \mathbb{R}$ una función. Consideremos la variable aleatoria $g(X_1, X_2, \dots, X_n)$. Entonces tenemos que

$$E\left[g(X_1, X_2, \dots, X_n)\right] = \sum_{x_1, x_2, \dots, x_n} g(x_1, x_2, \dots, x_n) p_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n)$$

si el vector es discreto y

$$E\left[g(X_1, X_2, \dots, X_n)\right] = \int \int \dots \int g(x_1, x_2, \dots, x_n) f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n$$

si el vector es continuo.

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

Mas generalmente,

$$E[h_1(X) + h_2(Y)] = E[h_1(X)] + E[h_2(Y)]$$

Este resultado se extiende a cualquier cantidad de variables:

$$E\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n E[X_i] ,$$

$$E\left[\sum_{i=1}^n h_i(X_i)\right] = \sum_{i=1}^n E[h_i(X_i)] .$$

Las variables aleatorias $X_1, X_2, \dots, X_n$ se dicen independientes si

$$P(X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_n) = \prod_{i=1}^n P(X_i \in B_i)$$

para todo $B_i \subset \mathbb{R}$.

Caracterización de independencia

Vectores Aleatorios

Vectores Aleatorios

Vectores Discretos

Vectores Aleatorios
Continuos

Las variables aleatorias $X_1, \dots, X_n$ son independientes si y solo si la función de distribución conjunta del vector $(X_1, \dots, X_n)$ se factoriza mediante la función de distribución de cada una de las coordenadas del vector:

$$X_1, \dots, X_n \text{ independientes} \iff F_{\vec{X}}(x_1, \dots, x_n) = \prod_{i=1}^n F_{X_i}(x_i)$$

Criterio de independencia para variables discretas.

Las variables aleatorias discretas $X_1, \dots, X_n$ son independientes si y solo si la función de probabilidad puntual del vector se factoriza mediante la función de probabilidad puntual de cada coordenada.

$$X_1, \dots, X_n \text{ independientes} \iff p_{\vec{X}}(x_1, \dots, x_n) = \prod_{i=1}^n p_{X_i}(x_i)$$

Criterio de independencia para vectores aleatorios continuos.

El vector aleatorio continuo $(X_1, \dots, X_n)$ tiene componentes independientes si y solo si la función de densidad conjunta del vector se factoriza mediante la función de densidad cada coordenada.

$$X_1, \dots, X_n \text{ independientes} \iff f_{\vec{X}}(x_1, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i)$$

Lemma

X, Y independientes entonces

$$E[XY] = E[X]E[Y]$$

Mas generalmente,

$$E[h_1(X)h_2(Y)] = E[h_1(X)]E[h_2(Y)]$$

Dadas dos variables aleatorias X e Y definimos la covarianza entre ellas mediante la fórmula

$$\text{cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] ,$$

siendo $\mu_X = E[X]$ y $\mu_Y = E[Y]$.

Fórmula Reducida:

$$\text{cov}(X, Y) = E[XY] - \mu_X \mu_Y ,$$

Corolario: X, Y independientes, entonces

$$\text{cov}(X, Y) = 0$$

$$\rho = \rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} .$$

- $Cov(X, Y) = Cov(Y, X)$
- $Cov(X, X) = V(X)$
- $Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$
- $Cov(aX, Y) = a Cov(X, Y)$
- $Cov(\sum_{i=1}^n a_i X_i, \sum_{j=1}^m b_j Y_j) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j Cov(X_i, X_j)$
- $|\rho| \leq 1$

Interpretación de la Covarianza

Vectores Aleatorios

Vectores Aleatorios

Vectores Discretos

Vectores Aleatorios
Continuos

Mejor predictor lineal para Y basado en X . Buscamos la mejor función lineal de X con la que predecir la variable Y . Es decir, buscamos $a, b \in \mathbb{R}$ de forma tal de minimizar el error cuadrático medio que cometemos al predecir la variable Y con la recta $aX + b$: queremos minimizar la función

$$G(a, b) = E [(Y - (aX + b))^2] .$$

Lemma

La función G se minimiza eligiendo

$$a^* = \frac{\text{Cov}(X, Y)}{\sigma_X^2}, \quad b^* = E[Y] - \frac{\text{Cov}(X, Y)}{\sigma_X^2} E[X].$$

EL mejor predictor lineal puede escribirse como

$$\mu_Y + \frac{\rho\sigma_Y}{\sigma_X}(X - \mu_X).$$

Además, el error cuadrático medio que cometemos al predecir Y con la mejor recta posible está dado por

$$G(a^*, b^*) = \sigma_Y^2(1 - \rho)^2.$$